JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 36, NO. 10, PP. 1110–1126 (1999)

Estimating the Impact of Instructional Practices on Student Achievement in Science

Clare E. Von Secker, Robert W. Lissitz

Department of Educational Measurement, Statistics, and Evaluation, College of Education, University of Maryland, College Park, Maryland 20742

Received 22 June 1998; accepted 6 July 1999

Abstract: This study used a hierarchical linear model (HLM) to estimate direct and indirect effects of instructional practices recommended by the National Science Education Standards on individual achievement. Three pedagogical reforms—namely, providing more opportunities for laboratory inquiry, increasing emphasis on critical thinking, and reducing the amount of teacher-centered instruction—were expected to account for variability in school mean achievement and explain why gender, racial-ethnic status, and socioeconomic status have more influence on achievement of students in some schools than in others. Results suggest that whereas the instructional policies recommended by the authors of the Standards may be associated with higher achievement overall, they are equally likely to have the unintended consequence of contributing to greater achievement gaps among students with different demographic profiles. Theoretical expectations about the impact of instructional practices on academic excellence and equity require further evaluation. © 1999 John Wiley & Sons, Inc. J Res Sci Teach 36: 1110–1126, 1999

Since publication of *A Nation at Risk* (National Commission on Excellence in Education, 1983), national, state, and local agencies have adopted reforms aimed at raising achievement of all students by setting higher standards and by changing the way science is taught. One national reform effort has been publication of the National Science Education Standards, a document that provides, among other things, guidelines for effective science instruction [National Research Council (NRC), 1996]. The Standards call for pedagogical shift from a teacher-centered to a student-centered instructional paradigm. Whereas teacher-centered instructional strategies such as large-group instruction, recitation, drill, and opportunities for controlled independent practice are successful for tasks that demand rote memorization, they have not been shown to be effective for teaching higher-order thinking and problem solving (Anderson, 1997; Darling-Hammond, 1996). The Standards recommend a student-centered instructional environment that engages students in socially interactive scientific inquiry and facilitates lifelong learning.

Experts recognize that teachers are central to the success of science reform efforts (Darling-Hammond, 1996; NRC, 1996; Rutherford & Ahlgren, 1990; Yager, 1992). Changing the way students are taught requires a sustained, long-term commitment from teachers to engage in active, student-centered instructional practices that provide academically challenging experiences for all students regardless of ability, motivation, and academic track. Instruction that emphasizes

---

Correspondence to: C. E. Von Secker

CCC 0022-4308/99/101110-17

inquiry as an essential precursor to scientific understanding is very different from teacher-centered courses and vocabulary-dense texts that are typical of high schools in the 1990s. Student-centered instruction, which is characterized by inquiry and discussion of open-ended questions, is expected to be more effective for promoting deep understanding of science (Anderson, 1997; Darling-Hammond, 1996). However, there is no verification of the extent to which teachers are already using strategies described in the Standards, nor are there accepted estimates of the magnitudes of effects of instructional practices on science achievement.

There are dozens of reports documenting educational inequities among individuals who are members of groups characterized by problems such as poverty and social disadvantage. In the past, an instructional strategy aimed at improving achievement was to group students in separate instructional settings, or tracks, where disadvantaged students could receive more intensive support and supplementary instruction to bridge gaps in student availability to learn (Wang & Reynolds, 1995). Essentially, this strategy translated in practice into disadvantaged students having less opportunity for inquiry and social interaction. Although well intentioned, this effort and others like it sometimes have resulted in greater gaps in achievement (Hoffer, 1992; Kulik & Kulik, 1982; Oakes, 1985, 1990).

The purpose of our study was to provide a baseline evaluation of whether teachers' decisions to implement the specific instructional emphases recommended in the National Science Education Standards are associated with science achievement and equity. That is, we wanted to estimate the direct influence of instruction on science achievement of all students and the interaction of instruction with student characteristics believed to affect science achievement.

We used data from a national study sponsored by the U.S. Department of Education to answer these questions and provide baseline information about associations of teachers' pedagogical emphases with science achievement prior to adoption of the Standards. We employed a research design that uses statistical methods that account for the nested nature of the data and allows evaluation of group effects on individuals. Our analysis explored both the main effects of instruction on mean science achievement of a school and the extent to which instructional practices of a school interact with gender, minority status, and socioeconomic status to affect individual science achievement.

Background

One assumption of the Standards is that changing the way science is taught will result in higher average science achievement for all students and decrease gaps in achievement among more and less advantaged groups of students. The Standards claim that learning is an active process which occurs when students have a chance to construct understanding through empirical investigation and social interaction with others (NRC, 1996). The Standards note that teachers' emphasis on presentation of information and covering science topics is in direct conflict with the goal of having students understand scientific knowledge. Emphasis on individual inquiry and emphasis on group interaction are two specific instructional strategies recommended by the Standards for achieving these aims. This prescription reflects a blend of two theoretical research strands: an empirical constructivist paradigm and a sociocultural one. This integrated view of learning has important pedagogical implications for reforms aimed at achieving academic excellence and equity.

Excellence

The theoretical underpinnings of learner-centered models reflect a constructivist tradition that focuses on individual activity and the importance of exploring physical phenomena as a starting point for personal construction of meaning (Carey, 1985; Carmichael, Driver, Holding, Phillips, Twigger, & Watts, 1990; Piaget, 1970; von Glaserfeld, 1984, 1987). Knowledge is not transmitted directly from one person to another, but actively constructed by the learner. Neurocognitive perspectives support a learner-centered model for science education (Anderson, 1998). Viewing learning primarily as a process of conceptual change is similar to what has been identified traditionally as discovery learning, where individuals make sense of the physical world through development of content-independent logical structures and operations. From this perspective, a teacher's role is (a) to provide opportunities for active investigation, and (b) to serve as a facilitator of student reflection and critical thinking.

Research indicates that students need to learn how to analyze the significance of what they experience and to practice developing arguments that support their position with respect to scientific issues. Students develop critical thinking skills by gathering evidence to support claims, analyzing data quantitatively, and constructing explanations for discrepant events. Empirical experiences that challenge learners to reorganize their personal theories and schema are consistently correlated with more sophisticated understanding of the nature and methods of science (Buzzelli, 1994; Glasson, 1989; Haury & Rillero, 1992; McDavitt, 1994; Morgan, 1991; O'Brien & Peters, 1994; Odubunmi & Belogun, 1991; Piburn, 1994; Pugh, 1991; Roth, 1993, 1994; Saunders, 1992; Shepardson, 1993; Sinclair, 1994; Steinberg, 1987; Thayer, 1990).

A central tenet of the Standards is that students cannot achieve high levels of performance without access to a rich array of learning materials and work spaces that offer hands-on and minds-on experiences. The Standards advise that a more student-directed program, rather than a traditional format of lecture, text, and demonstration, is essential for all students to achieve excellence. One of the expected benefits of student-directed instruction is the group collaboration that is essential for creating a community of learners.

The Standards assert that active learning must involve interaction with others. This sociocultural perspective of learning is supported by findings from a strand of research on social constructivism (Bruner, 1985; Edwards & Mercer, 1987; Johnston & Driver, 1990; Lemke, 1990; Rogoff & Lave, 1984; Scott, Asokko, Driver, & Emberton, 1995; Seely-Brown, Collins, & Duguid, 1989; Vygotsky, 1978). Sociocultural cognitive models portray learning as a process of cultural apprenticeship in which knowledge is constructed as a result of social interaction. Learning science thus involves learning the ideas and practices of the scientific community and making them meaningful at an individual level. Given this view, the teacher has two essential roles: (a) to provide opportunities for individuals to engage socially in talk and activity about shared problems or tasks, and (b) to serve as an expert who mediates social discourse and leads students to conventional science ideas (Driver, Asoko, Leach, Mortimer, & Scott, 1994).

An appropriate starting point toward achieving this goal is instruction that reflects the intellectual and cultural traditions that characterize the nature and practice of science. Novices are introduced to a community of knowledge through discourse in the context of relevant tasks. Students are socialized into a particular community of learners by engaging in discourse and developing a scientific way of thinking characterized by curiosity, precision of measurement, tolerance for ambiguity, use of supporting evidence for making claims, and a balance between open-mindedness and skepticism. Meaningful social interaction occurs within contexts where students have opportunities to make scientific judgments and practice the habits of mind that superintend scientific literacy (American Association for the Advancement of Science, 1989; Nussbaum & Novick, 1982; Rutherford & Ahlgren, 1990).

Equity

Most of the variability in student achievement in science traditionally is attributed to socioeconomic status, race, and gender [Coleman et al., 1966; Gibbons, 1992; Hilton & Lee, 1988;

Mason & Kahle, 1989; National Center for Education Statistics (NCES), 1992]. When groups of students with similar backgrounds are compared, students from families with high socioeconomic status (SES) outperform students from low-SES families; Asian and White students have greater achievement gains during high school than Black or Hispanic students, and males outperform females (Hoffer, Rasinski, & Moore, 1995; Madigan, 1997).

Providing all students with the opportunity to engage in authentic science inquiry is a tenet of the Standards and a key issue in reform efforts aimed at promoting academic equity. There are significant inequities in opportunities for access to laboratory experiences. Inadequate facilities and equipment and lack of money to purchase consumable supplies may create larger gaps among advantaged and disadvantaged students because students in disadvantaged schools generally have fewer opportunities for scientific inquiry (Lee & Burkham, 1996). Emphasis on laboratory inquiry and social interaction may be especially important for improving achievement of minority females. Evidence suggests that gender gaps in achievement can be reduced when girls' science classes have a laboratory component (Lee & Burkam, 1996; Lee, Chen, & Smerdon, 1996).

One of the criticisms of the Standards is that, although they emphasize the importance of promoting science achievement for all students regardless of demographic status, the proposed science education reforms do not directly address theoretical issues surrounding ethnic, socioeconomic, and gender equity (Rodriguez, 1997). Although there is some evidence that instructional emphases and curricular choices explain discrepancies in student achievement (McCauley, 1995; NCES, 1992), there is no empirical evidence to support the viability and utility of proposed science reforms for creating more equitable opportunities for students nationwide (Donmoyer, 1995; Riechard, 1994). Thus far, the goal of quality education for all students has not become a reality.

Research Questions

1. Does science achievement vary systematically among schools? We hypothesize that there is significant systematic between-school variability in science achievement that is most appropriately evaluated using a multilevel model that accounts for the common experiences of students who attend the same school.
2. To what extent do gender, minority status, and SES account for differences in science achievement? We expect demographic status to be strongly predictive of individual science achievement. However, we presume that school mean achievement will vary significantly even when the demographic composition of schools is adjusted to be the same.
3. Does instruction affect the average achievement of students within the same school? We believe that Standards-based instructional practices will be associated with significantly higher science achievement, and that disparities in teachers' instructional emphasis will explain variability in mean achievement of students who attend different schools.
4. Does instruction interact with demographic context to influence science achievement indirectly? To close achievement gaps as well as promote excellence for all, instruction should reduce inequities associated with gender, minority status, and SES. We theorize that the interaction of instructional practices with student characteristics will explain why some schools are more equitable than others.

Methodology

Sample and Data

This study used data from the 1990 High School Effectiveness Study (HSES), collected as part of the second wave of the National Education Longitudinal Study (NELS:90) sponsored by

Table 1
Comparison of respondents in HSES sample and study sample

	N = 7,642			N = 2,018		
Variable	n^*	Mean[†]	*SD*	n	Mean	*SD*
Science achievement	7,191	47.97	10.20	2,018	48.54	10.41
Percent male	7,642	0.52	0.50	2,018	0.54	0.50
Percent minority	7,503	0.49	0.50	2,018	0.45	0.50
Socioeconomic status	7,082	−0.01	0.84	2,018	0.01	0.82

[*] Number of cases is unweighted.
[†] Means ansd standard deviations are weighted by the normalized student weight.

the NCES, U.S. Department of Education. HSES was designed to support investigation of school effects issues including direct and indirect associations of instructional practices with student achievement. The total sample includes 7,642 students representing 790,810 tenth-grade students enrolled in 247 urban and suburban schools in the 30 largest metropolitan school districts. Within each school, science teachers and one administrator completed questionnaires that provided information about classroom instruction and school demographics. Detailed information about the design and analysis of the study is available in the **High School Effectiveness Study (HSES): Data File User's Manual** (Scott, Ingels, Pulliam, Sehra, Taylor, & Jergovic, 1995).

One of the advantages of using the HSES data for analysis of school effects is that sample weights are provided for student-level and school-level data. Weighting survey data is necessary to compensate for unequal probabilities of selection and adjust for the effects of nonresponse. Parameters in the hierarchical model were weighted by the cross-sectional student and school weights, S1STSCWT and S1SCWT1, respectively. An explanation of the derivation and appropriate use of sample weights is available in the **HSES Data File User's Manual** (Scott et al., 1996).

We selected for analysis a sample of 2,018 tenth-grade students in 163 schools based on the availability of four measures: (a) science achievement data; (b) student demographic data (i.e., gender, minority status, and SES); (c) science teacher questionnaire data; and (d) at least four students per school. Responses of teachers within each school were aggregated because there were insufficient data to support a third level of analysis (students within classrooms within schools). As shown in Table 1, students who were filtered from our sample are slightly more likely to be low achievers, females, minorities, and from families with low SES. The most likely consequence of this is underestimation of the effects of instruction on disadvantaged students. However, the means and standard deviations (**SD**s) in the two samples are close enough that the amount of bias, if any, should be small.

Measures

Tenth-Grade Science Achievement. Science questions were drawn from the fields of biology, earth science, physics, and chemistry. The tests were developed by the Educational Testing Service (ETS) with the express purpose of measuring higher-order thinking as well as understanding of fundamental concepts and mastery of basic skills. This test score constitutes the dependent variable in our analysis. Reports on the psychometric properties of the cognitive test may be obtained from NCES.

Student Characteristics. Our model included three demographic characteristics known to influence individual science achievement—namely gender, minority status, and SES. Gender (male) is represented by a dichotomous variable; values of 0 or 1 are assigned to females or males, respectively. Minority status (minority) is a dichotomous variable for which students in traditionally high-achieving racial-ethnic groups (i.e., White and Asian students) have a value of 0 and traditionally low-achieving minorities (i.e., Blacks, Hispanics, and Native Americans) have a value of 1. SES is a standardized composite variable prepared by NCES.

Student-level variables in the hierarchical model were centered around the group mean for their schools and allowed to vary nonrandomly. This means that gender reflects the average gap in achievement among males and females in the same school. Minority can be interpreted as the average difference in achievement between minority and majority students in the same school. SES is the average expected change in science achievement for every 1 **SD** change from average school SES.

Instructional Variables. The teacher questionnaire provides descriptive information about the classroom environments of the schools in this study. We created three composite variables that reflected instructional emphases recommended by the Standards. The first, teacher-directed, measures the extent to which teachers focus on teacher-centered instruction. The second, critical thinking, is a composite that reflects the amount of emphasis teachers place on habits of mind associated with scientific literacy, such as student understanding and application of scientific knowledge, ideas, and inquiry processes, and emphasis on scientific writing and advanced study. Laboratory emphasis is a composite variable that indicates how often teachers provide direct opportunities for science laboratory experiences where students engage in the active and extended scientific inquiry that is expected to foster scientific literacy.

All of the instructional variables used in the hierarchical models were standardized to have unweighted means of 0 and **SD**s of 1. Instructional variables in the hierarchical model were centered around their grand means. The level-two gamma coefficients are the expected changes in the intercept or slope coefficients associated with every 1 **SD** change from the average amount of a given instructional emphasis. A description of the results of principal components analysis used to create the composites is provided in Table 2.

Analytical Model

We selected a hierarchical linear model (HLM) as the statistical methodology for estimating the influence of instructional practices on science achievement. This method takes into account the nested nature of the data (i.e., students within schools). Conventional techniques based on ordinary least squares regression tend to systematically underestimate school effects (Kreft & Yoon, 1994; Prince & Taylor, 1995; Zigarelli, 1996). One reason for this is that regression analyses fail to decompose variance into that part that is unique to individuals (e.g., gender and racial-ethnic status) and variance between schools that can be attributed to school policies and programs.

The chief advantage of using HLM in this evaluation is that the statistical software produces design effect-corrected standard errors. This is particularly important since the sample design involves stratification and multistage probability sampling in which individuals are sampled with different probabilities of inclusion. Thus, standard errors of estimates must be adjusted to account for the complex nature of the survey design. HLM weights gamma coefficients by the least squares estimators in which each case is weighted by the inverse of its precision. The

Table 2
Principal components analysis used to construct instructional composites

Instructional composite	Item ID*	Loading†	Reliability
Emphasis on critical thinking			$\alpha = .80$
Emphasis on scientific methods	F2T19C	.781	
Emphasis on further study in science	F2T19D	.733	
Emphasis on problem-solving	F2T19E	.826	
Emphasis on observation skills	F2T19H	.741	
Emphasis on scientific writing	F2T19J	.672	
Emphasis on laboratory inquiry			$\alpha = .78$
How often experiments are done	F3T20C	.673	
How often reports on experiments written	F3T20E	.744	
Time spent conducting lab periods	F3T16G	.736	
Emphasis on lab techniques	F2T19F	.752	
Use of working in small groups	F3T18F	.629	
Emphasis on teacher-directed instruction			$\alpha = .73$
Time spent instructing small groups	F3T16B	.736	
Time spent instructing individuals	F3T16C	.877	

*Teacher survey items were standardized to have mean 0 and *SD* 1.
†Instructional composites were computed using standardized factor scores.

precision-weighted parameter variance depends on the weighted proportion of the cases and sample size. Failure to account for a design effect could inflate Type I error rates.

The HLM provides a means of explaining how variance in some outcome (e.g., a student's achievement score) is a function of both individual differences (Level 1) and differences in the contexts in which learning occurs (Level 2). HLM identifies both intercept and slope heterogeneity which enables researchers to investigate direct effects of school-level factors on average school achievement (the intercept) and indirect effects of school-level factors on individual characteristics that affect achievement (slope coefficients). Readers interested in a more complete description of HLM are directed to Bryk and Raudenbush (1992).

The unconditional Level 1 model uses a series of regression equations (one per school) to predict individual achievement. Within-group regression equations vary across schools as a function of mean group achievement scores (β_{0j}) and the relative strength of the effects of student-level variables (β_{pj}). In our model, the general Level 1 equation for predicting the achievement of student **i** in school **j** ($\mathbf{Y}_{ij}$) is:

$$\mathbf{Y}_{ij} = \beta_{0j} + \beta_{1j}(\text{male}) + \beta_{2j}(\text{minority}) + \beta_{3j}(\text{SES}) + \mathbf{r}_{ij.}$$

This model is essentially a random effects analysis of covariance (ANCOVA) in which three nonrandomly varying (fixed) predictors—male, minority, and SES—are centered on their school means. The model parameters can be interpreted as follows: β_{0j} = the mean achievement of School **j**; β_{1j} = the average gender gap in achievement across all schools; β_{2j} = the average minority gap in achievement across all schools; β_{3j} = the average gap in achievement among students in the same school who vary by 1 **SD** in SES; and $\mathbf{r}_{ij}$ = the deviation of the achievement of person **i** from mean achievement of all persons in school **j** when gender, minority status, and SES are controlled.

A conditional level-two model explains systematic variance in school excellence, equity, or both. The randomly varying intercept and nonrandomly varying slope coefficients are used as dependent variables in the Level 2 model, in which between-group regression equations use school instructional characteristics as independent variables to explain intercept and slope coefficient differences among schools. By partitioning variance into individual, group, and cross-level interactions, HLM can isolate the variance of school-level variables (e.g., instruction) when variance due to student characteristics is controlled.

In our conditional Level 2 model, predictors representing the extent to which teachers in school **j** report emphasizing teacher-centered instruction (W_{1j}), critical thinking (W_{2j}), or laboratory inquiry (W_{3j}) are used to explain academic excellence, measured by differences in mean school achievement (β_{0j}), and why the effects of gender (β_{1j}), minority status (β_{2j}), and SES (β_{3j}) are more equitable (i.e., have less influence on individual in achievement) in some schools than they do in others. The four between-schools equations are: $\beta_{0j} = \gamma_{00} + \gamma_{01}(W_{1j}) + \gamma_{02}(W_{2j}) + \gamma_{03}(W_{3j}) + \mu_{0j}$; $\beta_{1j} = \gamma_{10} + \gamma_{11}(W_{1j}) + \gamma_{12}(W_{2j}) + \gamma_{13}(W_{3j})$; $\beta_{2j} = \gamma_{20} + \gamma_{21}(W_{1j}) + \gamma_{22}(W_{2j}) + \gamma_{23}(W_{3j})$; and $\beta_{3j} = \gamma_{30} + \gamma_{31}(W_{1j}) + \gamma_{32}(W_{2j}) + \gamma_{33}(W_{3j})$. Since level-two predictors for each school (W_{pj}) were centered on their grand means, the Level 2 parameters can be interpreted as follows: γ_{00} = the grand mean achievement of all schools; γ_{0p} = the change in β_{0j} that can be expected by 1 **SD** difference from the average amount of an instructional practice, W_p (i.e., emphasis on teacher-centered instruction, critical thinking, or laboratory inquiry); μ_{0j} = the deviation of the mean achievement of school **j** from grand mean achievement when instruction is controlled; $\gamma10$ = the grand mean gender gap; γ_{1p} = the average change in the gender gap associated with 1 **SD** change in W_p; γ_{20} = the grand mean minority gap; γ_{2p} = the average change in the minority gap associated with 1 SD change in W_p; γ_{30} = the grand mean SES gap; and γ_{3p} = the average change in the SES gap associated with 1 **SD** change in W_p.

Results

Unconditional HLM

The first step in our analysis was to use a fully unconditional two-level HLM to partition the variance in science achievement into its between-school and within-school components. The result, shown in Table 3, can be interpreted essentially the same way as a one-way analysis of variance (ANOVA) with random effects. The adjusted intraclass correlation (ICC) for between-school variability indicates that about 45% of the variance in science achievement was due to systematic differences in the average achievement of students attending different schools. This

Table 3
Fully unconditional HLM for partitioning variance in science achievement

Between school variability (τ)	27.619
Within-school variability pooled across all students (σ^2)	53.169
Reliability of intercept β_0, (λ)	0.624
Proportion of variability between schools (intraclass correlation)*	0.342
Proportion of between-school variability, adjusted for reliability†	0.454

*The intraclass correlation, ICC, is calculated using the formula: $\tau/(\tau + \sigma^2)$.

†The adjusted ICC is calculated using the formula $\tau/(\tau + \sigma^2 \times \lambda)$.

outcome provides justification for using a hierarchical model to evaluate instructional effects. Ignoring this high ICC and analyzing disaggregated data could lead to underestimation of standard errors and inflation of Type I error rates (Burstein, 1980; Kromrey & Dickinson, 1996).

Information about direct and indirect associations of student demographic and school instructional variables with student achievement is displayed in Tables 4 and 5, in which results of conventional statistical testing for determining the significance of effects are reported. However, because of the large sample size, variables can achieve statistical significance even when they are of little substantive significance. For this reason, results of HLM are described in terms of effect sizes (ES), shown in Tables 4 and 5. Effect sizes are **SD** units that allow comparison of outcomes with different metrics. They can be interpreted as the **SD** change in the dependent variable associated with 1 **SD** change in an independent variable. Effect sizes of 0.5 **SD** or more are considered large; effect sizes between 0.3 and 0.5 **SD** are moderate; effect sizes between 0.1 and 0.3 **SD** are small; and effect sizes <0.1 **SD** are trivial (Rosenthal & Rosnow, 1984).

Unconditional Within-School HLM

The unconditional within-school model estimates individual science achievement as a function of the mean achievement of students in his or her school (γ_{00}), gender (γ_{10}), minority status (γ_{20}), and SES (γ_{30}). School mean achievement is modeled as a random parameter and demographic variables included as controls are fixed. Constraining between-school variances of the slope coefficients to zero allowed us to include many more schools with small numbers of students per school than would have been possible if we allowed slope coefficients to vary randomly.

Within-school effect sizes were computed by dividing each HLM gamma by the pooled within-group **SD** in the fully unconditional model (Table 3). These effect sizes represent changes in average science achievement when other variables are controlled. Thus, average science achievement of males is 0.288 **SD** higher than that of females, average science achievement of minorities is 0.578 **SD** lower than that of majority students, and average science achievement is expected to increase by 0.441 **SD** for every 1 **SD** increase in SES. The influence of demo-

Table 4

Unconditional within-school HLM for estimating effects of demographic status on science achievement

Independent variables		Beta Coefficient	Standard Error	*p*-value	Effect Size*
Mean achievement	β_0	49.194	0.520	0.000	—
Gender gap	β_1	2.100	0.407	0.000	.288
Minority gap	β_2	−4.216	0.570	0.000	−.578
Socioeconomic status	β_3	3.217	0.275	0.000	.441

Random parameter	Parameter variance	Degrees of freedom	Chi-square statistic
Mean achievement†	28.65780	162	1636.81932‡
σ^2	46.92259		

*Effect sizes are computed by dividing each beta coefficient by 7.292, which is the pooled within-school standard deviation. This value is computed by taking the square root of σ^2 from Table 3.

†In this random intercept model, only the intercept term is allowed to vary randomly across schools. The within-school independent variables are fixed.

‡$p \leq .000$.

graphic status is more pronounced when these effects are combined. For example, the average science achievement of minority females is expected to be more than 0.8 **SD** less than that of White males (0.288 + 0.578), even when the SES of the groups is the same.

Conditional Between-School HLM

The results shown in Table 5 summarize information about the direct and indirect impact of instruction on individual science achievement. The between-school model is used to determine whether instruction affects the average achievement of students within the same school and whether instruction interacts with demographic factors to influence science achievement indirectly.

The random intercept portion of our model suggested that two variables—emphasis on teacher-centered instruction and emphasis on laboratory inquiry—have moderate direct associations with science achievement. Because school-level variables were grand-mean centered (values of 0 represent the average emphasis across all schools), effect sizes can be interpreted as the change in mean science achievement per 1 **SD** change from average instructional emphasis. Teacher-centered instruction is negatively associated with achievement; mean science

Table 5

Conditional between-school HLM for estimating direct and indirect effects of instruction on science achievement

Independent variables		Gamma Coefficient	Standard Error	*p*-value	Effect Size
Mean achievement*	β_0				
Teacher-centered instruction	γ_{01}	−2.482	0.460	0.000	−0.472
Emphasis on critical thinking	γ_{02}	.311	0.514	0.545	0.059
Emphasis on laboratory inquiry	γ_{03}	2.040	0.511	0.000	0.388
Gender gap[†]	β_1				
Teacher-centered instruction	γ_{11}	−1.914	0.466	0.000	−0.368
Emphasis on critical thinking	γ_{12}	1.675	0.509	0.001	0.322
Emphasis on laboratory inquiry	γ_{13}	−1.093	0.520	0.035	−0.210
Minority gap[†]	β_2				
Teacher-centered instruction	γ_{21}	−0.118	0.708	0.868	−0.016
Emphasis on critical thinking	γ_{22}	−2.201	0.760	0.004	−0.303
Emphasis on laboratory inquiry	γ_{23}	0.944	0.761	0.215	0.130
Socioeconomic status[†]	β_3				
Teacher-centered instruction	γ_{31}	.843	0.369	0.020	−.240
Emphasis on critical thinking	γ_{32}	−.161	0.357	0.652	−.046
Emphasis on laboratory inquiry	γ_{33}	−.612	0.353	0.082	−.174

Random parameter	Parameter variance	Degrees of freedom	Chi-square statistic
Mean achievement	20.44658	159	1211.36768[‡]
σ^2	46.91647		

*Effect sizes of gamma coefficients for randomly varying coefficients (the intercept) are computed by dividing the gamma coefficient by 5.255, the standard deviation of the intercept of the within-school model.

[†] When beta coefficients are fixed, effect size is calculated by dividing the gamma coefficient in the conditional model by the standard deviation of the slope of the beta coefficient in the conditional model, calculated by multiplying the standard error of the beta in the within-school model by the square root of the sample size (*N* = 163 schools).

[‡] $p \leq .000$.

achievement is almost 0.5 **SD** lower in schools where the emphasis teachers place on teacher-centered instruction is 1 **SD** above average. On the other hand, mean science achievement is expected to increase by almost 0.4 **SD** for every 1 **SD** increase in the amount of emphasis placed on laboratory inquiry. The effect sizes represent associations when other variables are controlled. Thus, in a school where emphasis on laboratory inquiry is 1 **SD** above average and emphasis on teacher-centered instruction is 1 **SD** below average, average science achievement is expected to be 0.86 **SD** $[(0.472 \times 1) - (.388 \times -1)]$ higher than other schools.

The "slopes as outcomes" (Burstein, 1980) portion of our model examined indirect effects of instruction on achievement. Being male and coming from high-SES families were positively associated with science achievement. Being a member of a minority group was negatively associated with science achievement. Table 5 illustrates how instruction interacts with gender, minority status, and SES to alter the directions and magnitudes of these effects.

Although average science achievement of males was 0.288 **SD** higher than that of females, the instructional emphasis of a school can significantly affect estimates of gender equity. The negative association of teacher-centered instruction with gender ($\text{ES} = -0.368$) favored female achievement. When demographically equivalent groups of boys and girls were exposed to a 1 **SD** increase in the amount of teacher-centered instruction, the influence of gender on achievement was reduced to -0.1 **SD** $(0.288 - 0.368 \times 1)$. On average, science achievement of boys benefited more from the effects of emphasis on critical thinking ($\text{ES} = 0.322$). For every 1 **SD** increase in emphasis on critical thinking, the gender gap was expected to increase to 0.61 **SD** $(0.288 + 0.322 \times 1)$. On average, emphasizing laboratory inquiry was expected to promote gender equity ($\text{ES} = -0.210$). A 1 **SD** increase in the amount of laboratory emphasis reduced the average gap in performance to 0.078 $(0.288 - 0.210 \times 1)$, a trivial association.

The instructional strategies investigated in this model suggest that instruction may widen rather than close the gap in performance between minority and majority students ($\text{ES} = -0.578$). In schools where the amount of emphasis on critical thinking was 1 **SD** above average, minority students were expected to perform 0.88 **SD** $(-0.594 + -0.303 \times 1)$ lower than majority students. Emphasizing laboratory inquiry had a small equitable effect ($\text{ES} = 0.13$), but even when the amount of laboratory inquiry was the same, predicted average science achievement of minority students was still 0.448 **SD** less $(-0.594 + 0.13 \times 1)$ than that of majority students.

Two instructional strategies—emphasis on laboratory inquiry and teacher-centered instruction—significantly affected the strength of the association of SES with science achievement. When gender and minority status were controlled, science achievement was expected to increase 0.441 **SD** for students whose SES was 1 **SD** above the average school SES. The expected influence of SES was reduced by 40% for students in schools where emphasis on laboratory inquiry was 1 **SD** above average $(0.441 - 0.174 \times 1)$ and by 45% in schools where the amount of teacher-centered instruction was 1 **SD** below average $(0.441 + 0.240 \times -1)$. On the other hand, the inequitable influence of SES on science achievement was predicted to be almost twice as high in schools where emphasis on laboratory inquiry was 1 **SD** below and teacher-centered instruction was 1 **SD** above average $(0.441 + 0.240 \times 1 -0.174 \times -1)$.

Discussion

Statistical Analysis of Large-Scale Databases

Large-scale national databases such as HSES provide researchers with opportunities to investigate questions about the impact and implications of science policy and practice. The framers of the HSES teacher survey included items specifically designed to measure widely held epistemological beliefs about learning. These beliefs are consistent with theoretical underpinnings of the Standards. Even though data used in this study were collected 6 years prior to publication of the Standards, they provide compelling empirical evidence about associations of instructional strategies with science achievement and equity. Keeping these advantages in mind, we caution researchers to be aware of limitations of this database and others like it (e.g., National Education Longitudinal Study, National Assessment of Educational Progress).

The complexity of our model was limited by small within-classroom sample size. Because the average number of students within the 163 schools in our study was only 12, we used a two-level hierarchical linear model (i.e., students nested within schools). The reliability of estimates of school effects will be influenced by the extent to which teachers within a school are homogeneous with regard to their teaching practices. If between-teacher differences are substantial once student differences are controlled, a three-level model analyzing the impact of instruction on students nested within classrooms nested within schools could provide a better evaluation of the impact of instruction on science achievement and equity. Cross-validation of our findings using comparable data available from other large-scale national surveys (e.g., the National Assessment of Educational Progress) could provide a more complete picture of associations of science achievement with instructional strategies recommended in the Standards.

Another limitation was invariance in teacher self-reports of their instructional practices. An advantage of using this database was that the HSES teacher survey measured the extent to which teachers emphasize instructional practices consistent with those recommended by the Standards. However, preliminary examination of a number of individual items of interest revealed no significant differences among teachers. We were restricted to estimating the impact of instructional practices for which there were systematic differences between schools. Development of finer-grained measures of instructional practices would be useful for determining whether the lack of difference among teachers on emphasis of some types of instruction is real or a result of measurement error that might be explained by differences in teachers' conceptualizations of constructs of interest.

Impact of Instructional Practices

Because the literature on science education reform is so consistent about what instructional characteristics are associated with student achievement, we expected all of the instructional composite variables in our model to explain significant differences in student achievement in science. That was not the case. Nor did we find unwavering evidence to bolster claims that proposed instructional changes would foster equity: quite the contrary. For example, the combined effect of replacing an emphasis on teacher-centered instruction with one that focuses on developing critical thinking skills is expected to be associated with even greater gaps in achievement between boys and girls. We conclude that some widely accepted beliefs about the strength of associations of instruction with science achievement need further examination.

The strongest empirical support for instructional recommendations set forth in the Standards was observed for instruction that emphasized laboratory inquiry. This practice was invariably associated with higher achievement overall and with more equitable achievement among students with different demographic profiles. It may also explain why, in disadvantaged schools where teachers are more likely to lack resources and training to support this type of instructional practice, students are at greater risk of low achievement. Our results suggest that the Standards are more likely to promote equity if they are supported by national, state, and local efforts to provide equal opportunities for access to laboratory facilities, equipment, and supplies.

We found no evidence that emphasizing critical thinking is associated with significant differences in mean school achievement. We suspect this result is more a consequence of measurement error that fails to reveal systematic differences between schools (a limitation described above), rather than a true lack of an instructional impact on average achievement. Further study is necessary to verify this suspicion. We did find significant indirect effects of critical thinking on achievement as a result of interaction with student gender and minority status. Contrary to its intended outcome of more equitable distribution of achievement, emphasis on critical thinking was associated with a magnification of gender and minority gaps. This finding suggests that on average, females and minorities are more at risk for low achievement in schools where teachers are encouraged to embrace this instructional practice.

De-emphasizing traditional teacher-centered instruction is expected to increase average science achievement and minimize gaps in achievement between individuals of different socioeconomic status. However, we use this finding to remind the reader about the dangers of drawing causal inference from HLM results. Teacher-centered instruction does not cause inequity in achievement associated with SES, and multiple explanations for this association are reasonable. For example, examination of the student data file reveals that teacher-centered instruction is approximately 0.3 **SD** higher for low-SES than for high-SES students. Perhaps low-SES students are also more likely to be low achievers who are not yet able to work independently on complex tasks and actually benefit from cognitive scaffolding provided by more structured, teacher-centered environments. Replacing teacher-centered practices with more student-centered instruction will not necessarily result in more learning unless students have the basic vocabulary and scientific understanding essential for engaging in meaningful individual investigation and collaborative discussion.

Our findings suggest that instruction matters. School excellence and equity can be positively or negatively affected by the way science is taught. Although our estimates of the impact of instruction on science achievement are provocative, they need to be interpreted against a referential backdrop that depicts the total school setting. More sensitive measures of instructional practice and more contextual data about classroom dynamics are essential to evaluate more accurately the impact of instruction on science achievement and equity. The impact of specific instructional practices on student achievement remains an open question, particularly with regard to promoting equity for student populations at risk of low achievement.

Students from families of low SES, minority students, and females are more at risk of low achievement. However, conceptualizing equity in terms of independent variables that distinguish populations on the basis of broad demographic characteristics reduces the power of models to reveal small but significant cumulative effects that explain individual differences in science achievement. After all, individuals are not at risk of low science achievement because they are poor, female, or minorities. Rather, individuals with low SES, minorities, and females are members of highly variable populations with heightened probabilities of an undesirable achievement outcomes. Understanding the gap in science achievement and the processes that mediate it may be informed by taking into account different mechanisms involved in person–environment interactions and individual differences in perception of educational situations. Future analysis examining the relative impact of instruction on achievement of individuals within unique social contexts (e.g., low-SES minority females) could provide further insight into why science reforms have not always achieved desired results, and offer guidance for future policy and practice.

The consequences of changing the way students are taught are not self-evident. By 1990, teachers reported that they were using student-centered strategies that were recommended in the Standards 6 years later, with mixed results. Despite strong theoretical grounding, support for a Standards-based science program is still largely anecdotal. It seems reasonable to us that assessment of the effectiveness of instructional practices promoted by the authors of the Standards, however well informed conceptually, should be evaluated using the same rigorous evidentiary criteria that they recommend students apply to scientific inquiry in the classroom.

References

American Association for the Advancement of Science. (1989). Project 2061: Science for all Americans. Washington, DC: AAAS.

Anderson, O.R. (1997). A neurocognitive perspective on current learning theory and science instructional strategies. Science Education, 81, 67–89.

Bruner, J. (1985). Vygotsky: A historical and conceptual perspective. In J. Wertsh (Ed.), Culture, communication and cognition: Vygotskian perspectives (pp. 21–24). Cambridge, England: Cambridge University Press.

Bryk, A.S., & Raudenbush, S.W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. In D. Berliner (Ed.), Review of research in education. Washington, DC: American Educational Research Association.

Buzzelli, A.R. (1994). Using problem-based learning with large groups. Journal of Optometric Education, 19, 56–60.

Carey, S. (1985). Conceptual change in childhood. Cambridge, MA: MIT Press.

Carmichael, P., Driver, R., Holding, B., Phillips, I., Twigger, D., & Watts, M. (1990). Research on students' conceptions in science: A bibliography. Centre for Studies in Science and Mathematics Education, University of Leeds, United Kingdom.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). Equality of educational opportunity. Washington, D.C.: U.S. Government Printing Office.

Darling-Hammond, L. (1996). The right to learn and the advancement of teaching: Research, policy, and practice for democratic education. Educational Researcher, 25, 5–17.

Donmeyer, R. (1995) Rhetoric and reality of systemic reform: A critique of the proposed national science education standards. Cognosos, 4, 1–3, 7–8.

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. Educational Researcher, 23, 5–12.

Edwards, D., & Mercer, N. (1987). Common knowledge. London: Methuen.

Gibbons, A. (1992). Growing scientists for the 21st century. Science, 258, 1195.

Glasson, G.E. (1989). The effects of hands-on and teacher demonstration laboratory methods on science achievement in relation to reasoning ability and prior knowledge. Journal of Research in Science Teaching, 26, 121–131.

Haury, D.L., & Rillero, P. (1992). Hands-on approaches to science teaching: Questions and answers from the field and research. (Eric Document Reproduction Service No ED 349 185)

Hilton, T.L., & Lee, V.E. (1988). Student interest and persistence in science. Journal of Higher Education, 59, 510–526.

Hoffer, T.B. (1992). Middle school ability grouping and student achievement in science and mathematics. Educational Evaluation and Policy Analysis, 14, 205–227.

Hoffer, T.B., Rasinski, K.A., & Moore, W. (1995). Social background differences in high school mathematics and science course-taking and achievement. Washington, D.C.: United States Department of Education.

Johnston, K., & Driver, R. (1990). A constructivist approach to the teaching of the particulate theory of matter: A report on a scheme in action. Centre for Studies in Science and Mathematics Education, University of Leeds, United Kingdom.

Kreft, I.G.G., & Yoon, B. (1994). Are multilevel techniques necessary? An attempt at demystification. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. 371 033)

Kromrey, J.D., & Dickinson, W.B. (1996). Detecting unit of analysis problems in nested designs: Statistical power and type I error rates of the "F" test for groups-within-treatments effects. Educational and Psychological Measurement, 56, 215–231.

Kulik, C.C., & Kulik, J.A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. American Education Research Journal, 19, 415–428.

Lee, V.E., & Burkham, D.T. (1996). Gender differences in middle-grade science achievement: Subject domain, ability level, and course emphasis. Science Edition, February 1966.

Lee, V.E., Chen, X., & Smerdon, B.A. (1996). The influence of school climate on gender differences in the achievement and engagement of young adolescents. Washington, D.C.: American Association of University Women.

Lemke, J. (1990). Talking science. Norwood, NJ: Ablex.

Madigan, T. (1997). Science proficiency and course taking in high school: The relationship of course-taking patterns to increases in science proficiency between 8th and 12th grades. Washington, D.C.: U.S. Department of Education.

Mason, C.L., & Kahle, J.B. (1989). Student attitudes toward science and sciencerelated careers: A program designed to promote a stimulating gender-free learning environment. Journal of Research in Science Teaching, 26, 2539.

McCauley, K.J. (1995). A restructured biology classroom. Teaching and Change, 2, 152–168.

McDavitt, D.S. (1994). Teaching for understanding: Attaining higher order learning and increased achievement through experiential instruction. (Research/Tech. Rep., REI, ERIC Document Reproduction Service No. ED 374 093)

Morgan, D. (1991). New methods teach science by observation, hypothesis. Scientist, 5, 1, 8.

National Center for Educational Statistics. (1992). National educational study of 1988: A profile of American eighth-grade mathematics and science instruction. Statistical Analysis Report, NCECS 92-486. Washington, D.C.: United States Department of Education.

National Commission on Excellence in Education. (1983). A nation at risk. Washington, D.C.: U.S. Department of Education.

National Research Council. (1996). National science education standards. Washington, D.C.: National Academy Press.

Nussbaum, J., & Novick, S. (1982). Alternative frameworks, conceptual conflict and accommodation. Instructional Science, 11, 183–208.

Oakes, J. (1985). Keeping track: How schools structure inequality. New Haven, CT: Yale University Press.

Oakes, J. (1990). Multiplying inequalities: The effects of race, social class and tracking on opportunities to learn mathematics and science. Santa Monica, CA: RAND Corporation.

O'Brien, G., & Peters, J. (1994). Effect of four instructional strategies on integrated science process skill achievement of preservice elementary teachers having different cognitive levels. Journal of Elementary Science Education, 6, 30–45.

Odubunmi, O., & Belogun, T.A. (1991). The effect of laboratory and lecture teaching methods on cognitive achievement in integrated science. Journal of Research in Science Teaching, 28, 213–234.

Piaget, J. (1970). Genetic epistemology. New York: Columbia University Press.

Piburn, M. (1994). Three hypotheses about domain specific background knowledge and achievement in science. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Anaheim, CA, March 27–31. (ERIC Document Reproduction Service No. ED 368 570)

Prince, S.B., & Taylor, R.G. (1995). Should the correlates of effective schools be used as a prescription for improving achievement? Educational Research Quarterly, 18, 9–26.

Pugh, M.B. (1991). The effects of teaching problem solving on academic performance and retention. Community/Junior College Quarterly of Research and Practice, 15, 339–349.

Riechard, D.E. (1994). National science education standards: Around the reform bush . . . again? Clearing House, 67, 135–136.

Rodriguez, Alberto J. (1997). The dangerous discourse of invisibility: A critique of the National Research Council's national science education standards. Journal of Research in Science Teaching, 34, 19–37.

Rogoff., B., & Lave, J. (1984). Everyday cognition: Its development in social context. Cambridge, MA: Harvard University Press.

Rosenthal, R., & Rosnow, R.L. (1984). Essentials of behavioral research: Methods and data analysis. New York: McGraw-Hill.

Roth, W.M. (1993). The development of science process skills in authentic contexts. Journal of Research in Science Teaching, 30, 127–152.

Roth, W. (1994). Experimenting in a constructivist high school physics laboratory. Journal of Research in Science Teaching, 31, 197–223.

Rutherford, F.J., & Ahlgren, A. (1990). Science for all Americans. New York: Oxford University Press.

Saunders, W.L. (1992). The constructivist perspective: implications and teaching strategies for science. School Science and Mathematics, 92, 138–141.

Scott, L.A., Ingels, S.J., Pulliam, P., Sehra, S., Taylor, J.R., & Jergovic, D. (1995). High School Effectiveness Study: Data file user's manual. Washington, D.C.: U. S. Department of Education.

Scott. P., Asoko, H., Driver, R., & Emberton, J. (1994). Working from children's ideas: An analysis of constructivist teaching in the context of a chemistry topic. In P. Fensham, R. Gunstone, & R. White (Eds.), The content of science. London: Falmer.

Seely-Brown, J., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. Educational Researcher, 18, 32–42.

Shepardson, D.P., & Pizzini, E.L. (1993). A comparison of student perceptions of science activities within three instructional approaches. School Science and Mathematics, 93, 127–131.

Sinclair, A.S. (1994). Prediction making as an instructional strategy: Implications of teacher effects on learning, attitude toward science, and classroom participation. Journal of Research and Development in Education, 27, 153–161.

Steinberg, E.R. (1987). Effects of a laboratory design course on strategies for troubleshooting logic circuits. Engineering Education, 78, 191–193.

Thayer, R.J. (1990). Education for the 21st century: An elementary school perspective. Engineering Education, 80, 1023–1025.

von Glaserfeld, E. (1984). An introduction to racical constructivism. In P. Watzlawick (Ed.), The invented reality. New York: Norton.

von Glaserfeld, E. (1987). Learning as a constructive activity. In C. Janvier (Ed.), Problems of representation in the teaching and learning of mathematics. Hillsdale, NJ: Erlbaum.

Vygotsky, L.S. (1978). Mind and society: The development of higher psychological processes. Cambridge: Harvard University Press.

Wang, M.C., & Reynolds, M.C. (Eds.). (1995). Making a difference for students at risk: Trends and alternatives. Thousand Oaks, CA: Corwin.

Yager, R.E. (1992). What we did not learn from the 60s about science curriculum reform. Journal of Research in Science Teaching, 29, 905–910.

Zigarelli, M.A. (1996). An empirical test of conclusions from effective schools research. The Journal of Educational Research, 90, 103–110.