



When Science Counts as Much as Reading and Mathematics: An Examination of Differing State Accountability Policies

Eugene Judson

Mary Lou Fulton Teachers College, Arizona State University
United States

Citation: Judson, E. (2012). When Science Counts as Much as Reading and Mathematics: An Examination of Differing State Accountability Policies. *Education Policy Analysis Archives*, 20 (26). Retrieved [date], from <http://epaa.asu.edu/ojs/article/view/987>

Abstract: Although only results from mathematics and reading assessments are required to be used when Adequate Yearly Progress (AYP) of schools is calculated, some states have elected to include science achievement results either in their AYP calculations or as part of a separate dual accountability system. This study examined 2009 National Assessment for Educational Progress (NAEP) results based on how states use, or do not use, science in their accountability programs. Consideration was given to the idea that including science achievement might detract from efforts, and consequently results, in mathematics and reading. Results from both fourth- and eighth-grade data indicated that states choosing to use science in their accountability calculations did not lose ground in those other subjects. Fourth-grade data indicates that the states using science in their accountability programs additionally had significantly higher science achievement than the other states.

Keywords: accountability; assessment; educational policy; science education; large scale assessment

Cuando la ciencia cuenta tanto como lectura y matemáticas: un estudio de las políticas estatales diferenciadas de rendición de cuentas

Resumen: Aunque sólo resultados de las evaluaciones de matemáticas y lectura se requieren para calcular el Progreso Anual Adecuado (por sus siglas en inglés, AYP) de las escuelas, algunos estados han optado por incluir los resultados en el área de ciencias, ya sea en sus cálculos de progreso anual adecuado o como parte de un sistema doble de rendición de cuentas. Este estudio examinó los resultados para 2009 de la Evaluación Nacional de Progreso Educativo (por sus siglas en inglés, NAEP) en función si los estados utilizaron o no los resultados en el área de ciencias en sus programas de rendición de cuentas. Se tuvo en cuenta la idea de que como los esfuerzos en el área de ciencias podrían disminuir los esfuerzos y por tanto los resultados en las matemáticas y la lectura. Los resultados tanto en cuarto y octavo grados indican que los estados que optaron por utilizar la ciencia en sus evaluaciones no perdieron terreno en los otros temas. Además, para el cuarto grado los datos indican que los estados que utilizaron datos del área de ciencias en sus programas de rendición de cuentas tuvieron logros significativamente mayores en ciencias que otros estados.

Palabras clave: rendición de cuentas; evaluación; política educativa; educación científica; evaluación a gran escala

Quando a ciência conta tanto como a leitura e a matemática: um estudo de políticas públicas diferenciadas de prestação de contas

Resumo: Embora apenas os resultados para leitura e matemática sejam necessários para o cálculo de Progresso Anual Adequado (a sigla em Inglês, é AYP) de escolas, alguns estados optaram por incluir os resultados das ciências, em seus cálculos de AYP ou como parte de um sistema dual de prestação de contas. Este estudo analisou os resultados de 2009 da Avaliação Nacional do Progresso Educacional (a sigla em Inglês é NAEP), dependendo se os estados usaram ou não os resultados na área da ciência em seus programas de avaliação. Considerou-se que os esforços em ciência poderiam enfraquecer os esforços e, portanto, os resultados nas áreas de matemática e leitura. Os resultados tanto na quarta quanto na oitava série indicam que os Estados que decidem usar a ciência em suas avaliações não perderam terreno em outras disciplinas. Além disso, para a quarta série, a evidência sugere que os estados que utilizaram dados da área da ciência em seus programas de avaliação tiveram resultados na área de ciência significativamente maior do que outros estados.

Palavras-chave: prestação de contas; avaliação; política educacional; educação científica; avaliação em grande escala

Introduction

In the decade that has followed the passing of the No Child Left Behind (NCLB) Act, educators, policymakers, and researchers have often referred to a resulting narrowed curriculum. Typically, a narrowed curriculum refers to an overconcentration on the content areas of mathematics and reading and a diminished focus on other subjects, such as science and social studies. This is commonly believed to occur because achievement results from mathematics and reading contribute directly to the calculation of Adequate Yearly Progress (AYP) for schools and for local educational agencies (LEAs). When schools or LEAs miss AYP in successive years in the same area, such as mathematics achievement, then the institution is subject to intervention from the state department of education and if AYP is missed in the same area for six consecutive years, invasive actions such as removal of a school board or replacement of staff may occur (Northwest Regional

Educational Laboratory, 2004). Other measures such as the percentage of students tested, high school graduation rates, and attendance rates, also contribute to AYP calculations, but it is student achievement results from mathematics and reading that are the primary reasons schools and LEAs miss AYP (Hoff, 2009).

Criteria for the use of mathematics and reading achievement results were laid out in the NCLB Act. A chief objective for the designers of NCLB was that 100 percent of students would meet or exceed grade level expectations in mathematics and reading by the 2013-14 academic year (No Child Left Behind Act, 2002; Spellings, 2007). Regarding science, the NCLB Act specified that science standards were to be in place by the 2005-06 school year and, by the 2007-2008 school year, states were to have in place science assessments to be administered at least once during grades 3-5; grades 6-9; and grades 10-12 (Gross et al, 2005). Thirty-one states administer science assessments to only the three grades minimally required. Mathematics and reading assessments are however administered to at least seven grades in all fifty states. Significant to this study, although it is a requirement for the states to have science standards in place and to assess science, there is no stipulation within the NCLB Act requiring states to use the results from their science assessments as part of accountability calculations. Despite there being no federal requirement to include science achievement into accountability calculations, eleven states have chosen to do just that.

To take the figure of speech of the narrowing curriculum further, the author suggests the image of an accountability filtration system (Figure 1). On the left side of the accountability membrane it is envisioned there is an extensive curriculum scope where all content areas are represented equitably. However, the accountability membrane allows primarily just the mathematics and reading content to easily pass. The fraction of a content area present to the left or to the right of the membrane would perhaps be proportionate to the amount of effort dedicated to teaching and learning that particular subject. In this simplistic model it is predicted that student learning is balanced on the left side of the membrane where effort is distributed among the subjects and equivalent achievement is occurring. To the right of the membrane it may be believed that greater concentration of the two subjects of mathematics and reading leads to focused effort by teachers and students and subsequently greater achievement in these two most tested subjects. In this sense, the curriculum hasn't been narrowed as much as it has been restricted.

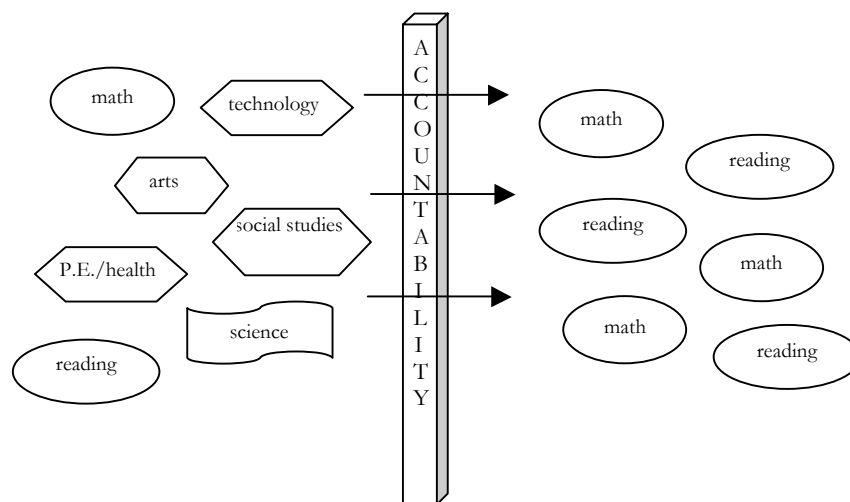


Figure 1. The Accountability Filter

However this filtration system operates differently among the states. In some states science is passing through the accountability membrane in addition to mathematics and reading. That is, a few states include science achievement directly in their accountability calculations alongside mathematics achievement and reading achievement. If a narrowed, or filtered, curriculum does lead to improved results in the subjects that pass through the accountability boundary, then there might be concern that achievement becomes more diffused in those few states where more than just two subjects are allowed to pass through. This might be a particular concern in elementary schools where greater latitude is often allowed regarding how much time and effort is spent on each subject. That is to say, if elementary teachers must be held accountable for three, instead of two subjects, the concern can immediately be that in order to attend to science, there will be a loss in mathematics or reading, or both.

This concern for how science is attended to in classrooms and how that attention may be affected by state accountability policies does not simply result from a sentiment for fairness among the subject areas, but stems from broader national interest. Various national-level policy documents have promoted the importance of supporting effective science education programs as a means to strengthen national pride and invigorate the United States' global competitiveness (e.g., National Governors Association, 2007; National Research Council, 2007). While these documents, and policies such as the America COMPETES Act, endorse an agenda of enriched science curricula and improved science teacher training, they have not impacted the practices of state departments of education calculating AYP.

The intent of this study was to investigate effects of allowing more than the two subject areas of mathematics and reading to be part of accountability calculations. In comparing the few states that integrate science into their accountability calculations with all other states, two research questions were pursued:

1. Does science achievement of students differ between states that use science as part of their accountability calculations and states that do not use science in their accountability calculations?
2. Does mathematics achievement and reading achievement of students differ between states that use science as part of their accountability calculations and states that do not use science in their accountability calculations?

Background

It can be generally agreed that a primary intention of the NCLB legislation was for all children to achieve specific proficiency standards in mathematics and reading. This goal was supported through several key provisions of the NCLB Act such as requiring schools that have not made AYP for two consecutive years to provide opportunities for their students to receive extra academic assistance and promoting teacher quality by ensuring teachers meet state certification requirements (Shaul & Ganson, 2005). Although it is not difficult to find commentaries and studies that contest the value of NCLB (e.g., Amrein-Beardsley, 2009), there does exist lukewarm support for the efficacy of NCLB in the research literature. Using estimates of accountability pressure among the states, Nichols, Glass, and Berliner (2006) found no relationship between accountability pressure with later cohort mathematics achievement on National Assessment of Educational Progress (NAEP) results at the fourth- and eighth-grade levels; however, the researchers did find a causal relationship between high-stakes testing pressure and subsequent achievement on non-cohort fourth-grade mathematics achievement (i.e., comparing achievement to a subsequent year's achievement from the same grade level). Nichols, Glass, and Berliner attributed this effect as likely

due to increased time spent on mathematics instruction. Also, using a rating scale to approximate states' strength of accountability, Carnoy and Loeb (2002) found substantial gains in eighth-grade mathematics scores when states raised external pressure on schools. Finally, in a comprehensive study of the effects of external accountability on student achievement, Lee (2008) examined the 76 effect size estimates from 14 large-scale studies. Results of Lee's meta-analysis showed modest positive policy effects on average, but the study did not address possible shifting of resources or attention from one subject area to another when accountability pressures mount.

Although a scholarly debate may persist for years to come regarding the value of policies that provide sanctions and rewards based largely on results from once-a-year multiple-choice tests, teachers and school administrators do not have the luxury of dismissing the NCLB related requirements that have been translated and set down by their state departments of education. While it is known that achievement of both high and low performing students can be affected by multiple variables, such as teacher quality (Darling-Hammond, 2000) and parental influence (Davis-Kean, 2005), there remains the question of what effect the distal variable of an accountability program has on different levels of student performance. Reback (2008) found that when school personnel believe the goal of attaining proficiency is achievable they are quick to respond to looming interventions and students at the lowest levels of performance make greater than expected gains in mathematics, yet high achieving students do not make similar gains in mathematics. However, this type of differentiated improvement may be due to school personnel attending to achievement in an educational triage manner – with students who are nearly passing or failing high stakes exams receiving greatest attention. Reback concluded that accountability incentives could influence achievement across subjects and across grades:

If a school has a relatively strong incentive to improve students' math performance in a particular grade, then the lowest achieving students in that grade outperform similar schoolmates. The other students in that grade, however, perform worse than similar schoolmates in the other grades, (unless their own performance is relatively important for the school's rating). If a school has a relatively strong incentive to improve some students' reading performance in a particular grade, then other students in this grade perform much worse than similar schoolmates. The findings are again consistent with schools sacrificing general performance in a classroom to focus on the performance of particular students. (p. 1411)

In support of this idea that accountability policies can lead to a shifting of resources, in a case study of one elementary school, Booher-Jennings (2005) also concluded that teachers' response to potential accountability consequences led to a focus on students close to performance thresholds and to diminished attention for other students. Similarly, Diamond and Spillane (2004) analyzed data from observations and interviews and inferred that low performing schools had a more limited focus on improving achievement among a narrow band of students who were at or near performance levels. This narrowed focus was also found within the Chicago Public Schools where Neal and Schanzenbach (2010) used data from standardized tests to learn that students in the third and fourth deciles of prior achievement made greater than expected gains in both mathematics and reading, while students in the first and second deciles remained stagnant.

However, other researchers have found that although failing schools may target students who are on the boundary of meeting performance expectations, this does not occur at the expense of the higher performing students enrolled at the same schools (Springer, 2007). Adding to the mixed results of the NCLB effect, Ballou and Springer (2008) found that improvements experienced at schools do not necessarily come at the cost of affecting high-performing students. In fact, they found pressured schools tend to increase achievement in most grades and not just in the grades

where low achievement had led to the schools missing performance targets. Furthermore, the findings of these researchers cast doubt on the notion that NCLB related gains are largest among schools most likely to face sanctions; in fact, Ballou and Springer found the opposite: that response to NCLB has been greatest among schools that are least threatened by interventions due to low test scores.

While the aforementioned research literature addresses how resources and attention can be transacted between the subjects of reading and mathematics, across grades, and among student groups, less has been studied regarding the effects on other subjects, namely science. There was optimism earlier in the history of NCLB that the requirement of states to assess science in at least three grades would actually lead to an increased focus on the subject (Cavanagh, 2007). However, a review of further reports leads to the conclusion that any early optimism turned to concern about science being relegated to a subject less important than reading or mathematics, especially in the elementary grades. Although it has been reported that elementary teachers describe their beliefs about teaching science to be unchanged by NCLB, and that they possess a generally positive attitude (Milner, Sondergeld, Demir, Johnson, & Czerniak, 2011), there are descriptions of schools diminishing the amount of time spent on science that have been reported repeatedly (Kingsbury, 2007; Linn, 2008; McMurrer, 2007; McMurrer, 2008). As might be expected, school and district personnel have reported cuts in time spent on science, as well as arts, social studies, physical education, and even lunch and recess. These cuts have been attributed to a shared perception for the need to spend more time on mathematics and reading. In fact, analysis of data from elementary schools found that in light of NCLB mandates, science was cut by at least 75 minutes per week in at least half of the reporting districts (McMurrer, 2008).

Yet, as discussed, accountability programs are not identical among the states and a few states do require science achievement to be part of calculations when determining if schools are meeting or not meeting targets necessary to avoid intervention from their state departments of education. In a comparison of fourth-grade and eighth-grade 2005 NAEP science achievement results, states were grouped based on whether they did or did not include science in their accountability calculations (Judson, 2010). Results of this study revealed that there was no appreciable difference between the groups of states when comparing eighth-grade NAEP science results. However, analysis of the fourth-grade NAEP science results revealed there were significant differences in favor of states that used science in their accountability programs. The medium effect size of the difference in fourth-grade results between the groups of states can on the one hand be taken as support for including science into accountability formulas. On the other hand, missing from this study was an examination of what was simultaneously occurring to achievement in mathematics and reading. If the states that use science in their accountability programs are shown to have significantly higher science achievement than other states on a common assessment but are also found to have inferior achievement on mathematics or reading achievement, then the argument can be made that allowing a third subject to pass through the accountability membrane leads to diffused results across the other high stakes content areas. The intent of this study was to then pick up where this earlier study left off. Using the more recent 2009 NAEP science achievement data, comparisons were to be made between states that choose to use and not use science in their accountability programs. Additionally, the analysis here would go further and examine if differences between these groups of states could be detected in the mathematics and reading achievement results of their fourth- and eighth-grade students.

Methods

Categorizing the States

The states were grouped into three categories based on how they use or do not use science achievement in their accountability program calculations. Although all of the states assess science achievement in at least three grades, only a few have mandated that the results from their state science assessments contribute to the determinations of whether schools and LEAs are meeting accountability benchmarks. Choosing to use science as part of accountability programs may be done in one of two ways. One way is for a state to directly integrate science achievement results into the federally required AYP calculations. Although the states are required to use high school graduation rate as a variable when determining if secondary schools are meeting expectations, the states are allowed to decide what variable they will use as an additional indicator when calculating AYP for the elementary grades. The large majority of states have selected attendance as their additional indicator for the elementary grades. However, a few states, such as New York, use science achievement as their additional indicator when calculating AYP.

The second way that science achievement can contribute to accountability is when it is part of a second accountability system that is parallel to the NCLB required AYP-based accountability system. Because most states did have some form of accountability and reporting in place prior to the commencement of NCLB legislation, several of those states have chosen to continue with, and adapt, their previous accountability systems. At present there are fifteen states that have dual accountability programs (i.e., the AYP accountability program plus a state accountability program) in place (Blank & Hovanetz, 2009). Among these fifteen states, a few require that, as part of their state accountability program, science achievement be included in the calculations. For example, within Utah's U-PASS accountability system, science achievement contributes 20 percent of a school's proficiency rating. For this study, to be labeled as a state that "uses" science in their dual accountability program, it was determined that this parallel accountability system needed to carry potential penalties when targets are not met, just as the AYP-based accountability program does. In other words, if the dual accountability program simply assessed and reported schools' status, but did not have the weight of possible intervention when schools failed, then that was viewed equivalent to the predominant AYP-based system in which states report science results but those results do not contribute to computing whether or not a school or LEA is subject to sanctions.

The states were categorized into three groups. Group 1 was comprised of the thirty-nine states that do not use science achievement in their accountability calculations. The definitions for the second and third groups of states were based on (a) the degree to which science is required as an accountability variable, and (b) the grade levels assessed by the states. Each of these criteria is clarified here. Among the eleven states that integrate science achievement results into accountability calculations, some of those states allow schools to select science from a menu of choices. This is the case in Georgia where science may be chosen as the additional indicator for AYP by the elementary schools, but most often the schools select to use attendance rate. If a state allows for science to be used in accountability, but does not require it, that state was placed in Group 2.

Because the NAEP science assessment achievement results were used in this study as the dependent variable when comparing states, the definitions for the second and third groups were also influenced by the grades tested by NAEP. That is, while some states do require that science achievement be included in accountability, the grades from which they require those results do

not always match the grades assessed by NAEP. NAEP tests science in only fourth- and eighth-grade. If a state required science to be included in accountability but did not require results from fourth- or eighth-grade, the state was placed in Group 2. There were cases such as that of North Carolina that requires fifth- and eighth-grade science results to be used in their accountability programs. In this instance, North Carolina matches only one NAEP grade. Therefore, North Carolina was placed in Group 2 for fourth-grade categorization and in Group 3 for eighth-grade categorization. Group 2 are the states that have either partially committed to using science in accountability by making it a choice or do require the use of science achievement in accountability, but do not require those results from a NAEP tested grade (i.e., fourth- or eighth-grade). The rationale for creating the category of Group 2, as opposed to aggregating these states with the states that required the use of science achievement results and did match the NAEP grades, was to maintain a clear focus on possible effects of a stricter definition of using science achievement in accountability. The Group 2 states were not coupled with the Group 1 states because there was interest to see if perhaps a spillover effect was in play in states that had made some movement in the direction of using science achievement in accountability. Group 3 are those states that require science to be used in accountability calculations and require achievement results from a grade matching a NAEP tested grade. The groupings of the states are provided in Table 1.

Table 1

Categories of States Based on Use of Science Achievement in Accountability Calculations

| | 4 th Grade | 8 th Grade |
|----------------------------------------------------------------------------------------------------------------------|------------------------|------------------------------------|
| Group 1 – Do not include science in accountability. | All other states | All other states |
| Group 2 – Include science in accountability, but either as a choice or do not use results from the NAEP tested grade | CA, GA, MI, NC, OH, VA | GA, KY |
| Group 3 – Require science in accountability from the NAEP tested grade | KY, NY, SC, TN, UT | CA, MI, NY, NC, OH, SC, TN, UT, VA |

Use of NAEP Data

As mentioned, NAEP data were utilized to address the research questions. The first research question is an inquiry of whether the process of including science into accountability calculations affects student achievement. NAEP science achievement results from 2000 and 2009 were available and considered well suited to address the first research question. In reauthorizing the Elementary and Secondary Education Act (ESEA), NCLB legislation was passed in 2001 and began to take effect in 2002 (No Child Left Behind Act, 2002). The NAEP

data from 2000 then provided a glimpse into pre-accountability austerity. Because the framework for the NAEP science achievement assessment was different in 2009 than it was in 2000 (National Assessment Governing Board, 2008), scale scores from the two years could not be strictly compared. Although the science framework had changed, certainly both the 2000 and 2009 assessments were built on the construct of science and the 2000 achievement results could be used as covariate variables when comparing 2009 fourth- and eighth-grade results across Groups 1, 2, and 3.

A point of clarification must be made regarding the use of NAEP data for this study that investigated how differing AYP practices among the states might yield dissimilar results on NAEP assessments. It is important to note that NAEP assessment results are not integrated into any AYP calculations and are not used as proxies for states' high stakes accountability tests. Studies have demonstrated that the percentage of students scoring proficient on state examinations can be extremely different from the percent of students scoring at Proficient level on NAEP (Stoneberg, 2007) and therefore using NAEP data when examining AYP practices may warrant caution. For example, although overall NAEP results have remained relatively stable over time, student proficiency on state assessments has increased in some states (Jacob, 2007). A cause of this discrepancy is likely due to the evidence revealed in a state level examination of proficiency standards conducted by the National Center for Education Statistics (NCES) indicating that for approximately half of the cases, examined in grades 4 and 8, the rigor of states' standards had decreased between 2005 and 2009 (Bandeira de Mello, 2011). However, this lack of corroboration between NAEP results and the proportion of students meeting states' proficiency standards was not viewed as a deterrent to this study. Within this study NAEP was not used as confirmatory evidence of states' assessment results; rather, NAEP was utilized only as a basis of comparison of achievement in the content areas of science, mathematics, and reading, as NAEP "offers the most reliable and equitable measures of student achievement across states" (Nicholas, 2005, p. 2). That is, it was not an objective to determine if states' methods of determining proportions of proficient students on their state assessments correlated with NAEP. Instead, the intent was to determine if the variable of including science results from state assessments in state-level accountability practices influenced achievement, as detected by NAEP.

To be consistent with the selection of the science NAEP analysis, mathematics and reading NAEP data were selected from 2009 and a pre-NCLB year. Similar to science, mathematics had been assessed in fourth- and eighth-grade in both 2000 and 2009. Reading had been assessed in both fourth- and eighth-grade in 1998 and again in 2009. The use of these data allowed the second research question to be addressed. If the inclusion of science into accountability programs reduced attention to the core subjects of mathematics and reading, then there was anticipation of a negative impact on mathematics and reading achievement in the Group 3 states.

Data Analysis

A series of analysis of covariance (ANCOVA) were conducted for the three subject areas of science, mathematics, and reading using the pre-NCLB NAEP mean scale scores as covariates and the 2009 NAEP mean scale scores as the dependent variables. The ANCOVA analyses were conducted for each of the three subjects of reading, mathematics, and science and for both fourth- and eighth-grade data. The use of ANCOVA analysis allowed the 2009 data to be adjusted on the pre-NCLB covariate data; this procedure yields adjusted means, also referred to as estimated marginal means, on the 2009 scale scores "as if" all groups had equivalent scale

scores in the pre-NCLB year. Significant omnibus F-test results were followed up with Fisher least significant difference (LSD) post-hoc comparisons.

NAEP achievement data from 1998 were used as the covariate when analyzing differences among the groups in reading and NAEP achievement data from 2000 were used as the covariates when analyzing differences among the groups in mathematics and science. To be included in the analyzed dataset, states needed to have reported NAEP data from the pre-NCLB year and from 2009. This requirement eliminated some states from the study. Fourth-grade data were available from 39 states for reading, 40 states for mathematics, and 37 states for science. Eighth-grade data were available from 36 states for reading, 39 states for mathematics, and 36 states for science.

Data were analyzed for all students and were also disaggregated and analyzed based on socioeconomic status (SES). To varying degrees, SES status has been shown to be related to academic achievement in multiple subject areas (e.g., McGraw, Lubienski, & Strutchens, 2006; Perry & McConney, 2010; Stipek & Ryan, 1997, Willms, 2003). Also of interest was determining if effects, attributable to the use of science in state accountability programs, could be detected among student groups based on ethnicity. Researchers have examined disparity in achievement among ethnic groups on standardized tests that is often related to SES differences among the groups (Flores, 2007; Magnuson, Rosenbaum, & Waldfogel, 2008). However, the NAEP data were inadequate to examine groups based on ethnicity. Disaggregation of NAEP science data, based on ethnicity, provided too few states meeting criteria of the National Assessment Governing Board (NAGB), that oversees NAEP administration and reporting, to be useable across all ethnic groups. For example, only a total of four states among the eighth-grade data met the NAGB criteria for both the pre-NCLB and post-NCLB data years for the groups of Asian and American Indian students. A reduced amount of states reported other ethnic groups (e.g., Hispanic), but not to the extreme of the example of Asian and American Indian student reporting. Because of the requirement that data needed to be available from both the pre-NCLB year and from 2009 had already vetted the states to be analyzed by approximately 30 percent, analysis remained focused on the broad category of all students and the SES-based categories that were generally well reported by the states.

Results

Results of ANCOVA analysis from the aggregate category of all students and SES-based categories (i.e., eligible and not eligible for free or reduced lunch) are provided for fourth-grade in Table 2.

Separate ANCOVA analysis of all students' NAEP scale scores for the three fourth-grade subject areas revealed that after controlling for the pre-NCLB scale scores (covariates) there were no significant differences among the groups in reading or in mathematics. There were, though, significant differences in the category of all students among the groups of states in their fourth-grade science results, $F(2, 34) = 4.831, p < .05$. A significant difference among the groups of states persisted when evaluating the fourth-grade data in the categories of students eligible for free or reduced lunch, $F(2, 34) = 3.639, p < .05$, and students not eligible for free or reduced lunch, $F(2, 34) = 4.286, p < .05$.

Table 2
 Fourth-grade NAEP Pre- and Post-NCLB, All Students and SES-based Groups

| Subject | Students | Group | n | Pre-NCLB | | Post-NCLB | | EMM | F | p |
|---------|------------------------------|-------|----|----------|------|-----------|-----|-------|-------|-------|
| | | | | M | SD | M | SD | | | |
| Reading | All | 1 | 29 | 214.7 | 8.2 | 220.0 | 6.9 | 219.7 | .090 | .914 |
| | | 2 | 5 | 211.4 | 6.0 | 218.3 | 5.9 | 220.1 | | |
| | | 3 | 5 | 214.0 | 3.6 | 220.4 | 4.4 | 220.5 | | |
| | Eligible free/reduced | 1 | 29 | 199.0 | 8.0 | 207.1 | 5.4 | 206.9 | .347 | .709 |
| | | 2 | 5 | 194.3 | 7.3 | 204.3 | 5.3 | 205.8 | | |
| | | 3 | 5 | 199.8 | 5.4 | 208.8 | 5.3 | 208.3 | | |
| | Not eligible free/reduced | 1 | 29 | 225.3 | 5.7 | 231.0 | 5.3 | 230.8 | .152 | .859 |
| | | 2 | 5 | 223.2 | 3.1 | 230.8 | 3.2 | 231.9 | | |
| | | 3 | 5 | 225.5 | 3.9 | 231.2 | 4.3 | 230.9 | | |
| Math | All | 1 | 29 | 224.9 | 6.6 | 239.4 | 6.5 | 239.2 | .066 | .936 |
| | | 2 | 6 | 225.1 | 7.4 | 239.1 | 5.1 | 238.7 | | |
| | | 3 | 5 | 222.2 | 3.5 | 237.5 | 3.7 | 239.4 | | |
| | Eligible free/reduced | 1 | 29 | 212.5 | 6.0 | 228.9 | 5.6 | 228.5 | .403 | .671 |
| | | 2 | 6 | 210.7 | 6.8 | 226.5 | 4.9 | 227.2 | | |
| | | 3 | 5 | 209.1 | 4.1 | 227.2 | 3.9 | 229.0 | | |
| | Not eligible free/reduced | 1 | 29 | 234.1 | 4.5 | 248.6 | 4.8 | 248.7 | .299 | .743 |
| | | 2 | 6 | 235.2 | 4.4 | 250.0 | 3.4 | 249.2 | | |
| | | 3 | 5 | 233.3 | 3.0 | 247.2 | 3.0 | 247.8 | | |
| Science | All | 1 | 26 | 149.4 | 8.3 | 149.9 | 7.8 | 149.3 | 4.831 | .014* |
| | | 2 | 6 | 146.6 | 10.1 | 149.6 | 9.1 | 151.5 | | |
| | | 3 | 5 | 147.8 | 5.5 | 152.0 | 5.5 | 152.9 | | |
| | Eligible free/reduced | 1 | 26 | 135.9 | 9.1 | 136.7 | 7.6 | 135.7 | 3.639 | .037* |
| | | 2 | 6 | 128.9 | 10.0 | 134.0 | 7.7 | 138.2 | | |
| | | 3 | 5 | 134.4 | 5.9 | 139.2 | 6.3 | 139.4 | | |
| | Not eligible free/reduced | 1 | 26 | 160.0 | 5.2 | 161.7 | 5.6 | 161.5 | 4.286 | .022* |
| | | 2 | 6 | 159.3 | 5.1 | 163.0 | 6.2 | 163.6 | | |
| | | 3 | 5 | 159.2 | 2.8 | 164.6 | 4.5 | 165.2 | | |

EMM: Estimated Marginal Mean

* $p < .05$

Fisher LSD comparisons were conducted to determine differences between the groups in fourth-grade science achievement. The post hoc tests revealed significant differences in fourth-grade between Group 1 (i.e., states not including science in accountability) and Group 3 (i.e., states that require science in accountability), $p < .05$, for the categories of all students, students eligible for free or reduced lunch, and students not eligible for free or reduced lunch. In all three cases, Group 3 had significantly higher mean scale science scores. Fourth-grade science achievement was significantly higher in states where science was required to be part of an accountability program. There were no significant differences in the fourth-grade data between Group 2 and either Group 1 or Group 3 for the all students and SES-based categories.

Partial eta-squared (η_p^2) measurements were used to determine effect sizes in which small, medium, and large effects were operationalized as .01, .06, and .14, respectively (Stevens, 1992). The effect of using science in accountability programs on the 2009 mean NAEP science scale scores was considered large for the fourth grade categories of all students ($\eta_p^2 = .226$), students eligible for free or reduced lunch ($\eta_p^2 = .181$), and students not eligible for free or reduced lunch ($\eta_p^2 = .206$).

The eighth-grade data were similarly analyzed. Table 3 provides results of the ANCOVA analyses of the three groups of states across the subjects of reading, mathematics, and science.

Table 3

Eighth-grade NAEP Pre- and Post-NCLB, All Students and SES-based Groups

| Subject | Students | Group | n | Pre-NCLB | | Post-NCLB | | EMM | F | p |
|---------|---------------------------|-------|----|----------|------|-----------|------|-------|-------|-------|
| | | | | M | SD | M | SD | | | |
| Reading | All | 1 | 27 | 261.4 | 6.2 | 262.5 | 6.5 | 262.3 | .468 | .625 |
| | | 2 | 2 | 259.8 | 3.6 | 263.5 | 4.7 | 264.6 | | |
| | | 3 | 7 | 260.3 | 5.3 | 260.8 | 4.8 | 261.5 | | |
| | Eligible free/reduced | 1 | 27 | 247.4 | 6.0 | 250.2 | 4.9 | 249.8 | .971 | .389 |
| | | 2 | 2 | 245.4 | 7.5 | 253.0 | 5.4 | 253.5 | | |
| | | 3 | 7 | 244.0 | 5.5 | 247.9 | 3.9 | 249.1 | | |
| | Not eligible free/reduced | 1 | 27 | 268.4 | 5.1 | 271.2 | 5.1 | 271.4 | 1.094 | .347 |
| | | 2 | 2 | 269.1 | 1.7 | 273.8 | 3.1 | 273.5 | | |
| | | 3 | 7 | 269.5 | 3.5 | 270.5 | 2.8 | 270.0 | | |
| Math | All | 1 | 28 | 272.8 | 9.2 | 282.4 | 8.9 | 281.9 | .029 | .972 |
| | | 2 | 2 | 267.6 | 3.2 | 278.4 | 1.2 | 282.1 | | |
| | | 3 | 9 | 271.1 | 7.4 | 280.7 | 5.3 | 281.6 | | |
| | Eligible free/reduced | 1 | 28 | 256.4 | 9.5 | 268.2 | 7.0 | 267.4 | .146 | .865 |
| | | 2 | 2 | 250.4 | 6.3 | 266.2 | 2.3 | 268.9 | | |
| | | 3 | 9 | 251.8 | 6.5 | 265.5 | 4.4 | 267.4 | | |
| | Not eligible free/reduced | 1 | 28 | 281.0 | 7.3 | 292.3 | 7.2 | 292.1 | .054 | .948 |
| | | 2 | 2 | 279.1 | 1.9 | 290.0 | 0.5 | 291.3 | | |
| | | 3 | 9 | 280.2 | 5.9 | 291.4 | 4.4 | 291.8 | | |
| Science | All | 1 | 25 | 148.6 | 8.9 | 149.5 | 8.5 | 149.0 | 2.711 | .082 |
| | | 2 | 2 | 145.8 | 5.4 | 151.6 | 6.7 | 153.6 | | |
| | | 3 | 9 | 147.0 | 9.3 | 149.4 | 7.4 | 150.3 | | |
| | Eligible free/reduced | 1 | 25 | 132.3 | 10.3 | 135.6 | 8.5 | 134.8 | 4.269 | .023* |
| | | 2 | 2 | 129.1 | 8.8 | 140.1 | 10.3 | 141.7 | | |
| | | 3 | 9 | 128.6 | 9.2 | 134.2 | 7.1 | 136.2 | | |
| | Not eligible free/reduced | 1 | 25 | 157.2 | 6.5 | 159.5 | 6.1 | 159.3 | 3.556 | .040* |
| | | 2 | 2 | 156.5 | 3.5 | 162.9 | 2.9 | 163.1 | | |
| | | 3 | 9 | 155.8 | 7.5 | 160.6 | 4.2 | 161.3 | | |

EMM: Estimated Marginal Mean

* $p < .05$

Similar to the fourth-grade analysis, the only significant differences among the three groups were found in the subject of science. There were no significant differences among

groups in either eighth-grade reading or eighth-grade mathematics data when comparing 2009 mean scale scores, with the pre-NCLB scores used as the covariates. Unlike the fourth-grade results, the eighth-grade results in science did not align neatly with the hypothesis that inclusion of science into accountability programs will lead to greater achievement. The omnibus F tests did not reveal a significant effect in the category of all students. However, there was a significant difference among the groups in the categories of students eligible for free or reduced lunch, $F(2, 33) = 4.269, p < .05$, and students not eligible for free or reduced lunch, $F(2, 33) = 3.556, p < .05$. Within the category of students eligible for free or reduced lunch, the significant differences were between Group 1 and Group 2 ($p = .008$), and between Group 2 and Group 3 ($p = .043$). Group 2 are those states that include science in some manner in their accountability programs, but either offer this as an option for their schools or do not use results from the NAEP grade being assessed – in this case, eighth-grade. Here, adjusting for the covariate 2000 NAEP data, it was the two states of Georgia and Kentucky, comprising Group 2, with significantly higher mean scale scores on the 2009 NAEP science assessment than the two other groups of states. Regarding the category of students not eligible for free or reduced lunch, although the overall model demonstrated significance ($p = .040$), the post-hoc tests did not reveal significant differences between groups of states at the criteria level, $p < .05$.

Discussion

In the presentation of the imagined accountability filtration system (Figure 1) the supposition was posed that allowing the third subject of science to pass through the membrane and be included in high-stakes accountability would lead to diffuse attention across mathematics and reading, and consequently lead to relatively lower achievement in those subjects. The data presented here does not support this supposition. In both fourth- and eighth-grade, 2009 NAEP mathematics and reading achievement scores were equivalent among the groups of states. What was different across the three groups of states was their 2009 NAEP science achievement.

The fourth-grade data are supportive of a hypothesis indicating that the inclusion of students' science achievement results into accountability calculations will promote higher achievement. Of course, these state level data provide only a mile high view of learning and do not indicate what fourth-grade practices and informal policies may be different among the groups of states. However, the results from fourth-grade data are consistent with the previous study of 2005 NAEP data (Judson, 2010). That earlier study can be viewed as an intermediate data inspection wherein 2000 and 2005 NAEP science data were compared and, as is the case here, the hypothesis that inclusion of science results into accountability formulas will promote science achievement was supported only at the fourth-grade. Yet this study additionally demonstrates that states venturing to include science are not losing a step in reading or mathematics. This latter finding may at first seem to run counter to commonsensical logic that in order to attend well to an additional subject, some resources must be drawn from mathematics and/or reading. It is offered that this line of thought might be too limited and that schools in the Group 3 states are not simply robbing Peter in order to pay Paul – or stealing resources from mathematics and reading, in order to attend to science, as is the case here. Instead, further investigation must consider if fourth-grade classrooms in the Group 3 states are incorporating science through an overall enriched curriculum. Integration of science with mathematics and with literacy has been shown to have benefits across all of these subjects, so the question is now raised if any form of interdisciplinary curriculum is more prevalent among the Group 3 states in their fourth-grade classrooms. Of course, what must also be considered is that a shifting of resources is occurring in the Group 3 states' fourth-grade classrooms, but not at the cost

of reading or mathematics, but at the cost of the subjects still left on the left side of the accountability membrane, such as physical education and art. Prior research has been presented indicating that non-high stakes subjects may be receiving less attention, so while these fourth-grade data can be hailed as support for including science into accountability calculations, further analysis is needed to determine if other subjects have been even more disregarded.

Less straightforward to scrutinize are the eighth-grade results. The eighth-grade data were similar to the fourth-grade data in that the 2009 NAEP mathematics and reading achievement were equivalent among the groups of states. However, unlike the fourth-grade data, the eighth-grade data did not reveal Group 3 states to have greater relative achievement on the 2009 NAEP science assessment. While it might be thinly argued that some positive effect was found in the Group 2 states, namely Kentucky and Georgia, because plausibly there is some spillover effect into eighth-grade science results when these states allow schools to choose to include science into their accountability formulas, it is doubtful that such an effect would be detected only in the Group 2 states. More likely what is occurring is a lack of a science accountability effect registering in eighth-grade. It is believed that the effect found in the fourth-grade data, but not the eighth-grade data, is likely due to the nature of science characteristically taught in these grades. Across the United States, far more often than not, eighth-grade students have multiple teachers and their schedules provide allotted instructional time for their courses. Although anecdotes exist of eighth-grade science teachers being instructed by school administrators to stop teaching science days before a state's high stakes test so as to help drill students on mathematics skills, generally in eighth-grade there is blocked and defined time for science. This is not necessarily the case in fourth-grade. In a typical elementary school, where fourth-grade teachers must teach multiple subjects, the terrain of the curriculum can be more flexible. This may mean that more resources are devoted to providing professional development for teachers to improve instructional practices in the high-stakes subjects. It may also simply mean that more time is devoted to those high-stakes subjects. Further investigation of the amount of time spent on science in the three groups of states would further this line of reasoning.

Regarding implications for policymakers, this study may have weight when members of Congress consider reauthorization of the Elementary and Secondary Education Act (ESEA) or when state decision-makers determine revisions of their state accountability policies. There have been past attempts to require science results be used in AYP calculations. The Science Accountability Act (2006, 2007, 2009) has been introduced in Congress three times in attempts to include science in AYP calculations, but each time the bill has failed to make it out of committee. The National Science Teachers Association (NSTA) organized 61 organizations to back the Make Science Count petition (2007) to Congress and in 2011 the K-12 STEM Education Policy Conference included among its key talking points with Congressional members the imperative to include science on par with mathematics and reading when ESEA is reauthorized. There have been other recent recommendations to include science into accountability, such as that from the National Research Council (NRC) (2011), in which it was recommended that "policy makers at the national, state, and local levels should elevate science to the same level of importance as reading and mathematics."

Science education is in the midst of its next wave of change. The recently released Framework for K-12 Science Education (National Research Council, 2012) provides the guidelines from which will emerge the Next Generation Science Standards, pending to be released in 2012 (Robelen, 2011). This new conceptual framework lays down a valuable foundation for teaching and learning science that will steer science standards and consequently classroom curriculum. In all likelihood, a majority of the states will adopt the new science standards and revisit their science

assessments. Yet, the impact on accountability policy is unknown. Sensibly, this would be an occasion for states to simultaneously re-examine their accountability policies; the time is ripe for policymakers to deliberate on research findings as they make their decisions. Hopefully this study is among those included in the register of informative studies used by policymakers.

References

- Amrein-Beardsley, A. (2009). The unintended, pernicious consequences of “staying the course” on the United States’ no child left behind policy. *International Journal of Education Policy & Leadership*, 4(6), 1-13.
- Ballou, D., & Springer, M. G. (2008, April). *Achievement trade-offs and no child left behind*. Paper presented at the annual meeting of the American Finance Association, Denver, CO. Retrieved March 19, 2012 from http://www.vanderbilt.edu/schoolchoice/documents/achievement_tradeoffs.pdf
- Bandeira de Mello, V. (2011), *Mapping state proficiency standards onto NAEP scales: Variation and change in state standards for reading and mathematics 2005-2009* (NCES 2011-458). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC: Government Printing Office.
- Blank, R. K., & Hovanetz, C. (2009, April). *State accountability systems under NCLB at year 6: Do they report useful information? What is the role of state policy?* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Booher-Jennings, J. (2005). Below the bubble: “Educational triage” and the Texas accountability system. *American Educational Research Journal*, 42(2), 231–268.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305-331.
- Cavanagh, S. (2007, October 13). Federal rule yields hope for science: Testing mandate is expected to increase time for subject. *Education Week*, 27(7), pp. 1, 13.
- Darling-Hammond, L. (2000). How teacher education matters. *Journal of Teacher Education*, 51(3), 166-173.
- Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, 19(2), 294-304.
- Diamond, J. B., & Spillane, J. (2004). High-stake accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record*, 106(6), 1145–1176.
- Flores, A. (2007). Examining disparities in mathematics education: Achievement gap or opportunity gap? *The High School Journal*, 91(1), 29-42.
- Gross, P.R., Goodenough, U., Haack, S., Lerner, L. S., Schwartz, M., Schwartz, R. (2005). *The state of state science standards 2005*. Washington (DC): The Thomas B Fordham Institute.
- Hoff, D. J. (2009, January 7). Schools struggling to meet key goal on accountability. *Education Week*, 28(16), pp. 1, 14-15.
- Jacob, B. (2007). Test-based accountability and student achievement: An investigation of differential performance on NAEP and state assessments. NBER Working Paper 12817. Retrieved March 19, 2012 from <http://closup.umich.edu/files/closup-wp-17-test-based-accountability.pdf>
- Judson, E. (2010). Science education as a contributor to adequate yearly progress and accountability programs. *Science Education*, 94(5), 888-902.

- Kingsbury, A. (2007, July 25). Schools cut other subjects to teach reading and math. U.S. News & World Report. Retrieved March 19, 2012 from <http://www.usnews.com/usnews/edu/articles/070725/25nclb.htm>
- Lee, J. (2008). Is test-driven external accountability effective? Synthesizing the evidence from cross-state causal-comparative and correlational studies. *Review of Educational Research*, 78(3), 608-644.
- Linn, R. L. (2008). Toward a more effective definition of adequate yearly progress. In G. L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, and school reform* (pp. 27-42). Thousand Oaks, CA: Corwin Press.
- Magnuson, K., Rosenbaum, D. T., & Waldfogel, J. (2008). Inequality and black-white achievement trends in the NAEP. In K. Magnuson and J. Waldfogel (Eds.), *Steady Gains and Stalled Progress: Inequality and the Black-White Test Score Gap* (pp. 33-65). New York, NY: Russell Sage Foundation.
- McGraw, R., Lubienski, S. T., & Strutchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race/ethnicity, and socioeconomic status. *Journal for Research in Mathematics Education*, 37(2), 129-150.
- McMurrer, J. (2007). *Choices, changes, and challenges: Curriculum and instruction in the NCLB era*. Washington, DC: Center on Education Policy. Retrieved March 19, 2012 from <http://www.cep-dc.org/displayDocument.cfm?DocumentID=312>
- McMurrer, J. (2008). *Instructional time in elementary schools: A closer look at changes in specific subjects*. Washington, DC: Center on Education Policy. Retrieved March 19, 2012 from <http://www.cep-dc.org/displayDocument.cfm?DocumentID=309>
- Milner, A. R., Sondergeld, T. A., Demir, A., Johnson, C. C., Czerniak, C. M. (2011). Elementary teachers' beliefs about teaching science and classroom practice: An examination of pre/post NCLB testing in science. *Journal of Science Teacher Education*. Retrieved March 19, 2012 from <http://www.springerlink.com/content/5954494r44ll6825/>
- National Assessment Governing Board. (2008). Science framework for the 2009 national assessment of educational progress. Retrieved March 19, 2012 from <http://www.nagb.org/publications/frameworks/science-09.pdf>
- National Governors Association. (2007). *Innovation America: A final report*. National Governors Association, Washington DC.
- National Research Council. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. National Academies Press, Washington DC.
- National Research Council. (2011). *Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics*. Committee on Highly Successful Science Programs for K-12 Science Education. Board on Science Education and Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- National Science Teachers Association (2007, October). *Make science count – Include science in the No Child Left Behind adequate yearly progress*. Retrieved March 19, 2012 from http://science.nsta.org/nstaexpress/nstaexpress_2007_10_22_makesciencecount.htm
- Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics*, 92(2), 263-283.
- Nicholas, J. (June, 2005). Use of the 2003 National Assessment of Educational Progress results as an indicator of state adequate yearly progress. Paper presented at the CCSSO National Conference on Large Scale Assessment, San Antonio, TX.

- Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, 14(1). Retrieved March 19, 2012 from <http://epaa.asu.edu/epaa/v14n1>
- No Child Left Behind Act of 2001. Public Law 107-110, 115 Stat. 1425 (2002). Retrieved March 19, 2012 from <http://www.ed.gov/policy/elsec/leg/esea02/index.html>
- Northwest Regional Educational Laboratory. (2004). *Making sense of adequate yearly progress*. Portland, OR: Author. Retrieved March 19, 2012 from <http://nwrcc.educationnorthwest.org/filesnwrcc/webfm/docs/AYP.pdf>
- Perry, L. B., & McConney, A. (2010). Does the SES of the school matter? An examination of socioeconomic status and student achievement using PISA 2003. *Teachers College Record*, 112(4), 1137-1162.
- Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, 92(5-6), 1394-1415.
- Robelen, E. W. (2011, July 19). New science framework paves way for standards. *Education Week*, 30(37). Retrieved March 9, 2012, from <http://www.edweek.org/ew/articles/2011/07/19/37science.h30.html?tkn=NXLF84%2BGWb96XT8cVTiWy9AqtICk76XPN0FL&cmp=clp-edweek>
- Science Accountability Act of 2006, H.R. 5442, 109th Cong. (2006).
- Science Accountability Act of 2007, H.R. 35, 110th Cong. (2007).
- Science Accountability Act of 2009, H.R. 2511, 111th Cong. (2009).
- Shaul, M. S., & Ganson, H. C. (2005). The No Child Left Behind Act of 2001: The federal government's role in strengthening accountability for student performance. *Review of Research in Education*, 29, 151-154.
- Spellings, Margaret (2007). *Building on results: A blueprint for strengthening the No Child Left Behind Act*. Washington, DC.: U.S. Department of Education.
- Springer, M.G. (2007). Accountability incentives: Do failing schools practice educational triage? *Education Next*, 8(1), 74-79.
- Stevens, J. C. (1992). *Applied multivariate statistics for the social sciences*. Hillside, NJ: Lawrence Erlbaum.
- Stipek, D. J., & Ryan, R. H. (1997). Economically disadvantaged preschoolers: Ready to learn but further to go. *Developmental Psychology*, 33(4), 711-723.
- Stoneberg, B. D. (2007). Using NAEP to confirm state test results in the No Child Left Behind Act. *Practical Assessment, Research & Evaluation*, 12(5). Retrieved March 19, 2012 from <http://pareonline.net/pdf/v12n5.pdf>
- Willms, J. D. (2003). Literacy proficiency of youth: Evidence of converging socioeconomic gradients. *International Journal of Educational Research*, 39(3), 247-252.

About the Author

Eugene Judson
Mary Lou Fulton Teachers College, Arizona State University

Email: Eugene.Judson@asu.edu

Eugene Judson is an assistant professor of science education. His research focuses on the crossroads of educational policy and science, technology, engineering, and mathematics (STEM) teaching and learning.

education policy analysis archives

Volume 20 Number 26 September 3rd, 2012

ISSN 1068-2341



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-sa/3.0/>. All other uses must be approved by the author(s) or **EPAA**. **EPAA** is published by the Mary Lou Fulton Institute and Graduate School of Education at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, H.W. WILSON & Co., QUALIS A2 (Brazil), SCImago Journal Rank; SCOPUS, SOCOLAR (China).

Please contribute commentaries at <http://epaa.info/wordpress/> and send errata notes to Gustavo E. Fischman fischman@asu.edu

education policy analysis archives
editorial board

Editor **Gustavo E. Fischman** (Arizona State University)

Associate Editors: **David R. Garcia** (Arizona State University), **Rick Mintrop**, (University of California, Berkeley)

Jeanne M. Powers (Arizona State University)

Jessica Allen University of Colorado, Boulder

Gary Anderson New York University

Michael W. Apple University of Wisconsin, Madison

Angela Arzubiaga Arizona State University

David C. Berliner Arizona State University

Robert Bickel Marshall University

Henry Braun Boston College

Eric Camburn University of Wisconsin, Madison

Wendy C. Chi* University of Colorado, Boulder

Casey Cobb University of Connecticut

Arnold Danzig Arizona State University

Antonia Darder University of Illinois, Urbana-Champaign

Linda Darling-Hammond Stanford University

Chad d'Entremont Strategies for Children

John Diamond Harvard University

Tara Donahue Learning Point Associates

Sherman Dorn University of South Florida

Christopher Joseph Frey Bowling Green State University

Melissa Lynn Freeman* Adams State College

Amy Garrett Dikkers University of Minnesota

Gene V Glass Arizona State University

Ronald Glass University of California, Santa Cruz

Harvey Goldstein Bristol University

Jacob P. K. Gross Indiana University

Eric M. Haas WestEd

Kimberly Joy Howard* University of Southern California

Aimee Howley Ohio University

Craig Howley Ohio University

Steve Klees University of Maryland

Jackyung Lee SUNY Buffalo

Christopher Lubienski University of Illinois, Urbana-Champaign

Sarah Lubienski University of Illinois, Urbana-Champaign

Samuel R. Lucas University of California, Berkeley

Maria Martinez-Coslo University of Texas, Arlington

William Mathis University of Colorado, Boulder

Tristan McCowan Institute of Education, London

Heinrich Mintrop University of California, Berkeley

Michele S. Moses University of Colorado, Boulder

Julianne Moss University of Melbourne

Sharon Nichols University of Texas, San Antonio

Noga O'Connor University of Iowa

João Paraskveva University of Massachusetts, Dartmouth

Laurence Parker University of Illinois, Urbana-Champaign

Susan L. Robertson Bristol University

John Rogers University of California, Los Angeles

A. G. Rud Purdue University

Felicia C. Sanders The Pennsylvania State University

Janelle Scott University of California, Berkeley

Kimberly Scott Arizona State University

Dorothy Shipps Baruch College/CUNY

Maria Teresa Tatto Michigan State University

Larisa Warhol University of Connecticut

Cally Waite Social Science Research Council

John Weathers University of Colorado, Colorado Springs

Kevin Welner University of Colorado, Boulder

Ed Wiley University of Colorado, Boulder

Terrence G. Wiley Arizona State University

John Willinsky Stanford University

Kyo Yamashiro University of California, Los Angeles

* Members of the New Scholars Board

archivos analíticos de políticas educativas
consejo editorial

Editor: **Gustavo E. Fischman** (Arizona State University)

Editores. Asociados **Alejandro Canales** (UNAM) y **Jesús Romero Morante** (Universidad de Cantabria)

Armando Alcántara Santuario Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

Claudio Almonacid Universidad Metropolitana de Ciencias de la Educación, Chile

Pilar Arnaiz Sánchez Universidad de Murcia, España

Xavier Besalú Costa Universitat de Girona, España

Jose Joaquín Brunner Universidad Diego Portales, Chile

Damián Canales Sánchez Instituto Nacional para la Evaluación de la Educación, México

María Caridad García Universidad Católica del Norte, Chile

Raimundo Cuesta Fernández IES Fray Luis de León, España

Marco Antonio Delgado Fuentes Universidad Iberoamericana, México

Inés Dussel FLACSO, Argentina

Rafael Feito Alonso Universidad Complutense de Madrid, España

Pedro Flores Crespo Universidad Iberoamericana, México

Verónica García Martínez Universidad Juárez Autónoma de Tabasco, México

Francisco F. García Pérez Universidad de Sevilla, España

Edna Luna Serrano Universidad Autónoma de Baja California, México

Alma Maldonado Departamento de Investigaciones Educativas, Centro de Investigación y de Estudios Avanzados, México

Alejandro Márquez Jiménez Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

José Felipe Martínez Fernández University of California Los Angeles, USA

Fanni Muñoz Pontificia Universidad Católica de Perú

Imanol Ordorika Instituto de Investigaciones Economicas – UNAM, México

María Cristina Parra Sandoval Universidad de Zulia, Venezuela

Miguel A. Pereyra Universidad de Granada, España

Monica Pini Universidad Nacional de San Martín, Argentina

Paula Razquin UNESCO, Francia

Ignacio Rivas Flores Universidad de Málaga, España

Daniel Schugurensky Universidad de Toronto-Ontario Institute of Studies in Education, Canadá

Orlando Pulido Chaves Universidad Pedagógica Nacional, Colombia

José Gregorio Rodríguez Universidad Nacional de Colombia

Miriam Rodríguez Vargas Universidad Autónoma de Tamaulipas, México

Mario Rueda Beltrán Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

José Luis San Fabián Maroto Universidad de Oviedo, España

Yengny Marisol Silva Laya Universidad Iberoamericana, México

Aida Terrón Bañuelos Universidad de Oviedo, España

Jurjo Torres Santomé Universidad de la Coruña, España

Antoni Verger Planells University of Amsterdam, Holanda

Mario Yapu Universidad Para la Investigación Estratégica, Bolivia

arquivos analíticos de políticas educativas
conselho editorial

Editor: **Gustavo E. Fischman** (Arizona State University)
Editores Associados: **Rosa Maria Bueno Fisher** e **Luis A. Gandin**
(Universidade Federal do Rio Grande do Sul)

Dalila Andrade de Oliveira Universidade Federal de Minas Gerais, Brasil
Paulo Carrano Universidade Federal Fluminense, Brasil
Alicia Maria Catalano de Bonamino Pontifícia Universidade Católica-Rio, Brasil
Fabiana de Amorim Marcello Universidade Luterana do Brasil, Canoas, Brasil
Alexandre Fernandez Vaz Universidade Federal de Santa Catarina, Brasil
Gaudêncio Frigotto Universidade do Estado do Rio de Janeiro, Brasil
Alfredo M Gomes Universidade Federal de Pernambuco, Brasil
Petronilha Beatriz Gonçalves e Silva Universidade Federal de São Carlos, Brasil
Nadja Herman Pontifícia Universidade Católica –Rio Grande do Sul, Brasil
José Machado Pais Instituto de Ciências Sociais da Universidade de Lisboa, Portugal
Wenceslao Machado de Oliveira Jr. Universidade Estadual de Campinas, Brasil

Jefferson Mainardes Universidade Estadual de Ponta Grossa, Brasil
Luciano Mendes de Faria Filho Universidade Federal de Minas Gerais, Brasil
Lia Raquel Moreira Oliveira Universidade do Minho, Portugal
Belmira Oliveira Bueno Universidade de São Paulo, Brasil
Antônio Teodoro Universidade Lusófona, Portugal
Pia L. Wong California State University Sacramento, U.S.A
Sandra Regina Sales Universidade Federal Rural do Rio de Janeiro, Brasil
Elba Siqueira Sá Barreto Fundação Carlos Chagas, Brasil
Manuela Terrasêca Universidade do Porto, Portugal
Robert Verhine Universidade Federal da Bahia, Brasil
Antônio A. S. Zuin Universidade Federal de São Carlos, Brasil