

Looking Back and Looking Forward: Inclusion of All Students in the National Assessment of Educational Progress

**Rebecca Kopriva, University of Wisconsin
with Julia Lara, Independent Consultant**

Recognizing that the nation's economic well being is linked to the level of educational competency of its citizenry, policy makers at least since World War II have re-examined educational policies and practices to support the goal of a quality education for all. The re-examination has led to passage of notable pieces of legislation such as the Elementary and Secondary Education Act (ESEA) of 1964, the Individuals with Disabilities Education Act (IDEA) of 1975, and the Equal Education Opportunities Act of 1974.

Also during this period, a major demographic shift has taken place in the nation's public schools, resulting in greater representation of disabled students and in those who are racially and linguistically diverse. In 2006, racial and ethnic minorities comprised 43 percent of public school enrollment (an increase from 31 percent in 1986); children who spoke a language other than English represented 20 percent of all enrollments (up from 9 percent in 1979); and students receiving special education services comprised 9 percent of all children and youth from ages 3-21—up from 4 percent in 1977 (NCES, 2008a). In reviewing figures from the 2000 U.S. Census, a reporter for *The New York Times* noted that “the increase in the immigrant population (a 57 percent increase from 1990 Census figures), which many state officials believe was undercounted, surpassed the century's greatest wave of immigration, from 1900 to 1910, when the number of foreign born residents grew by 31 percent” (Scott, J, 2002, pg. A1, A20). Further, “for the first time

immigrants moved far beyond the big coastal cities and Chicago and Denver and Houston, into the Great Plains, the South and Appalachia.... ‘These numbers represent an enormous social experiment with very high stakes,’ said Steven A. Camarota, director of research for the Center for Immigration Studies,....‘and the experiment is not over’” (U.S. Census Bureau, 2002, pg. A20).

Partly as a consequence of federal leadership during the past 40 years, much has been accomplished to ensure that students enrolled in the public schools are participating meaningfully in all aspects of schooling and in eliminating barriers to a quality education. Nonetheless, policymaker and educators continue to confront challenges in their attempts to expand educational opportunities for all. One such issue is ensuring meaningful participation of English language learners (ELL) and students with disabilities (SD) in large scale student assessment programs, specifically the National Assessment for Educational Progress (NAEP). For the National Assessment Governing Board (NAGB) which is charged with overseeing the development, implementation and reporting of NAEP, the challenge is how to properly include special needs students in the NAEP assessments and ensure valid comparisons of scores across states. This issue is significant because NAEP is the barometer used by numerous public and private entities to judge the degree to which the nation’s school age population is meeting world class standards of educational performance. Given how the NAEP scores are used, it is critically important to NAEP to produce data on student achievement that is representative of the nation and the states assessed.

In this paper we will discuss briefly how NAGB has attempted to make NAEP participation representative of the population enrolled in the nation’s public schools while

maintaining the ability to defensibly make comparisons across states, the challenges it has confronted in this journey, and the issues that remain unresolved. The discussion focuses on two different, yet related aspects. The first is including ELL and SD students in the assessment so that their participation is consistently representative of the states and large urban districts on which NAEP reports. The focus in this aspect is on “who to include” so that results can be compared defensibly across jurisdictions. The second aspect involves “how to include”, and addresses how these students might meaningfully and validly access the test through the use of accommodations which are properly selected, administered and monitored. In this way results can be considered comparable across diverse types of students. It is argued in this paper that *both* of these aspects are necessary in order to draw proper comparisons. The third section will present a series of recommendations about how to proceed that is meant to augment suggestions introduced by others and outlined in the first two sections.

1. Who to Include: Representational Participation Rates

NAGB has been wrestling with how to best represent students with disabilities and English learners in the National Assessment of Educational Progress for over ten years, and the section below will touch on some of the primary considerations. Underpinning these considerations are the approaches that states and schools use to make decisions about inclusion on statewide tests, and the fact that these approaches have matured over time in response to crucial federal legislation. Specifically, state policies of inclusion have adapted to two sets of Elementary and Secondary Education Act reauthorization regulations and guidance (first under IASA and then under NCLB) for including special needs students on state tests, specifying the types of testing for some SD students, and

constraining how both SD and EL students identified to be included on the mainstream tests might be accommodated.

Nevertheless, two related factors have remained relatively constant for states over this time. The first factor is at what educational agency level the decisions about inclusion are made, and the second factor is who the decision makers are. For students with disabilities, the level of decision-making is mandated by the 1997 IDEA legislation, which specifies that local IEP teams are to make the decision for all testing, including tests originating at the federal level (although guidelines are not often explicitly defined at this level (Shakrani & Roeber, 2009)). IDEA also identified the decision-makers for students with disabilities. For ELs, states generally followed suit and produced guidelines which specified that educators at the local level should be in charge of making inclusion decisions. As compared to students with disabilities, the exact configuration of the decision-makers (e.g. the content teacher, the ELL specialist and/or a team) differs from state to state and often within states.

The primary question regarding who to include in NAEP has been how to accurately reflect the educational agencies' school populations, including most SD and EL students. For various well-considered reasons NAEP has increasingly reflected the effects of the state and local policies in their guidelines. However, ESEA authorizations mandate participation and NCLB specifies explicit summative percentages that must be met at the state level. Both of these requirements have the effect of constraining the local decisions to include or not include students. NAEP, on the other hand, is a voluntary test administered to a well-defined sample of students per state, or in some cases districts. Its legislation does not mandate specific summative percentages or other avenues. If one or

both of these were to be added to the legislative language, they also could either have the effect of limiting the non-consistent aspects associated with a local decision-making culture and/or otherwise encourage representative participation on a voluntary test. As it stands currently, inclusion rates in NAEP continue to radically differ among states with similar profiles, and efforts to minimize these differences to-date have been largely unsuccessful.

Historical Summary

Following is a brief synopsis of how inclusion has been handled to-date, some problems that have emerged, and some possibilities that have been studied. In general, just as concerns about interpretation of results have increased, rates of inclusion have dramatically increased as well. For instance, participation of special populations went from 40% in 4th grade reading in 1992 to 94% in 2007; in 4th grade math, rates increased to 97% in 2007 (Fields, 2008).

Prior to 1996, students with disabilities and English learners were allowed to participate in NAEP, however no accommodations were allowed and so large numbers were waived out of the test. An analysis of EL and SD participation rates prior to 1998 showed that approximately half of all these students enrolled in US schools were excluded from NAEP in the 1992 and 1994 assessments (Mazzeo et al., 2000). Often, those participating reflected English learners with high English proficiency and SD's with less severe disabilities, clearly not representative samples, and centrally at odds with the core mission of the NAEP program. As noted above, changes in the federal disability law (IDEA, 1997) encouraged movement towards inclusion and test accommodations for students with disabilities. The amended IDEA required that states report the types of

accommodations offered to students in state tests, and the IASA authorization of ESEA (in 1994) required inclusion of all students, both SDs and ELs. In an attempt to align itself with state policies, conform to IDEA and ESEA educational trends, and inform local decision makers, NAEP released inclusion guidelines for the 1998 administration. The development of the guidelines was preceded by several working meetings and research studies designed to inform NAEP staff thinking and assess the effects of inclusion policies on NAEP results.

The guidelines instructed local administrators to include EL and SD students in NAEP and offer accommodations (until 2002, a split sample design was used in reporting where accommodations were permitted for half of the sample; NCES, 2008b). The guidelines for students with disabilities directed local officials to include students unless one of the following conditions was present: the IEP team recommended student not participate; student's cognitive function was severely impaired, **or** IEP team recommended accommodation(s) not offered by NAEP. For English learners the rules *suggested* that they be included unless: student had received reading and mathematics instruction in English for less than three years prior to NAEP administration **and** student could not demonstrate knowledge of subject in English even with accommodations permitted by NAEP (NCES, 2008b). A budget was created to assist local level officials with inclusion questions and with the administration of NAEP under accommodated conditions, including training coordinators at the state level, developing additional source materials, and providing support for increased personnel, space, or time associated with using accommodations during the test.

As predicted, a combination of selection guidelines and the availability of a very limited number of accommodation options resulted in an increase in the rate of inclusion nationally. However, across states there was a wide variability in the rates in which EL and SD students were included in NAEP. This raised questions about the degree to which states were applying the rules consistently, and the results called into question the degree to which the state NAEP scores could be defensibly compared nationally. If states with lower rates of inclusion had higher scores in NAEP assessment, was it because they left out the students most likely to perform at lower levels of the performance scale?

In response, NCES and the interested staff from the state of Kentucky re-examined selected NAEP performance data with studies that produced conflicting findings. The NCES sponsored study (Lutkus, Mazzeo, Zhang, & Jerry, 2004) concluded that there was a relationship between higher exclusion rates in KY and the higher scores. The study commissioned by the state of Kentucky disputed this finding, reporting that that the excluded students performed above the lowest levels in other exams and would likely not have lowered the NAEP scores had they been included (Feinberg, 2008). Other related investigations found similar and different findings (e.g. see Hoff, 1999). As a result, in 2000, NAGB recommended that when exclusion rates exceeded 3% that a cautionary note be added to NAEP reports. However, the policy was reversed in 2001, and a sentence alerting consumers to possible comparability questions were added to all state reports. Since that time, it appears that a number of language alerts have been variously considered and used.

Meanwhile, changes in the ESEA federal educational policy landscape facilitated the development of greater uniformity in state practice regarding the inclusion of English

learners and students with disabilities in state assessments. In 2001, with the NCLB reauthorization of ESEA there was a strengthening of the accountability requirements for states, districts and schools receiving funds under this Act, mandating that, for accountability purposes, 95% of all subgroups must be included in the statewide academic assessments. Additional interpretations led to the “1% and 2% rules”. These rules specified that no more than 1% of significantly cognitively impaired students could be included using an alternate assessment, while no more than 2% of students could be included in other tests which measure grade level content standards but are scored according to modified achievement standards. Further, a study of the 2001-06 PIRLS international assessment using the NCLB standard for inclusion found that the exclusion rates for most countries was below 5% (Mullis, Martin, Kennedy, & Foy, 2007). Therefore, both nationally and internationally, a 5% cap in exclusion was, de facto, gaining currency and credibility.

Efforts to bolster representative inclusion in NAEP continued. In 2002 a panel was convened whose charge was to improve the representation rates of inclusion in NAEP (Shakrani & Roeber, 2009). They produced adjustments to the initial decision rules for participation of SD and ELs and provided further guidance regarding who to include and how. In this revision, both the SD and EL decision trees first ask if and how the students are tested within the state assessment system. If they take the state test and were not accommodated in the state system it is expected these students would take the NAEP assessment. For those students who are tested in the state assessment system and receives accommodations on the state test (or, for some SDs, if they took a modified or alternate test), the local decision-maker(s) are asked to review the NAEP

accommodations and decide if the particular student can be tested on NAEP. The decision maker can choose to include the student because either the state allowed accommodations are the same as those NAEP allows or that the NAEP set is acceptable for testing this student on NAEP. Otherwise, the student is waived. For EL students, even if they did not participate in the state test, the decision-maker could recommend they take NAEP if they judge that the accommodations are acceptable. These revised rules were implemented in NAEP state and district administrations beginning in 2003.

An examination of the 2003, 2005, and 2007 exclusion rates in reading at the 4th grade level for both ELs and SD students shows that the average rates of exclusion at the national level in 2003 was 6% versus 8% for large central cities; in 2005 it was 7% at the national level and 8% at the district level; and in 2007 it was 6% nationally and 7% at the district level. However, once again, release of the state and district level reports indicated that the numbers of students included in the assessment varied significantly across jurisdictions (NCES, 2008), and that for some states with similar profiles, discrepancies even increased (Shakrani & Roeber, 2009). When state trends were analyzed, low to moderate correlations were found between changes in the rate of exclusion and average score gains from 2003 to 2005 (Wise, Hoffman, & Becker, 2006). One study of the 2005 results found that 4th and 8th grade reading and math test exclusion rates in states ranged on average from (1% to 13%), and that, in some states, 50% or more of the percentages of EL or SD students enrolled in those grades did not participate in NAEP reading (School Matters, December 2005). Trial Urban District Assessment (TUDA) results also reflected variability. For example, in the 2007 4th grade NAEP reading assessment, Houston, (17%), Cleveland (17%), and Austin (20%) had higher exclusion

rates than San Diego (4%), LA (3%), and Charlotte (4%; NCES, 2008). Not surprising, questions continued to be raised about the rigor and accuracy of the cross-state and cross-district score comparisons.

In 2007, the GAO analyzed the state inclusion rates to investigate the source of state variability on the 2003 and 2005 NAEP, and concluded that state inclusion rates of SD differ because “the proportions of students with different types and severities of disability differ across states, and accommodations offered by the states in state tests also vary whether or not they are permitted by NAEP (Kitmitto & Bandeira de Mello, 2008).” Over 13 studies have been commissioned to examine how to address the dual impacts of varying exclusion rates and variability in enrollment, and several solutions have been proposed. Using imputing algorithms, one option to address the exclusion issue is to produce official report scores using estimates of full inclusion (as if all students had participated), as compared to the current practice where rates of excluded special needs students vary. Full population estimates (FPEs) from 2003 for state results and district results from 2005 are already computed but are not currently used as the official scores. Analysis by the National Institute of Statistical Sciences (McLaughlin & Bandiero de Mello, 2008, as cited by Goldstein, 2008) of the full population estimates indicates the impact of using them appears modest—most of the changes in scores are less than 2 points from the previous testing year, sometimes a relevant difference and sometimes not. ETS researcher, Henry Braun, used a slightly different method from McLaughlin and Bendiero de Mello but found similar results. Dr. Goldstein from NCES reported that findings for 4th grade reading comparisons from 2005 to 2007 indicated 10 jurisdictions had significant gains under the current approach which were not significant when FPE

were used, 8 had significant current and FPE gains, and none had non-significant gains under both conditions.

Second, in a letter to NAGB the chief state school officer from the state of West Virginia, suggested the following approaches to address both the exclusion and enrollment variations: a) identification of enrollment numbers for students with disabilities for all states; b) an exclusion of students whose IEP call for an alternate assessment; and c) based on 2007 NAEP data, a definition of a nationwide maximum, consistent special education participation for all performance calculations based on the percentage found in the state with the lowest identified percentage (Paine, 2007 as cited in XX, 2008). Third, Cavanagh (2008) reported that Andrew Porter and some other NABG members would like to see standard NAEP accommodation policies set across jurisdictions to address the exclusion issue, but others question whether states would agree as participation in NAEP is voluntary. Fields (2008) suggested that the 95%, 1% and 2% rules for SD students under NCLB may provide some ideas of a criterion for NAEP, and that, perhaps jurisdictions not meeting this criterion would be flagged in some way. Goldstein (2008) reported that a study has been commissioned to develop an estimate of what proportion of disabled students could reasonably be expected to be included given a) the prevalence of students with different types and severities of disabilities, and b) given that the variation in accommodations permitted by states in their testing programs. Actual versus expected rates of inclusion will be compared.

Fourth, Gorman (2008) recently discussed an approach for identifying ELs earlier proposed by Bernstein to NAGB but subsequently rejected (and adopted later by the National Assessment of Adult Literacy in 2003). In this approach an English language

proficiency screener exam would measure oral fluency to determine, at an established cutpoint, which ELs should be included in NAEP. While Gorman cautioned that logistics, cutpoints, conflicts with school policies, and possible legal issues would need to be addressed, there is also concern that oral literacy is not closely associated with written literacy as K-12 students increase beyond very rudimentary proficiency levels (see Kopriva, 2008 for related literature).

2. How to Include: Meaningful Access

Beyond issues associated with variability in participation, the second core consideration has, to-date, been largely absent from the controversies surrounding NAEP. This is the *quality* of inclusion and how English learners and students with disabilities are allowed to take the assessments so that their scores accurately reflect their knowledge and skills. This consideration is key to producing scores that are actually comparable in meaning, although producing scores where common inferences can defensibly be made across student populations is complex.

It seems that NAGB and NAEP are increasingly comfortable with holding jurisdictions accountable for all their students, as they thoughtfully increase the range of optional conditions to accommodate students with various kinds of challenges. The challenge at this point seems to be in improving how the connections between student and proper accommodations are made, and in increasing oversight associated with decisions and implementation. To-date this is something that the states have not suitably addressed, although federal reviews are increasingly inspecting state policies, procedures and documentation of accommodation effectiveness for particular students in the hopes of improving the quality of the scores for populations who take the test under varying

conditions. NAEP could take a leadership role in improving the quality but it takes a comprehensive set of strategies to implement properly. What is clear is that both representative participation and accurate measurement are essential components for making proper comparisons. Invalid measurement with representative participation rates, or the converse of effective measurement with skewed participation should be equally untenable.

Historical Summary

As noted in Section 1, a limited number of accommodations were allowed for some students beginning in 1998, although prior to 2002 accommodations were allowed only in subjects where new trend lines were being introduced (writing and civics). However, starting in 2002 accommodations were offered in NAEP for all subjects and all students, and the practice of drawing non-accommodated samples was terminated (NCES, 2008b). It is apparent that the number and appropriateness of accommodations continues to grow as a suitable set of empirically-based and doable options are identified, and as the states and districts become more comfortable with using accommodations for their students. For English learners, the primary adjustment has been to recognize that the test-taking challenges of students with disabilities and English learners are not the same, and that ELs do not benefit from many of the accommodations listed for SDs. Further, recent decision trees for English learners distinguish between accommodations which provide direct linguistic support and those which provide indirect support (see Rivera & Collum, 2006 for an explanation). Indirect accommodations without the direct linguistic support are typically not helpful for this population.

The latest decision trees identify approximately 20 accommodations that are offered to students with disabilities in at least some content areas, and about 12 that are presently offered to English learners. At the present time, though, NAEP provides only Spanish oral and written translations of the assessments and directions for ELs or dual Spanish/English texts, although the school can provide bilingual word-for-word glossaries in other languages as well as Spanish. English read-aloud is provided by NAEP as well for both ELs and SDs.

Nonetheless, while the range of accommodations that can be used during NAEP testing has improved, and more SDs and ELs are using these accommodations, ongoing evidence regarding their effectiveness is sobering. Researchers, including the work cited by Thurlow and others (2002, 2004), and Rivera and Collum (2006), document that, over the last several years, usage continues to vary tremendously within and across states in accommodations offered to and implemented with students with similar profiles. These analyses suggest that quite a bit more must be done to address the inequities associated with how students are included.

NAEP seems poised to continue to increase the number of allowable accommodations as proper evidence and feasibility warrants their inclusion. Recent background questionnaires ask questions about accommodations offered on the state test and who is providing the NAEP-allowed accommodations during administration of the tests. Fields (2008) suggested to NAGB committee members charged with addressing the special populations issue that periodic examinations of accommodations permitted in NAEP may be in order for both populations. Also in 2008, NAGB convened a panel of SD and EL accommodation experts to discuss how these populations might be best

accommodated on the NAEP writing assessment that will be implemented in 2011. Discussions not only revolved around this assessment, but broached the complexity of how to meaningfully include these students in various NAEP tests, and what might be done to address this complexity. Further, since the development of the 2004 NAEP mathematics framework, guidelines for writing items in all subject areas has included an addendum on how to construct more accessible items, explained how reviews might better use the expertise of special needs populations, and made suggestions for how data from piloted items might be interpreted for these subgroups (for example, see Cook & others, 2008). Over the last few years increased training of coordinators and improvements in test administration materials associated with issues of who to include and how signals NAEPs interest as well.

To-date, what hasn't undergone a comprehensive examination within the NAEP literature appears to be what all needs to be consistent regarding improvements in how students are accommodated and how this consistency needs to be managed. In order to address these issues, it seems that related literature inside and outside of the NAEP experience might provide some guidance. For special needs students to be properly included in mainstream tests, it is proposed that three elements need to be present (see Kopriva, 2008). First, consistent policies need to be articulated which identify a parsimonious set of test accommodations (or other conditions) which are appropriate to use for students with disabilities and a set of accommodations (or other conditions) for English learners. Second, reputable sources which target how appropriate accommodations are matched to particular EL and SD students who require them need to be identified and disseminated,

and use of these sources to make assignment decisions should be monitored for consistency within and across jurisdictions. Third, consistent implementation of accommodation conditions across jurisdictions and students with similar profiles needs to be mandated and monitored.

The argument for common inferences from traditional tests where no accommodations were generally permitted had been on procedural grounds: in order to generalize the interpretation of scores from different students, common content in items and a common approach for synthesizing and summarizing items and response data over items needed to occur. This required one standardized conditions of observation as a key aspect of synthesizing item data. The new argument for condition variations for special populations can be made from theories surrounding Evidence Centered Design (Mislevy & others, 2002, 2004). Here the case is made for how different kinds of observations (across any kind of students) can be built into a comparable testing framework. It is built on evidencing appropriate inter-relationships between target inferences, the knowledge and skills of interest, properties of tasks or items designed to elicit the observations, and the assessment situations where students interact with assessment requests. For NAEP, this means that these kinds of evidence need to be collected and evaluated. This in turn requires that a comprehensive plan associated with condition variations needs to be put into place to specify, oversee and analyze the data collections.

In addressing the first element, NAGB has demonstrated that it is interested in identifying which test accommodations are appropriate for English learners and students with disabilities. Investigations into both SD and EL accommodation options was hampered initially by researchers who typically focused on an imprecise question (Which

accommodations are effective or not for ELs or SDs?) as compared to the related but more central and targeted question (Which accommodations are effective for EL or SD students with a particular set of needs?). As the field has matured, NAEP appears to be open to guidance about which accommodations might be most appropriate to choose from for both of these populations. Additionally, as noted above, procedural guidance about writing accessible items, training, and test administration for students with disabilities or ELs have been put in place. These steps provide a firm foundation for building a stable of appropriate accommodation choices.

The second element addresses consistency in how specific accommodations are matched to students with specific sets of needs, and monitoring to make sure the decision-making is implemented properly. Over the years various researchers have argued and provided evidence that teams or individual teachers did not satisfactorily or consistently have the expertise to assign particular accommodations to individual students that are most appropriate for them (for example see, Fuchs, Fuchs, Eaton, Hamlett, Binkley, et al., 2000; Helwig & Tindal, 2003; Sirici, Li, & Scarpati, 2003; Douglas, 2004; Koran & Kopriva, 2006). Recently, two evaluations investigated how educators with expertise in assigning accommodations for students with disabilities (Plake & Impara, 2006), and EL teachers (Koran & Kopriva, 2006) would assign accommodations when they were given clear and detailed guidance. In both cases, results were troubling. In Koran and Kopriva findings illustrated that teacher judgments were no different than random assignment of accommodations to students. For NAEP, it seems clear that cross-state inferences would be influenced by assignment errors made at the local level within states, compounded with assignment errors made systematically across states.

There has been some movement in ‘mapping’ or otherwise providing very detailed guidance which students should receive which condition variations. Kopriva and others (2005, 2006, 2007, 2009) have designed and empirically tested a mapping system which recommends specific accommodations to English learners with particular needs. In a small sample, Kopriva, Emick et al. (2007) found that individual EL students who received proper accommodations suitable to their needs (as defined by the assignment system that took into consideration several aspects of the students’ challenges and strengths) performed significantly better than their peers on a mathematics test. On the other hand, those who received inappropriate accommodations scored no better than students with no accommodations. For students with disabilities, Fuchs, Fuchs and others (2000, 2000, 2005) have developed an inductive approach but no maps per se; Tindal, Ketterin-Geller, and Helwig (2003, 2003, 2006) have drafted and tested out versions but nothing definitive has been made available. South Carolina has recently completed a process-oriented document for assigning particular accommodations to students with disabilities which takes decision-makers step by step through a series of specific student background and schooling questions (Foster, Hall, & Elsmann, 2008). Because of the history of relying on IEP teams, producing even guidance (much less definitive) maps for IEP teams to help them make final decisions is controversial for this population. A detailed summary of the research associated with assignment issues for both populations can be found in Kopriva and Koran (2008).

The third element focuses on consistent implementation of accommodation conditions. As NCLB has continued to review the states’ summative assessment systems to ensure all students are being represented properly and proper score inferences can be

made for all students, one central element in their reviews has been to evaluate how states are overseeing the implementation of accommodations statewide (USED, 1999, 2004). According to NAEP materials (e.g. WESTAT materials disseminated at coordinator training sessions), this aspect is being addressed by NAEP state coordinators and contractors as well, although it is unknown how the oversight is documented, especially over states, what over-site evaluations are being done to ensure that oversight is properly handled, and what are the consequences if a locale is not in compliance. It appears that NAEP is aware of the importance of this element, and that steps are being taken to ensure the consistent implementation of accommodations over sites.

3. Recommendations

Several well-considered recommendations have been posited, some of which have been mentioned above. Below are a few recommendations to think about in conjunction with the others, as the National Assessment Governing Board wrestles with these important issues.ⁱ

Who to Include

Proposed is a three step process for increasing the representative participation of SDs and ELs in NAEP assessments:

1. The focus of participation is that state/district samples would reflect the percentages of ELs and SD's tested on their state test, with the following adjustments. This step ensures the proper participation rates.
 - a. For students with disabilities, the number of students in the 1% and 2% categories would be subtracted from the percentages of SDs tested in

the state testing system until modified versions are available (It is questionable whether additional versions should be a top priority as compared to addressing the other issues). Compliance would be monitored by NAEP.

b. ELs who have been in the state schools less than a year should be omitted from NAEP participation. More states now have completed or are in the process of completing state-level identifications for students, and these, rather than district numbers, would be the focus of compliance.

Compliance would be monitored by NAEP.

c. In addition, independent NAEP sponsored panels would determine student and schooling profiles where students who fall within these profiles would not be adequately served by the existing accommodations. States could also subtract these students if they can show that the students fit the profiles. Also, states can make case-by-case claims if they believe a student who falls outside the profiles should not be tested on NAEP, but accepted claims should be relatively rare—this means that unless the student receives some specialized classroom accommodation clearly not offered by NAEP, reasonable substitutions don't count...The burden of proof is on the states. Review of claims and compliance would be monitored by NAEP.

2. After the numbers of students corresponding to 1a, 1b and 1c are subtracted from the state percentage of students tested, samples of SDs and ELs corresponding to the final rates would be sampled by WESTAT as part of their sampling plan per state (Oversampling is suggested). It will be important

that WESTAT not only disaggregate by EL or SD but also that they develop more complex strata to reflect the representative mix of types of ELs and SDs in the state. In other words, a greater proportion of high English proficient ELs than the state school demographics reflect is not acceptable, and should not be substituted for a representative mix of low English proficient, middle English proficient, and high English proficient students. This step ensures the proper participation ranges. Compliance would be monitored by NAEP.

3. If states waive too many students or otherwise do not provide representative percentage rates or ranges, there would be oversampling in place to address the first shortfall. If the percentages and ranges are still not adequate, the suggestion is to impute or otherwise re-sample or both. This step is backup to ensure representational rates and ranges. Compliance would be monitored by NAEP.

How to Include

1. NAEP should continue to vet promising accommodations (including other forms or methods of collecting information), add them to their stable of allowable accommodations, and as possible have them provided by NAEP.
2. Systematic selection of appropriate accommodations for students who need them must be improved. NAEP is especially vulnerable because results will be compared directly across states, and systematic biases could provide a non-negligible source of error. NAEP, with advice, should decide on a plan of action for providing much more precise tools and/or guidance (such as what was summarized above) to local decision-makers to address this need.

Final assignments should be monitored to reflect that tools/guidance are used properly (albeit with local oversight possible and welcome). Like several other elements of NAEP, attention here would probably result in models that could be used by other testing systems. Compliance would be monitored by NAEP.

3. As elements 1 and 2 are put into place, compliance by NAEP regarding the proper implementation of the accommodations at the school sites should continue. Because of the increased complexity over what is being done currently, coordinators or other staff would need to be trained properly to monitor and address any problems that arise.

References

- Cook and others (2008). *Addendum to the 2011 Writing Specifications for the National Assessment of Educational Progress*, ACT Inc., Iowa City, IA.
- Cavanagh, S. (2008, July 16). Testing Officials again Tackle Accommodations and Exclusions for Special Student Populations, *Education Week*, vol 27, issue 43, pp 1,16.
- Douglas, K. (2004). Teacher ideas on teaching and testing English language learners: Summary of focus group discussions. Center for the Study of Assessment Validity and Evaluation, University of Wisconsin.
- Fields, R. (2008). Inclusion of special populations in the national assessment: a review of relevant laws and regulations. NAGB Report.
- Fuchs, L.S, Fuchs, D., Eaton, S.B., & Hamlett, C.B. (2005). Dynamic Assessment of Test Accommodations. San Antonio, TX: PsychCorp. Retrieved May 30, 2005 from the World Wide Web: www.unl.edu/buros/
- Fuchs, L.S., Fuchs, D., Eaton, S.B., Hamlett, C.B., Binkley, E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children*, 67(1), 67-81.
- Fuchs, L.S., Fuchs, D., Eaton, S.B., Hamlett, C.B., & Karns, K.M. (2000). Supplementing teacher judgments of mathematics test accommodations with objective data sources. *School Psychology Review*, 29(1) 65-85.
- Goldstein, A. (July 31, 2008). Overview of full population estimates. Presentation to NABG ad hoc committee. NCES PowerPoint.
- Gorman S (July 31, 2008). Uniform criteria for English language proficiency: Possible screener exam. Presentation to NABG ad hoc committee. NCES PowerPoint.
- Helwig, R., & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children*, 69(2), 211-225.
- Hoff, D. (May 19, 1999). Board Won't Revise State NAEP Scores, *Education Week*, p. 1.
- Ketterlin-Geller, L.R. (2003). *Establishing a validity argument for using universally designed assessments*. Unpublished doctoral dissertation, University of Oregon, Eugene.
- Kitmitto, S., and Bandeira de Mello, V. (2008). Measuring the Status and Change of NAEP State Inclusion Rates for Students with Disabilities (NCES 2009-453). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Kopriva, R.J. (2008). *Improving Testing for English Language Learners: A Comprehensive Approach to Designing, Building, Implementing, and Interpreting Better Academic Assessments*. Routledge/Lawrence Earlbaum Publishers, New York, NY.

Kopriva R.J. & Carr, T. (2005). Introducing STELLA. Paper presented at the annual meeting of the Council of Chief State School Officers Large-Scale Assessment, San Antonio, TX.

Kopriva, R.J., Carr, T., & Koran, J. (2006, April). Documenting validity of a matching system for assigning accommodations for individual English language learners. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Kopriva, R.J. & Carr, T (2009). Mapping defensible selection criteria for assigning condition variations to EL students who need them. Paper to be presented at the National Council of Measurement in Education, San Diego, CA.

Kopriva R. J., Emick, J., Hildago-Delgado, C.P., and Cameron, C. (2007). Do proper accommodation assignments make a difference? Examining the impact of improved decision-making on scores for ELLs. *Educational Measurement: Issues and Practices*, pp 21-20.

Koran, J., & Kopriva, R.J., (2006). *Teacher assignment of test accommodations for English language learners*. Presentation at the National Council of Measurement in Education, San Francisco, CA.

Lutkus, A., Mazzeo, J., Zhang, J., Jerry, L. (2004). *Including Special-Needs Students in the NAEP 1998 Reading Assessment: Part II, Results for Students with Disabilities and Limited-English-Proficient Students*, ETS-NAEP report number 04R01. Educational Testing Service, Princeton, NJ.

Mazzeo, J., Carlson, J.E., Voelkl, K.E. and Lutkus, A.D. (2000). *Increasing the Participation of Special Needs Students in NAEP: A Report on 1996 NAEP Research Activities*. U.S. Department of Education. Office of Educational Research and Improvement. National Center for Educational Statistics, 2000-473. Washington, DC.

Mullis, I., Martin, A., Kennedy, Foy, P. (2007). *PIRLS 2006 International Report: IEA's Progress in International Reading Literacy Study in Primary Schools in 40 Countries*. Lynch School of Education, Boston College.

NCES (accessed October, 2008a). <http://www.nces.ed.gov/nationsreportcard/about/inclusion.asp>

NCES (accessed October, 2008b). History of NAEP inclusion policy. <http://www.NCES.ed.gov/>

Plake, B.S. and Impara, C. (2006). Final report for funded proposal to the South Carolina Department of Education, Achieving Accurate Results for Diverse Learners (AARDL). *Beginning the Process: Assigning accommodation for students with disabilities*. South Carolina.

Rivera, C. & Collum, E. (Eds.). (2006). A national review of state assessment policy and practice for English language learners. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Scott, J. (2002, February 7). Foreign Born in U.S. at Record High. *New York Times*, pp.26 (A1, A20).

Sireci, S.G., Li, S. & Scarpati, S. (2003). The effects of test accommodations on test performance: A review of the literature. (Center for Educational Assessment Research Report no. 485) Amherst, Ma: School of Education, University of Massachusetts.

Tindal, G. (2006). The journey through the reliability of a decision-making model for testing students with disabilities. Presentation at the annual meeting of the American Educational Research Association, San Francisco, CA.

Thurlow, M., Lazarus, S., Thompson, S., & Robey, S. (2002). *State participation and accommodation policies for students with disabilities: 2001 update*. Minneapolis, MN: National Center on Educational Outcomes.

Thurlow, M., Moen, R. & Wiley, H.I. (2004). *Biennial performance reports: 2002-2003 state assessment data*. Minneapolis: University of Minnesota, National Center on Educational Outcomes.

XX (December, 2005). *Leveling the Playing Field 2005: Identifying Outperforming and Underperforming States on the NAEP in Demographic Context*. School Matters, www.Schoolmatters.com.

XX, (2008). Steven L. Paine, Letter to chair of the NAGB from Steven L. Paine, February 21, 2007.

U.S. Department of Education. (1999). Peer Reviewer Guidance for Evaluating Evidence of Final Assessments Under Title I of the Elementary and Secondary Education Act. Washington, DC.

U.S. Department of Education, Office of Elementary and Secondary Education (2004). Standards and Assessments Peer Review Guidance: Information and Examples for Meeting Requirements of the No Child Behind Act of 2001. Washington, DC.

ⁱ These recommendations reflect the views of Rebecca Kopriva. They may or may not reflect recommendations by Julia Lara.