

*NAGB Conference on Increasing the Participation of SD and LEP Students in NAEP*

**Commissioned Paper Synopsis**

*The attached paper is one of a set of research-oriented papers commissioned by NAGB to serve as background information for the conference attendees. The authors bear sole responsibility for the factual accuracy of the information and for any opinions or conclusions expressed in the paper.*

**Measuring the Performance of Students with Disabilities and English Language Learners in the National Assessment of Educational Progress**

**Albert E. Beaton**

Boston College

December 2003

- The purpose of this paper is to stimulate discussion of how to measure the educational progress of SD and ELL students who are deemed unable to sit for the NAEP assessment, even with accommodations.
- Opinions about excluding students with severe disabilities vary substantially from: “testing these children is unsound and hurtful to students and their parents” to “these students must be tested because with no testing there would be not accountability and with no accountability there would be not teaching and no learning.”
- Different decisions about testing students with disabilities might alter test results and skew the measurement of growth with a state or the interpretation of differences among states.
- In the early days of NAEP when results were not reported by state, students who were not deemed testable were just not put on the roster of those who could be sampled for assessment. During the design changes of 1983–84, the procedure was changed so that all eligible students were placed on the sampling roster and ten untestable students were excluded. After the introduction of reporting and comparing NAEP results by state, the procedures were tightened up considerably.
- The problem of including all students comes from setting the same standards—one size fits all—for all students. The idea of a common curriculum with common tests for all students is unlikely to meet the also attractive goal of having every student doing the best that he or she can possibly do.
- Using the analogy of a marathon running race where prizes are awarded separately for men, women, men in wheelchairs, and women in wheelchairs, the author proposes that each student with and IEP should have clear and objectively measurable goals and that NAEP develop an assessment to measure progress on how well the students accomplish reasonably achievable goals. NAEP could also develop and administer to ELL students an English language acquisition test that is standardized and agreed upon by participating states.

- NAEP would administer the new assessment to a random sample of excluded students in each state. The results of this data gathering would be a report on the accomplishments of SD and ELL students. This report would include estimates of the size of the population each state and the types of conditions that make exclusion necessary. Most importantly, this report would contain estimates of how well these students were reaching their academic goals.

Measuring the Performance of  
Students with Disabilities and  
English Language Learners in  
The National Assessment of Educational Progress

Albert E. Beaton

Boston College

January 9, 2004

## Introduction

The measurement of students who have disabilities (SDs) and those who are learning the English language (ELLs) are a matter of concern for the National Assessment of Educational Progress (NAEP) as well as the state testing programs required by the No Child Left Behind Act (NCLB). The stakes of the state testing programs are becoming higher as rewards are given to states or districts where students are improving and sanctions are placed on those that are not improving. The results of the NAEP assessment do not affect individual students or districts directly but have a role to play in establishing the credibility of state testing programs. The decisions about whether or not to test SD or ELL students may have an important effect on the results of state tests or national assessments.

The effect of excluding students from testing or assessment is obvious. If the SD and ELL students are very likely to do poorly on the test, then their poor scores will not be in the sample and thus the average score and the percentages of assessed students reaching basic or advanced levels will be increased. Having a large number of students who are excluded is not necessarily gamesmanship; a state with a strong concern for children with special needs might provide a large budget for remediation and make special services available to more children. The result might remove more students from testing situations.

The purpose of this paper is to stimulate discussion of how to measure the educational progress of SD and ELL students who are deemed unable to sit for the NAEP assessment, even with accommodations. As NAEP has become more relevant as a state and national educational indicator, it must adapt to the changing situation without sacrificing its accuracy. The next section of this paper will give a little background for the testing of these students, including mention of some other factors that are current in educational accountability. The following section sketches a proposal for measuring and reporting on the performance of excluded students.

## Background

Opinions about excluding students vary substantially. I recently spoke separately with two professors at Boston College<sup>1</sup> who have worked with SDs and I was quite surprised at how different their opinions were. One professor said that testing the Campus School children was educationally unsound and hurtful to the students and their parents. He added that failing was certain and that neither the student nor the parents needed another reminder. The other professor said that these students must be tested because with no testing there would be no accountability and with no accountability there would be no teaching and no learning. He added that the advocates for these children had worked hard to ensure their education and were not going to give up now.

The experience with the Campus School is extreme but underscores the different beliefs about excluding students from tests or assessments; such students cannot perform on the same tests even with accommodations. A more common problem is the practical question about marginal students, students that would be tested in one state or district but not in another. Different decisions might alter test results and skew the measurement of the growth of students within a state or the interpretation of differences among states.

The decision to exclude or not exclude a student is bounded by state laws. In some states, such as Massachusetts, all students must be tested whether or not they are SD or ELL students<sup>2</sup>. Even the students in the Boston College Campus School must be tested. In some other states, the testing decision for SD students is based on the student's Individual Educational Plan (IEP), and the exclusion considerations in such a decision vary from state to state. The exclusion policy for ELL students also varies from state to state. It is unlikely that a uniform exclusion policy will be agreed-upon by all states but my guess is that there will be fewer differences in the future.

Handling of excluded students in NAEP has changed over the years. In the early days of NAEP when results were not reported by state, students who were not deemed testable were just not put on the roster of those who could be sampled for assessment. Thus, the number and percent of excluded students could not be estimated. During the design changes of 1983-84, the procedure was changed so that all eligible students were placed on the sampling roster and then untestable students were excluded. Some minimal information was collected on these excluded students so that the number and percentage of excluded students could be estimated and some analyses could be done. After the introduction of reporting and comparing NAEP results by state, the procedures were tightened up considerably. Also, methods for computing "full population estimates" were explored for imputing the state results if the excluded students were actually assessed but this procedure has not been used in the main NAEP reports.

---

<sup>1</sup> Boston College has a Campus School that caters to multiply handicapped children, many of whom cannot speak or communicate with the outside world. Some students communicate only through computer devices.

<sup>2</sup> The Massachusetts legislature is reviewing this requirement at this time.

## Proposal

The opinions of the two professors at Boston College are seemingly so different as to be irreconcilable. However, we must recognize that the NCLB Act is concerned with educational accountability; it leaves the curriculum decisions and the standard setting with the states but requires that progress toward their goals be objectively measured. There seems to be a mistrust of subjective judgments of student progress.

The problem with including all students comes from setting the same standards—one size fits all—for all students. In many states, students must pass the same tests no matter whether they are in a trade or academic school, whether they have had an opportunity to learn the test material or not. The idea of a common curriculum with common tests for all students is unlikely to meet the also attractive goal of having every student doing the best that he or she can possibly do.

Fortunately, we do have procedures for many analogous situations. Consider the running of a marathon. In many marathons, the entrants are classified by gender. Within each gender some entrants with special needs are accommodated by allowing them to use wheelchairs during the race. The race is then run with four groups: men, women, men in wheelchairs, and women in wheelchairs. The four groups travel the same course. At the end, four first prizes are awarded, one for the winner of each group. Each prize is highly esteemed and reported by the press. There is no attempt to “equate” the timings, that is, to say what the timing of a wheel chaired man would have been if the entrant had run in the unaccommodated race.

Trying to set the same standards for all students without concern for their initial abilities is fraught with difficulties. At present, students are excluded from NAEP because they are deemed unable to respond meaningfully to the test. Unless there is some reason to doubt the exclusion judgment, it seems inappropriate to go ahead and test them anyway. Also, legally binding IEPs may force the exclusion of testing or assessment. It seems that we must look elsewhere to find a way to assure accountability for the education of SD and ELL students.

Therefore, the challenge is to find a way of setting individual goals for SD and ELL students and objectively measuring the attainment of these goals. For the SD students, using the present individualized educational plans is a reasonable way to start. Although these plans are often similar in intention and format, they are not identical across various states and districts and, as mentioned above, they are not likely to become so. They are, after all, *individual* educational programs and should be expected to vary somewhat. ELL students do not have IEPs and so a starting place is needed for determining progress toward appropriate attainment standards.

What is necessary is that each student with an IEP should have clear and objectively measurable goals. It is not the purpose of this paper to propose a blueprint for such goals but their development would clearly need to involve many persons who are knowledgeable about the education of SD and ELL students but also persons with

knowledge of educational measurement. The blueprint for IEPs will have to address the concern about the accountability of educators for the growth of SD and ELL students. Developing ground rules for such a blueprint will be a major task that will not be done overnight nor without the cooperation of various stakeholders.

I hope that this paper will generate suggestions about how a meaningful assessment could be produced. I think that it is clear that estimating the proficiencies or achievement levels of the excluded students on the NAEP scales is logically contradictory and should not be pursued. In many cases, giving the excluded students the usual test would be like giving a Sanskrit test to students who never studied the language. The aim on NAEP is to measure progress and I believe that progress should be measured by how well the students accomplish reasonably achievable goals.

Developing such an assessment will take time. Assuring the reliability and validity of the measurement tools will be difficult. It will also be important that stakeholders who have an interest in the education of special education students believe that the assessment will secure and advance the educational gains made over the last decades.

Such an assessment of special education students would ultimately require measurement of these students in tandem with the regular NAEP assessments. There should also be a period of feasibility testing while new ideas were tried out in the field.

This expanded assessment could be done with a fairly modest extension of the present NAEP sample. The present system requires that all SD and ELL students be listed along with the rest of the assessment population within a school. If possible, the students are assessed, with accommodations as required; otherwise, the students are excluded. A brief questionnaire is administered to the students or, more likely, their most relevant teacher. NAEP now supplies a trained assessment administrator who is responsible for proctoring the assessment and assuring quality control. He or she also is responsible for gathering the information about excluded students.

The quality control for an assessment of excluded students might work as follows. Using the lists of excluded students within a state, a random sample of these students would be selected. The sample would have to be carefully selected to assure that state estimates are unbiased. The size of the random sample might start at 100 students within each state but this number should change as more information becomes available. Specially trained observers would visit with each student and review the information of the questionnaires. Information about the reason for exclusion and the degree of handicap could be recorded for state and national population estimates.

Most importantly, the trained observer would review or measure the attainment of the special education students on their individual educational goals. Although the aims for individual students would vary depending on the type and level of the special need, in the end we would like an objective measure of whether the student reached the National Assessment Governing Board (NAGB) achievement levels; that is, reach a basic level of

accomplishment, was fully proficient, or achieved at an advanced level. These categories would be relative to the students' own abilities and not comparable to the categories in the regular NAEP.

The ELL students would also be included in the sample of excluded students. Perhaps an English language acquisition test that is standardized and agreed-upon by the participating states could be developed and administered. The student's performance could then be classified according to the NAGB achievement levels developed for students with equivalent opportunity to learn the English language.

The results of this data gathering would be a report on the accomplishments of SD and ELL students. This report would include estimates of the size of the populations in each state and the types of condition that made the exclusion necessary. Most importantly, this report would contain estimates of how well these students were reaching their academic goals.

The report on excluded students could be part of a regular NAEP report but might receive more attention if it were published separately.