Science
Education

The Relationship Between Time Allocated for Science in Elementary Schools and State Accountability Policies

EUGENE JUDSON
Mary Lou Fulton Teachers College, Arizona State University, Mesa, AZ 85212, USA

Received 27 June 2012; accepted 26 January 2013
DOI 10.1002/sce.21058
Published online 14 June 2013 in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: In the United States, student achievement results from mathematics and reading are always included in high-stakes accountability calculations. Because of this, it has been argued that other subjects have been minimized due to assessment and accountability policies. In this study, the accountability practices of states were differentiated into three groups based on the degree to which science achievement contributed to accountability formulas. Data from National Assessment of Educational Progress Teacher Questionnaires were evaluated to assess the amount of time fourth-grade teachers devoted to mathematics and reading in 2003 and 2011, as well as to science in 2005 and 2009. Teachers from the three groups of states reported spending equivalent amounts of time on mathematics and reading. However, the frequency of teachers reporting spending at least 4 hours of weekly instructional time on science was significantly higher in states that integrated fourth-grade science achievement into accountability formulas versus states where science did not figure in high-stakes accountability. Implications of this significant difference are discussed and are related to states' applications to receive waivers from No Child Left Behind and to alignment with the goals of the Next Generation Science Standards. © 2013 Wiley Periodicals, Inc. *Sci Ed* **97:**621–636, 2013

Correspondence to: Eugene Judson; e-mail: Eugene.Judson@asu.edu

INTRODUCTION

Unlike many other nations, education in the United States is not under precise national control. While the federal government certainly has an influential role that it can leverage through funding, plenary authority over education rests with the states. During the past two decades, there has been a growing national emphasis on high-stakes testing and accountability, and within this landscape states have varied in how they have chosen to use, or not use, science in their school accountability practices. The focus of this study was to assess whether variation in how states use science in their accountability programs could be linked to the amount of time fourth-grade students spend engaged in science learning.

A narrowed curriculum is commonly typified as being tapered in favor of concentrating on reading and mathematics. The rationale most often follows that curriculum has been narrowed because student achievement results from reading and mathematics are always included in states' accountability formulas that are used to determine whether schools and local educational agencies (LEAs) have met Adequate Yearly Progress (AYP). Beyond a simple stigma of missing AYP, the consequence of not meeting AYP over consecutive years has been quite significant as schools and LEAs could be subject to their state department of education restructuring the institution by removing teachers, recommending curriculum changes, firing a principal, and even removing authority of control from the local school board.

A good deal of literature in the past several years has supported the idea that No Child Left Behind (NCLB) has narrowed curriculum in terms of other subjects, such as science and social studies receiving less attention in classrooms, particularly elementary classrooms (e.g., Kingsbury, 2007; McMurrer, 2008; Milner, Sondergeld, Demir, Johnson, & Czerniak, 2012). Researchers also discovered that a narrowed NCLB curriculum could lead to instructional styles being altered in support of specific problem types and formats found on standardized assessments (Hamilton et al., 2007).

Past studies have typically relied on single surveys administered to teachers (e.g., Griffith & Scharmann, 2008) that prompted teachers to reflect on the impact of NCLB and the consequential distribution of instructional time among various subjects. Collectively, these studies provide evidence in support of the sentiment that, in the wake of NCLB, the need to support mathematics and reading achievement led to diminished time spent on other subjects. Where this study differs is that instead of attempting to examine the overall effect of NCLB on instructional time, a finer grain was examined to evaluate the effect of different types of state accountability policies. Whether those state policies were a direct result of NCLB was not consequential to this study, only that the differences existed.

Looking forward, it is anticipated that reauthorization of the Elementary and Secondary Education Act (ESEA) will not be addressed by Congress until 2014 (Klein, 2012). In the interim, waivers from NCLB measures have been granted to states that have developed alternate accountability plans that have been approved by the U.S. Department of Education (USDOE). However, the majority of the states that either have already been granted these waivers or are in the process of applying have not included subjects other than the traditionally assessed areas of reading and mathematics (McNeil, 2012). On the whole, while the approved waiver plans do emphasize new teacher accountability plans, the plans typically incorporate data from only reading and mathematics. That is, while a state may assess science, the results are not necessarily incorporated into accountability formulas.

Given that the imminent reauthorization of ESEA will provide policy makers an opportunity to reexamine the value of integrating subjects such as science into required accountability plans, it is befitting to look back at the recent history of how accountability policies have affected teaching and learning and to then evaluate how science has been treated differently among state accountability policies. These variations of policies and related effects on science learning and teaching can then provide some guidance when new policy is developed at federal, state, and local levels.

By and large, previous research has inspected the impact of accountability measures and its underpinning push to improve mathematics and reading achievement. Several researchers have focused locally on administrators (e.g., Zellmer, Frontier, & Pheifer, 2006) or teachers in one state or even one city (e.g., Finnigan & Gross, 2007; Jones & Swanson, 2009) and have highlighted teachers' reports of a narrowing curriculum leading to greater focus on mathematics and reading at the cost of time spent on science and other subjects that are not included in accountability calculations. The hazard of such regionally centered research is attempting to generalize the findings to a larger scale. In addition, regionally centered research does not account for variation of accountability practices. Scrutinizing accountability policies within the United States, it is clear that states, and at times school districts, have applied varying practices of accountability. How those practices are executed, one would expect, affect classroom routines. A few studies have included national level data, and researchers have deduced implications from analysis of those data. Notable among these is the report from the Center on Education Policy (CEP) that examined surveys from a nationally representative set of 349 school districts (McMurrer, 2008). Here too, even though the CEP study highlighted that nearly a third of school districts reported a decrease in time spent on science in the elementary classrooms, averaging 75 minutes per week, there was no distinction made among the districts in the sample regarding how their states executed accountability plans. While the researchers did importantly draw attention to science instructional time being diminished in many school districts since enactment of NCLB, there was little acknowledgment that the data also demonstrated that 62% of the reporting school districts did not indicate that their time spent on science had decreased, nor was there speculation as to why these differences might exist.

The focus of this study was to examine the relationship between the amount of time spent on science instruction in fourth-grade classrooms and the varied use of science achievement in accountability calculations among the states. The emphasis of this study was therefore not to determine whether NCLB had affected state policies, but instead to assess whether differences in the allocation of classroom instructional time devoted to science existed when there were *differences* in accountability policies specific to science. Waivers notwithstanding, the states have had latitude when establishing their accountability procedures and have been allowed to include achievement results from subjects beyond mathematics and reading in their high-stakes measurements. Regarding the subject of science, a minority of states have included students' science achievement results in their accountability formulas.

Previous research has indicated that states that do use science achievement results in their accountability procedures yield significantly higher fourth-grade achievement results on the science assessment of the National Assessment of Educational Progress (NAEP) versus those states that do not integrate science achievement results into their accountability procedures (Judson, 2010). Furthermore, examination of NAEP data has shown that states that integrate science into accountability achieve significantly higher fourth-grade student achievement in science while also maintaining equivalent achievement in mathematics and reading (Judson, 2012). These prior studies then support the idea that integrating science achievement into accountability measures manifests in greater student achievement in the elementary classroom. That is, in states where science achievement has "counted," alongside reading and mathematics achievement, students achieve relatively greater science results in fourth-grade classrooms as measured by NAEP. This prior research did not indicate a parallel in eighth-grade classrooms. That is, in eighth-grade science, achievement was not found to correlate with whether or not a state integrated science achievement into its accountability formula. It has been conjectured that this difference between the fourth- and eighth-grade findings is due to the nature of conventional science instruction in these two grade levels. Eighth-grade students typically move from one subject-based classroom to another throughout the day and have a dedicated amount of time for science, just as they do for other subjects. However, fourth-grade teachers often teach multiple subjects and are given latitude in deciding how to budget instructional time. So while prior research indicates that science is a diminished subject in elementary classrooms, there exists the possibility that when accountability systems do include science achievement in calculations that instructional time devoted for science in elementary classrooms benefits.

What is missing from the research is a connection between the generalized research that demonstrates science has been minimized in classrooms and the national scope research that points to evidence that states with science integrated into accountability practices outperform other states on fourth-grade NAEP science assessments. We do not know whether science is treated differentially when accountability practices include science results. The purpose of this study was to investigate whether the amount of time devoted to science instruction in fourth-grade classrooms differed among groups of states, based on how they have or have not integrated science into their accountability formulas. Similarly, this study sought to determine whether the amount of time devoted to mathematics and reading differed among these states in their fourth-grade classrooms. Specifically, these two research questions were addressed:

1. Does instructional time devoted to science 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 instructional time devoted to mathematics and reading differ between states that use science as part of their accountability calculations and states that do not use science in their accountability calculations?

BACKGROUND

The Value of Time

It is commonsensical that to learn about a subject time needs to be devoted to that goal. Whether the approach to learning is through inquiry or direct instruction, time is a fundamental ingredient. Of course, the amount of time needed to achieve learning goals has been often raised in educational research literature. In his seminal model of learning, Carroll (1963, 1989) described how academic achievement was dependent on variables representing the amount of time available to learn, the time needed to learn, and the time a student is willing to spend learning. The factor of time in learning has been viewed as "an exceedingly obvious variable" (Carroll, 1989, p. 27), yet it is a factor that is not always well addressed in the research literature.

Bloom (1984) contended in his model of mastery learning that as student achievement increases there is a decrease in the variance of time needed to learn additional content. This idea that instructional time possesses a type of elasticity to achieve goals was affirmed by Brown and Saks (1986) who found that time had a positive effect on achievement and that how time was allocated across grade levels and subjects varied considerably and included apportioning time to pursuits such as time for students to self-pace and time for group work. Of course, learning occurs when allotted instructional time becomes a period when learners are engaged with the content. Determining what is going on inside of a student's mind would then be a better predictor of learning than the amount of minutes spent on a topic. However, the amount of time dedicated to learning about a subject can be used to interpret the value placed on that subject.

The seemingly simple concept that time is tied directly to the opportunity to learn (Carroll, 1963; Sorenson & Hallinan, 1977) implies for researchers that learning theories, which emphasize only individual learning determinants and/or teaching methods, overlook the important aspect of quantity. Yet time alone is certainly not a reliable predictor of achievement; an examination of data from the Programme for International Student Assessment (PISA) and the Third International Math and Science Survey (TIMSS) from multiple countries revealed no significant relationships between total numbers of instructional hours and achievement (Baker, Fabrega, Galindo, & Mishook, 2004). However, when these researchers examined within-country relationships between time and achievement, significant relationships were revealed. This suggests that educational systems are using time differently. It is more appropriate then to view time as a type of resource that can be managed rather than as just a substance that is applied to children. In that light, time is an extremely valuable resource that can be mismanaged.

Accountability and Instructional Time

Reports of less time being devoted to science instruction in elementary classrooms since the inception of NCLB are not difficult to find (e.g., Kingsbury, 2007; Linn, 2008; McMurrer, 2007, 2008). In a comprehensive examination of all 50 states, Rentner et al. (2006) found that, in response to NCLB, 71% of school districts had reduced elementary school instructional time in at least one subject to have more time for reading and mathematics. Science and social studies were the subjects that were most often shortchanged (Rentner et al., 2006). A survey of teachers by the National Science Teachers Association revealed nearly half of the teachers reported that time allotted for science had decreased in the 2010–2011 school year compared to the year before (Petrinjak, 2011). There were some indications that reductions in time for science were not always due to the time being reapportioned for mathematics or reading. Instead, sometimes science was suspended to allow time for pictures and assemblies. However, the overall implication was that these reductions occurred because science was often viewed as being not as important as mathematics or reading and this was due to science not being included in accountability calculations. This finding was parallel to Griffith and Scharmann's (2008) reporting of a majority of elementary teachers reducing the amount of time spent on science due to the enactment of NCLB. In their study, teachers who indicated they had decreased instructional time for science were questioned why the time had been reduced. Their responses suggested that "cuts were due to the need to improve performance in the assessed content areas" (p. 40). The message that curriculum has been narrowed due to high-stakes accountability and that science is among those subjects often squeezed out is a consistent finding in the research literature (Cawelti, 2006).

However, NCLB alone is not likely the singular reason science has been neglected in favor of other subjects. Prior to NCLB, data indicated that, in first- through fourth-grade classrooms, science received the least amount of dedicated time compared to the other core subjects of reading, mathematics, and social studies (Perie, Baker, & Bobbitt, 1997). This was aligned with findings from the late 1990s, when standards-based learning was taking a greater foothold, and second-, fifth-, eighth-, and 12th-grade teachers all estimated needing less time to address the science standards than was needed for either language arts or mathematics (Florian, 1999). Yet these pre-NCLB reports of fewer hours being dedicated to science in elementary schools, as compared to mathematics and reading, are perhaps not surprising. In fact, one would be hard pressed to find reports of science instruction running amuck in elementary schools prior to NCLB. Pre-NCLB researchers reported a recognizable story of science often being avoided by elementary teachers due to lack of familiarity with the content and teachers having negative attitudes and feelings of inadequacy related to science (Tilgner, 1990). So while research indicates NCLB did have an impact on science instructional time, we must also consider the idea that when NCLB was enacted many teachers were quite willing to allow attention to be drawn toward other subjects.

Of greater interest is the contrast of time allocated for science in elementary classrooms from pre-NCLB era to after enactment. Comparing the 1987–1988 school year to the 2003–2004 school year (a couple of years after the implementation of NCLB), the total percentage of a student's school week allocated for the subjects of English, mathematics, social studies, and science remained relatively stable—67.1% in 1987–1988 and 66.7% in 2003–2004 (Morton & Dalton, 2007). Nevertheless, the proportion of the school week allotted for science decreased from 8.1% to 7.1%. While this is a defined decrease, overall it may not be the precipitous decline implied by some reports and many anecdotes.

Although some research indicates that generally less time was spent on nontested subjects than was spent prior to NCLB, other reports somewhat challenge this conception. The CEP (2005) released a report based on a survey of districts from across the country that indicated the majority of school districts reported that instructional time spent on subjects other than reading and mathematics had been reduced only minimally or not at all. However, this same report highlighted that school districts that did report diminished time spent on science specified this was due to the pressure to perform well on high-stakes reading and mathematics assessments.

While reports of schools placing greater emphasis on mathematics and reading are prevalent (e.g., Dee & Jacob, 2010; McMurrer, 2007; Srikantaiah & Kober, 2009), there are some indications that subjects that are not frequently part of high-stakes accountability programs are not completely overlooked. In an analysis of interviews conducted with state assessment directors, Pederson (2007) revealed that after the enactment of NCLB the integration of nontested subjects into tested subjects increased. However, Pederson's findings that science became more prevalent in state accountability systems was tied directly to the NCLB requirement that science be assessed at minimum once annually in three different grade levels (3–5, 6–9, and 10–12). So while it can safely be said that NCLB led many states to begin administering science assessments, because the results of those science assessments lacked the consequence of those from mathematics and reading, the question remains as to how classrooms were affected.

Because some researchers have connected high-stakes accountability practices with improved student achievement in mathematics (Carnoy & Loeb, 2002; Dee & Jacob, 2011) and in reading (Jacob, 2005; Kober, Chudowsky, & Chudowsky, 2008), it is also important to assess the connection with science. Determining simply if the amount of time spent on science is affected by accountability practices is an initial step toward understanding how accountability and the practices of teaching science in the elementary classroom are linked. A following step would then be to evaluate how new accountability policies that are developed in tandem with the reauthorization of ESEA and adoption of the Next Generation Science Standards (NGSS) can be shaped by this and related research.

METHODOLOGY

Classifying States

Before classifying the states, it was important to establish a definition of what it meant for science achievement results to "count." Although most states during the years examined only assessed their students in the content area of science, some did go further and integrated those results into some type of accountability program. It was established that for a state to be considered one that appreciably used science in an accountability system, that the accountability system had to carry with it the possibility of sanctions and/or rewards. For instance, Texas did include science in a reporting system of school performance, but there were no sanctions or rewards associated with achievement level within that reporting system. From this point forward, when it is stated that a state "used" science in an accountability program, it means that the state integrated the quantitative student achievement results from a mandated science assessment and that those results were incorporated into an algorithm that potentially led to reward or punitive consequences for a school.

To attempt to determine whether a link existed between how states have used, or not used, science achievement in their accountability programs and how much time fourth-grade teachers allocated for science, the states were categorized into three groups. The groups were based on how the states used science achievement in their accountability program calculations. Here, the definition of an accountability program was broadened beyond just that of AYP. Even though it was required for all of the states to assess science in at least three grades, there was no requirement within NCLB that these results be integrated into AYP calculations. Although very few states did this, a state may have chosen to integrate science directly into its AYP calculations as an "additional indicator" of achievement. States were required to use graduation rate as the additional indicator for their secondary schools but had some liberty to have an additional indicator approved by the USDOE for their elementary schools. The great majority of states used attendance rate as their additional indicator. However, some states used science achievement as their AYP additional indicator for schools that included elementary grades.

In addition to integrating science achievement directly into AYP calculations to meet federal requirements, it was possible to include these achievement results in an alternate accountability system. A total of 23 states maintained a separate accountability program after the federal AYP-based accountability programs became a requirement in 2002. Of those 23 states, several required that science achievement results be included. For example, in South Carolina state statutes required that science assessment results contribute 20% of the accountability formula in Grades 3–5 and 25% in Grades 6–8. In all, 11 states integrated science achievement into either their AYP federal accountability system or into a separate dual accountability system.

Straightforwardly, the 39 states that did not integrate science into an accountability system comprised one group. However, because NAEP teacher survey data exist only for fourth-grade teachers in the elementary grade range, the states were further grouped based on this parameter. Specifically, for the 11 states that used science in an accountability system, not all of those states used fourth-grade results and some allowed science to be selected from a menu of choices as an indicator of success. For example, while North Carolina did require science results from the elementary grades to be used in their accountability program, those results were drawn from fifth and eighth grades. As a further example, in Georgia a school could select science as their additional indicator when deciding what would be used to determine AYP.

TABLE 1
Groups Based on Use of Fourth Grade Science Achievement in Accountability Calculations

Group	States
Group 1: Require science in accountability from fourth grade	Kentucky, New York, South Carolina, Tennessee, Utah
Group 2: Include science in accountability, but either as a choice or do not use fourth-grade results	California, Georgia, Michigan, North Carolina, Ohio, Virginia
Group 3: Do not include science in accountability	All other states

The 11 states that integrated science results into accountability programs were then divided based on whether they used fourth-grade results and whether that was a requirement. This led to one group of five states that had a requirement of using fourth-grade science results. The other six states that used science in some manner comprised a second group; however, these states either did not use fourth-grade results or they allowed the schools to choose science from a menu of options. Table 1 indicates how the states were grouped for comparison. The primary focus of this study was obviously to examine differences between states that required the use of fourth-grade science achievement results (Group 1) and states that did not integrate science at all (Group 3). The decision to include the Group 2 states in analyses was to determine whether their intermediate integration of science into accountability policies could be detected as a transitional point.

Comparing Instructional Time

Instructional time allocation was based on the responses of fourth-grade teachers to NAEP Teacher Questionnaires. These questionnaires have been administered by NAEP for many years as part of their comprehensive data collection system. Although the primary focus of this study was to compare differences between sets of states based on how they integrated science into accountability, a subordinate enquiry was to consider whether NCLB had affected those practices. However, only national level data were available for years prior to 2002; state level data were available beginning in 2003. Because this study was focused on comparing groups of states based on the way they integrated, or did not integrate, science into their accountability formulas, the lack of pre-NCLB data precluded any reasonable pre- to post-NCLB comparisons of time allocation among the groups of states. Nevertheless, it was decided that data would be selected so that early-NCLB enactment could be compared to later-NCLB enactment. In either case, the focus would still be on comparing the amount of time devoted to science, relative to how the states integrated science into their accountability formulas.

The examination of data primarily focused on the most recently available data years—2009 for science and 2011 for mathematics and language arts. The related analysis to examine data from the available year immediately following the enactment of NCLB meant that data from 2003 for mathematics and language arts, and from 2005 for science, were to be used. At first glance, the use of the 2005 did not seem ideal since NCLB was signed into law by President Bush in January 2002, and the 2002–2003 academic year was the first year of enactment. However, the given state of data was considered acceptable in view of how NCLB was phased in over several years. Specifically, NCLB began with the requirement of assessments for mathematics and language arts being administered once between Grades 3–5, 6–9, and 10–12. Administration of assessments in mathematics and language arts in all Grades 3–8 were not required until the 2005–2006 academic year. Having science assessed in at least three grades did not become a requirement until the 2007–2008 academic year. Therefore, the purpose of this related analysis that focused on the first available year following the enactment of NCLB was to assess whether there were differences among the groups of states at this earlier juncture when NCLB was still making its way through the educational system. It is emphasized that the data from 2003 and 2005 were not considered to represent "pretest" points, but instead simply denoted earlier points in time on the NCLB timeline. The recent years of 2009 and 2011 would then provide a more complete view of the differences, if any, among the groups of states after NCLB had been in place for approximately a decade.

The NAEP Teacher Questionnaire data are from teachers' self-reports regarding their practices, and these reports are not corroborated by the National Center for Education Statistics that oversees the administration of NAEP. As with any study that incorporates self-reported data, there are concerns that these types of data may have inherent inaccuracies, as some researchers have found self-reported data to be not as reliable as data collected from observations (Mueller & Zeidler, 1998; Simmons et al., 1999). However, there was no reason to believe that any possible error in reporting would be skewed more toward one group of states than another. The NAEP Teacher Questionnaires were considered additionally useful because they did not include any questions that prompted teachers to consider NCLB policies, as instruments of other studies have done, which might bias respondents.

A series of analyses of variance (ANOVAs) were conducted to examine potential differences among the three groups of states. The survey data were available as percentages of teachers reporting in each response category to questions about how they allocated instructional time. For example, teachers could respond as spending *less than 1 hour* per week on science, *1–1.9 hours, 2–2.9 hours*, and so on. The ANOVAs were carried out for each response category per content area. The null hypotheses were that groups of states would report equivalently within each response category. To account for the multiple comparisons and to reduce possible familywise Type I error, a Bonferroni correction was applied. This led to the typical p value of significance of .05 being divided by a value of three, which represents the three dependent variables (i.e., state groups). The adjustment led to results being considered significant only at a level less than or equal to a p value of .017. Significant omnibus F-test results were followed up with Tukey HSD (honestly significant difference) post hoc comparisons to determine where the significant differences were.

RESULTS

Results from the ANOVAs related to the amount of time spent on language arts indicated that there were no differences in any of the reporting categories among the fourth-grade teachers in 2011 (Table 2). For all three groups of states, the plurality of teachers reported in 2011 dedicating 10 or more hours of weekly instructional time to language arts. The response options in 2003 (*less than 7 hours*, *7–9.9 hours, 10–12.9 hours,* and *13 or more hours*) were different than those provided in 2011 and a straightforward contrast is not allowable. However, as the point of this study was to assess whether there were differences among the groups of states related to how instructional time was spent in fourth-grade classrooms and not necessarily assess changes over time, the data from 2003 were still useful. ANOVA results indicated that there were no significant differences among the groups of states in 2003 regarding how much instructional time was spent on language arts. That is, states that used science in their accountability systems (i.e., Group 1) did not report spending instructional time on language arts any differently from the other two groups of

TABLE 2
Weekly Instructional Time Spent on Language Arts in Fourth-Grade Classrooms, 2011

Category	State Group	*n*	Mean Percentage	*SD*	*F*	*p*
Less than 3 hours	1	5	2.8	0.8	1.00	.38
	2	6	3.5	1.9		
	3	39	2.7	1.3		
3–4.9 hours	1	5	6.6	2.2	0.44	.65
	2	6	5.3	1.6		
	3	39	5.8	2.3		
5–6.9 hours	1	5	14.4	4.1	2.57	.09
	2	6	15.2	4.5		
	3	39	11.7	4.0		
7–9.9 hours	1	5	34.8	8.9	0.22	.80
	2	6	31.5	9.6		
	3	39	33.0	8.0		
10 or more hours	1	5	42.0	13.3	0.61	.55
	2	6	45.2	12.5		
	3	39	47.0	9.0		

TABLE 3
Weekly Instructional Time Spent on Mathematics in Fourth-Grade Classrooms, 2011

Category	State Group	*n*	Mean Percentage	*SD*	*F*	*p*
Less than 3 hours	1	5	2.4	1.7	0.50	.61
	2	6	3.7	1.6		
	3	39	3.4	2.4		
3–4.9 hours	1	5	9.2	4.4	1.37	.26
	2	6	6.0	3.2		
	3	39	6.6	3.5		
5–6.9 hours	1	5	59.2	3.9	0.15	.86
	2	6	62.7	9.7		
	3	39	60.8	11.3		
7 or more hours	1	5	29.2	3.9	0.05	.95
	2	6	27.7	10.6		
	3	39	29.2	11.4		

states. The 2003 results in language arts then paralleled the findings of 2011 in language arts.

Similar analyses were conducted for the 2011 and 2003 fourth-grade teachers' responses regarding how much time was spent weekly on mathematics. ANOVA results of the 2011 data indicated there were no significant differences in any of the response categories (Table 3). For all three groups of states, the smallest proportion reported allocating *less than 3 hours* of weekly instructional time to mathematics and the majority of the teachers indicated spending *5–6.9 hours* of weekly time on mathematics.

As was the case with language arts, there were no significant differences in 2003 between the groups in any of the response categories related to how much time was spent on

TABLE 4
Weekly Instructional Time Spent on Science in Fourth-Grade Classrooms, 2005 and 2009

		2005					2009				
Category	State Group	*n*	Mean percentage	*SD*	*F*	*p*	*n*	Mean percentage	*SD*	*F*	*p*
Less than 1 hour	1	5	3.2	1.5	1.22	.304	5	2.9	1.5	1.50	.233
	2	6	5.3	4.6			6	5.2	4.0		
	3	39	6.6	4.9			35	6.8	5.1		
1–1.9 hours	1	5	11.4	8.3	2.07	.137	5	11.8	7.7	1.95	.154
	2	6	13.8	11.2			6	11.6	8.1		
	3	39	18.9	8.8			35	18.0	9.6		
2–2.9 hours	1	5	26.6	15.0	3.12	.053	5	24.6	15.4	2.11	.133
	2	6	31.2	5.7			6	32.0	4.8		
	3	39	35.2	6.8			35	33.1	7.9		
3–3.9 hours	1	5	27.8	9.8	.603	.552	5	25.2	8.7	.57	.568
	2	6	31.3	9.9			6	30.2	9.2		
	3	39	27.2	8.4			35	26.4	8.5		
4 or more hours	1	5	30.6	29.1	5.88	.005*	5	35.4	9.2	4.77	.013*
	2	6	18.3	10.0			6	21.1	8.6		
	3	39	12.2	8.1			35	15.7	11.0		

*$p \leq .017$.

mathematics. Again, the response categories for mathematics in 2003 (*less than 7 hours, 7–9.9 hours, 10–12.9 hours*, and *13 or more hours*) differed from those in 2011 (Table 3). The majority of the fourth-grade teachers in 2003 reported spending less than 7 hours weekly on mathematics, with the scope ranging from 55.8% among Group 1 teachers to 56.4% among Group 3 teachers.

It was only in the content area of science that the fourth-grade teachers from the groups of states provided significantly different responses regarding the allocation of instructional time. The response categories (l*ess than 1 hour, 1–1.9 hours, 2–2.9 hours, 3–3.9 hours,* and *4 or more hours*) were the same in 2005 and 2009 so those data are presented together in Table 4. Inspection of the ANOVA results from the 2005 science data revealed that the groups of teachers responded significantly different in the category of *4 or more hours*, $F(2, 47) = 5.88$, $p = .005$. Tukey HSD comparisons were then conducted to determine where the differences between the groups existed. The post hoc Tukey HSD tests revealed significant differences in 2005 between Group 1 and Group 3 in the category of *4 or more hours* ($p = .005$). In 2009, the teachers' reporting of weekly class time devoted to science remained significantly different among the states in the category of *4 or more hours*, $F(2, 47) = 4.77$, $p = .013$. Again the post hoc tests revealed that there was a significant difference in the reporting rates between the Group 1 teachers and the Group 3 teachers in this category of *4 or more hours* ($p = .011$).

Fourth-grade teachers from states where the use of fourth-grade student science achievement was a requirement of state accountability procedures (i.e., Group 1) reported allocating 4 or more hours of classroom time weekly for science at a rate more than twice that of fourth-grade teachers from states that did not include student science achievement at all in accountability calculations (i.e., Group 3) in both 2005 and 2009. The *4 or more hours* category was the uppermost limit among the response choices. Partial eta-squared ($\eta_p{}^2$)

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 being a member of a state that integrated science into accountability was small for the *4 or more hours* category in both 2005 ($\eta_p^2 = .200$) and 2009 ($\eta_p^2 = .181$). Using the corrected Bonferroni p value of .017, retrospective analysis, related to the category of teaching *4 or more hours* of science, indicated power was 0.723 in 2005 and 0.604 in 2009. It is also noted, although the differences were not statistically significant, that fourth-grade teachers from the Group 3 states reported allocating *less than 1 hour* to science weekly at a rate more than twice that of the Group 1 fourth-grade teachers.

Because the response categories (e.g., *2–2.9 hours*) remained stable for science, but not in mathematics and language arts, it was possible to examine the changes from 2005 to 2009 for the content area of science. Paired samples t tests were conducted for each group of states to determine whether the changes from 2005 to 2009 were significant. There were no significant changes for either Group 1 or Group 2 for any of the time allocation reporting categories for science. It is noted that Groups 1 and 2 have small n values and detection of any significant differences would involve substantial changes from 2005 to 2009. There were significant changes between 2005 and 2009 for the Group 3 states ($n = 35$) in the categories of *2–2.9 hours* ($t = 3.32$, $p < .01$) and *4 or more hours* ($t = 4.75$, $p < .01$). The fourth-grade teachers from Group 3 reported a slight but significant decrease in the middle category of *2–2.9 hours* while also reporting a significant increase of more than 3 percentage points in the highest category of *4 or more hours*.

DISCUSSION

Before attempting to interpret any implications from the findings of this study, limitations must be noted. These NAEP data provide at best only a glimpse into the relative weight of instructional time placed on the subjects of mathematics, language arts, and science. Yet, NAEP is designed to be a national representation and in the 2011 assessment, fourth-grade teachers represented 8,500 schools, so the representation is fairly extensive. What is not extensive is an understanding of the degree to which any of the subjects of mathematics, language arts, and science were integrated and possibly "double counted." Also, the teacher data open up questions regarding more precise measurements of time. Particularly, categories such as *4 or more hours*, leads one to wonder how many teachers were spending 5, 6, or even more hours on science on a weekly basis. The categorical data also did not permit more precise analyses beyond the ANOVA comparisons for each ordinal response.

Despite these limitations, there are notable inferences that can be drawn from two key results of this study. First, the results indicate that fourth-grade teachers who work in states where the achievement results from their fourth-grade students are included in accountability calculations report allocating at least 4 hours of weekly classroom time to the subject of science at a significantly higher rate than teachers from other states. Teacher data were available only as proportions of teachers responding to time allocation response categories (e.g., *less than 1 hour*, *1 –1.9 hours*), and this did not permit an exact calculation of the average amount of hours distributed for science. However, the ordinal nature of the responses does make possible an overall assessment of time allocation. Simply, the Group 1 teachers are more often spending larger swaths of classroom time devoted to science. This finding is perhaps not surprising and in a sense follows the logic of earlier studies that reported that accountability has led to instructional time being shifted to those subjects that are counted. The difference here is that the environments of accountability have been differentiated, and it appears that science is receiving more classroom attention when it is included *robustly* in accountability measures. The robustness of how science is integrated into policy proved to be key in that no differences were found when comparing Group 2 (where science was integrated mildly) to either Group 1 or Group 3.

The primary focus of this study was to determine whether different types of accountability policies affected the amount of instructional time devoted to science. It is clear that teachers in the Group 1 states reported allocating at least 4 hours weekly for science more often than teachers in the other states. What is not clear is whether teachers in the Group 1 states made significant changes to their allocation of instructional time as a result of their states adapting to NCLB or if these teachers were in states where there already existed a culture of directing more attention to science in the elementary classroom than existed in the other states. This study also leaves us unsure whether the accountability policies of Group 1 states are related to practices beyond allocation of instructional time. Speculation that components, such as professional development and funding for elementary science, are notably different between the groups of states is only guesswork without further study.

Comparison of data also signifies that not only did the Group 1 teachers report the allocation of 4 or more hours of weekly instructional time for science at a rate significantly greater than Group 3 teachers in 2005 and in 2009, but the Group 1 teachers increased the proportion of teachers reporting the allocation of 4 or more hours for science during this period of time (from 30.6% to 35.4%). This increase from 2005 to 2009 was not statistically significant ($p = .207$), yet it is still noted because of the upward direction. This then parallels the small but significant increase in the reporting of at least 4 hours of classroom time for science among the Group 3 teachers from 2005 to 2009. Without further investigation, it is difficult to conjecture why this slight upward trend of science instructional time occurred among states where science was not integrated into high-stakes formulas. One possibility, though, is that the establishment of science assessments alone (required by the 2007–2008 school year) may have been enough to have a minor effect on instructional time, even though the assessments were not tied to school accountability.

Another important result from this study was the understanding that the teachers from all three groups of states did not allocate their time differently in the subjects of mathematics or language arts. The only significant differences among the states were found in how time was allocated for science. Therefore, although the condition of these data do not allow any strict ranking, it can be said that teachers from all three groups of states reported equivalently in all categories of time allocation in both mathematics and language arts in 2003 and again in 2011. Of course, an examination of the percentage of teachers allocating instructional time in the different categories does lead to a simple inference that language arts and mathematics do receive more classroom attention than science. Specifically, in 2011, in all three of the groups, approximately three-fourths of the teachers reported spending at least 7 hours of weekly classroom time on language arts and approximately 90% of the teachers in all of the groups reported spending at least 5 hours of classroom time on mathematics. These data can then be compared broadly with 35.4%, 21.1%, and 15.7% of teachers from Groups 1, 2, and 3, respectively, in 2009 reporting spending 4 or more hours weekly on science.

The findings of this study parallel those of earlier research related to how fourth-grade NAEP achievement compares among the states based on how the states used, or did not use, science achievement results in their accountability formulas. Prior research indicated that states incorporating science into accountability yielded significantly greater fourth-grade NAEP science achievement than states that did not include science achievement in their accountability calculations (Judson, 2010). In a sense, this study merely confirms a commonsense deduction that what gets tested (and counted) gets taught. More time was allotted for science in states where the results of fourth-grade students' science achievement contributed to determining whether a school would be subject to punitive consequences.

The line that connects the dots between instructional time and achievement is far from being a bold and straight line, so it cannot be implied that increasing instructional time for any subject is a quick route to higher achievement. There are simply too many variables that can contribute to achievement to allow such a simplistic conclusion. However, when juxtaposed against other reports, this research does support the idea that instructional time is one key contributor to science achievement in fourth-grade classrooms.

It appears also that when science is included in accountability in fourth-grade, that instructional time is not pilfered from language arts or mathematics. This leads to questions that need to be addressed with further study. It is not clear from this research if the teachers in the Group 1 states have managed to integrate science into other content, specifically language arts and mathematics. If the Group 1 states are indeed integrating science with these other two subjects, that would explain the significantly higher proportion of Group 1 fourth-grade teachers spending 4 or more hours weekly on science while also reporting no differently the proportion of time spent on language arts or mathematics. However, following the argument that accountability leads to a narrowed curriculum, the supposition arises that teachers in the Group 1 states are perhaps merely reducing time spent on other subjects, such as physical education and art, to a greater degree than teachers in Groups 2 and 3.

Both of the explanations presented to account for how the Group 1 fourth-grade teachers are able to allocate more time to science, while apparently not subtracting from language arts and mathematics, are viable: (a) science is more often being integrated with the other tested subjects and (b) time is being reallocated more dramatically from the other nontested subjects. These two conjectures paint very different pictures of the effect of including science in accountability. Going forward, it is important to get a better handle on the classroom effects of including science in accountability measures.

IMPLICATIONS

The results of this study occur at a time when the landscape of science education policy is in flux. For many states, due to their waiver applications being accepted, the AYP federal accountability system has been temporarily suspended; the NGSS will soon be adopted; aligned student science assessments will then be developed; and accountability systems will be revised to integrate these new standards and assessments. The NGSS developers indicate that state entities will decide whether or not they will develop science assessments aligned with NGSS. How achievement results from any new science assessments are integrated into accountability will also be a decision at the state level. As we move forward, it is fitting to consider the relationships among accountability, achievement, and classroom practice. Prior research suggests that when science is included alongside mathematics and reading in accountability formulas, that achievement is positively affected in elementary schools. This study indicates that the actual amount of time dedicated to science in fourth-grade classrooms is also related to accountability policy. The implication comes back to the adage that what gets tested gets taught, but the ramification of not following this maxim, which is now supported with evidence, can lead to science continuing the second-class status it has had in most of the states.

Recently, several states that applied for waivers from NCLB proposed to include assessments data from other subjects, particularly science, in their accountability plans (McNeil, 2011; Robelen, 2012). At first glance, this would seem to be a solid move toward broadening the scope of accountability measures by including science—a move that would align well with national goals of promoting science, technology, engineering, and mathematics. However, if science achievement is to be more commonly included in accountability, then educators and policy makers must be cautious that the intended and actual effects are aligned. Most educators would likely agree that a purpose of including science in accountability measures is to develop greater comprehension of science concepts. Yet if the vision of promoting science is regulated too heavily by high-stakes assessments and by a fear of possible penalties from accountability practices, then it is a concern that there will be too great of a concentration on test preparation and not enough of a focus on developing scientific thinking and interests in science fields. The NGSS support the development of students' abilities to engage in scientific and engineering practices such as carrying out investigations and communicating information. These are elements that are not easily assessed on single paper and pencil tests. So while including science in accountability formulas appears to be a step toward gaining more classroom attention for the subject, a great deal more consideration must be paid to developing the habits of mind promoted by the NGSS; allocation of classroom time is but one important step in this direction.

REFERENCES

Baker, D. P., Fabrega, R., Galindo, C., & Mishook, J. (2004). Instructional time and national achievement: Cross-national evidence. Prospects, 34(3), 311–334.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13(6), 4–16.

Brown, B. W., & Saks, D. H. (1986). Measuring the effects of instructional time on student learning: Evidence from the beginning teacher evaluation study. American Journal of Education, 94(4), 480–500.

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. Educational Evaluation and Policy Analysis, 24(4), 305–331.

Carroll, J. B. (1963). A model of school learning. Teachers College Record, 64(8), 723–733.

Carroll, J. B. (1989). The Carroll model: A 25-year retrospective and prospective view. Educational Researcher, 18(1), 26–31.

Cawelti, G. (2006). The side effects of NCLB. Educational Leadership, 64(3), 64–68.

Center on Education Policy (2005). NCLB: Narrowing the curriculum? Washington, DC: Author. Retrieved March 9, 2013, from http://www.cep-dc.org/displayDocument.cfm?DocumentID=239.

Dee, T. S., & Jacob, B. A. (2010). The impact of No Child Left Behind on students, teachers, and schools. Brookings Papers on Economic Activity, 41(2), 149–207.

Dee, T. S., & Jacob, B. A. (2011). The impact of No Child Left Behind on student achievement. Journal of Policy Analysis and Management, 30(3), 418–446.

Finnigan, K. S., & Gross, B. (2007). Do accountability policy sanctions influence teacher motivations? Lessons from Chicago's low-performing schools. American Educational Research Journal, 44(3), 594–630.

Florian, J. (1999). Teacher survey of standards-based instruction: Addressing time. Aurora, CO: Mid-continent Research for Education and Learning.

Griffith, G., & Scharmann, L. (2008). Initial impacts of No Child Left Behind on elementary science education. Journal of Elementary Science Education, 20(3), 35–48.

Hamilton, L. S., Stecher, B. M., Marsh, J. A., McCombs, J. S., Robyn, A., Russell, J. L., et al. (2007). Standards-based accountability under No Child Left Behind experiences of teachers and administrators in three states. Santa Monica, CA: RAND Corporation.

Jacob, B. (2005). Accountability, incentives and behavior: Evidence from school reform in Chicago. Journal of Public Economics, 89(5–6), 761–796.

Jones, R., & Swanson, E. (2009). Understanding elementary teachers' use of science teaching time: Lessons from the big sky science partnership. The Journal of Mathematics and Science: Collaborative Explorations, 11, 163–192.

Judson, E. (2010). Science education as a contributor to adequate yearly progress and accountability programs. Science Education, 94(5), 888–902.

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 March 9, 2013, from http://epaa.asu.edu/ojs/article/view/987.

Kingsbury, A. (2007, July 25). Schools cut other subjects to teach reading and math. U.S. News & World Report. Retrieved March 9, 2013, from http://www.usnews.com/usnews/edu/articles/070725/25nclb.htm.

Klein, A. (2012). No matter its makeup, new congress faces policy, fiscal logjam. Education Week, 32(7), 18.

Kober, N., Chudowsky, N., & Chudowsky, V. (2008). Has student achievement increased since 2002? State test score trends through 2006–07. Washington, DC: Center on Education Policy. Retrieved March 9, 2013, from http://www.cep-dc.org/displayDocument.cfm?documentid=320.

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.

McMurrer, J. (2007). Choices, changes, and challenges: Curriculum and instruction in the NCLB era. Washington, DC: Center on Education Policy. Retrieved March 9, 2013, 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 9, 2013, from http://www.cep-dc.org/displayDocument.cfm?documentid=309.

McNeil, M. (2011). Waiver plans would scrap parts of NCLB. Education Week, 31(13), 1, 28–29.

McNeil, M. (2012). Waiver applicants steer wary course. Education Week, 31(26), 1, 22–24.

Milner, A. R., Sondergeld, T. A., Demir, A., Johnson, C. C., & Czerniak, C. M. (2012). Elementary teachers' beliefs about teaching science and classroom practice: An examination of pre/post NCLB testing in science. Journal of Science Teacher Education, 23(2), 111–132.

Morton, B. A., & Dalton, B. (2007). Changes in instructional hours in four subjects by public school teachers of grades 1 through 4. Washington, DC: National Center for Education Statistics.

Mueller, J. C., & Zeidler, D. L. (1998). A case study of teacher beliefs in contemporary science education goals and classroom practices. Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Diego, CA.

No Child Left Behind Act of 2001. Public Law 107-110, 115 Stat. 1425 (2002). Retrieved March 9, 2013, from http://www2.ed.gov/policy/elsec/leg/esea02/index.html.

Pederson, P. V. (2007). What is measured is treasured: The impact of the No Child Left Behind Act on nonassessed subjects. Clearing House: A Journal of Educational Strategies, 80(6), 287–291.

Perie, M., Baker, D. P., & Bobbitt, S. (1997). Time spent teaching core academic subjects in elementary schools: Comparisons across community, school, teacher, and student characteristics. Washington, DC: National Center for Education Statistics.

Petrinjak, L. (2011, July 19). Elementary teachers getting less time for science. NSTA Reports, 17.

Rentner, D. S., Scott, C., Kober, N., Chudowsky, N., Chudowsky, V., Joftus, S., et al. (2006). From the capital to the classroom: Year 4 of the No Child Left Behind Act. Washington, DC: Center on Education Policy.

Robelen, E. W. (2012, April 25). Waivers to widen test menu. Education Week, 31(29), 1, 21.

Simmons, P. E., Emory, A., Carter, T., Coker, T., Finnegan, B., Crockett, D., et al. (1999). Beginning teachers: Beliefs and classroom actions. Journal of Research in Science Teaching, 36(8), 930–954.

Sorenson, A. B., & Hallinan, M. T. (1977). A reconceptualization of school effects. Sociology of Education, 50(4), 273–289.

Srikantaiah, D., & Kober, N. (2009). How state and federal accountability policies have influenced curriculum and instruction in three states. Washington, DC: Center on Education Policy. Retrieved March 9, 2013, from http://www.cep-dc.org/displayDocument.cfm?DocumentID=217.

Stevens, J. C. (1992). Applied multivariate statistics for the social sciences. Hillside, NJ: Erlbaum.

Tilgner, P. J. (1990). Avoiding science in the elementary school. Science Education, 74(4), 421–431.

Zellmer, M. B., Frontier, A., & Pheifer, D. (2006). What are NCLB's instructional costs? Educational Leadership, 64(3), 43–46.