Science
Education

CIENCE ED CA ION POLIC

Testing Predictors of Instructional Practice in Elementary Science Education: The Significant Role of Accountability

KATHRYN N. HAYES,[1] CARY J. TREXLER[2]
[1]Department of Educational Leadership, California State University, East Bay, Hayward, CA 94542-3004;[2]School of Education and College of Agricultural and Environmental Sciences, University of California, Davis, CA 95816, USA

Received 22 November 2014; revised 28 August 2015; accepted 1 September 2015
DOI 10.1002/sce.21206
Published online in Wiley Online Library (wileyonlinelibrary.com).

ABSTRACT: Many resources have been committed to research on science teaching ped-

constraining or supporting such best practices at the elementary level. This study attempts to Þll this need through a multilevel model of how teacher traits, socioeconomic context (SE context), and accountability pressures predict studentsÕ opportunity to engage in hands-on and laboratory science education. Results indicate accountability pressure eclipsed all other predictors, including SE context, in accounting for variance in the model. Final analysis in-

INTRODUCTION

Recent policy developments in the United States invoke the economic importance of student preparation for Science, Technology, Engineering, and Math (STEM) careers as reported as hands-on or laboratory activities. With this caveat, we situate the study in the existing literature on both inquiry and hands-on approaches to science education.

Inquiry, project-based learning, and various forms of experiential learning have deep roots in educational practice and literature, starting with Dewey, Kilpatrick, and turn of the 20th-century progressives (Dewey, 1916; Montgomery, 1994). Inquiry speciÞcally has been a hallmark of excellent science education (Abd-El-Khalick et al., 2004; Anderson, 2012; Marshall, Horton, Igo, & Switzer, 2009). Inquiry was deÞned by the NRC (1996) as involving students in investigation and experimentation activities to Òdevelop knowledge and understanding of scientiÞc ideas, as well as an understanding of how scientists study the natural worldÓ (p. 23).

DeÞnitions of inquiry have evolved to include students conducting data collection and analysis, engaging in reasoning, explanation and argumentation, and communicating results (Abd-El-Khalick et al., 2004; Duschl & Osborne, 2002), all of which served as a foundation for the NRC (2012a) framework for KÐ12 Science Education of the NGSS science and engineering practices. In specifying the practices of science, the framework (NRC, 2012a) laid out three spheres of activity: investigating, evaluating, and developing explanations and solutions. In both literature and the present study, teachersÕ descriptions of hands-on and laboratory science correspond most closely to the Þrst, ranging from Òcookbook labsÓ to investigation activities that engage students in critical thinking and meaning construction (Dorph et al., 2011; Ginns & Watters, 1999; NRC, 2012a).

Arguably, children should have the opportunity to participate in the full range of science education activities (Duschl et al., 2007), including direct instruction, demonstration, and inquiry or laboratory activities. Yet evidence suggests that inquiry and the opportunity for inquiry provided by hands-on, lab-based activities are neglected in many elementary classrooms, particularly in high-poverty contexts (Capps & Crawford, 2013; Dorph et al., 2011; Fulp, 2002). Inequities in childrenÕs exposure to hands-on learning may limit science career preparation and their ability to participate as full citizens in an increasingly technoscientiÞc society. Differential distribution of science pedagogical practices at the elementary level, however, is not well documented.

Differentiating the Role of Teachers, Social Context, and Policy Milieu

Scholars who attend to multiple factors that guide instructional practices have described a combination of internal elements (a teacherÕs content preparation, conÞdence, attitude, beliefs about students, classroom management, and other elements of individual discretion) and external elements (resources, materials, student population, leadership support, and policy directives) (Biggers, 2013; Lee & Houseal, 2003; Valli & Buese, 2007). The present study draws from these elements in demonstrating the role of teacher traits (internal; Level 1), including experience, attitude, hours of science professional development (PD), and degree; it also draws from policy/contextual factors (external, Level 2), including accountability pressures and SE context. This review brießy describes each of these in turn before turning to the model.

Teacher Traits: Well-Researched but Still Uncertain.

Certainly there have been valid concerns regarding the lack of science content and pedagogical preparation among elementary teachers. Preservice elementary teachers tend to take few science courses in college or during teacher preparation (Fulp, 2002; Lee & Luykx, 2005; Ramey-Gassert, Shroyer, & Staver, 1996), and they may lack preparation in teaching using inquiry pedagogies (Schneider & Plasman, 2011). Yet the relationship of this lack of preparation to inquiry or hands-on practice is inconsistent. Lack of scientiÞc content knowledge may affect teacher

Shaver, Cuevas, Lee, & Avalos, 2007; Warburton & Torff, 2005). The spotty and mixed results in this area indicate a clear need for modeling the inßuence of student SE context on teacher instructional strategies in elementary science.

Bringing It All Together: What Is Known of the Predictors of Hands-On or Inquiry Practice

If hypotheses one and two are substantiated in the primary model, teacher traits would have a less substantial relationship to instructional practices than accountability pressure. If that is the case, teacher preference for certain practices should not differ signiÞcantly across accountability pressure, but their ability to carry out those pedagogies would differ. It follows that:

Hypothesis 3. Accountability pressure will predict the difference between reported and preferred instructional practices to a greater extent than teacher traits or SE context.

METHODS

Sampling Procedure

District Role. Districts play a major role interpreting and setting policy by allocating time, supporting PD, setting priorities, and choosing curriculum (Hamilton et al., 2007). Thus, a typical sampling procedure employed when sampling many districtsÐoften only one to Þve schools per district and only a few teachers at each schoolÐalthough more generalizable, is less able to delineate teacher and school effects within a given district policy context. Because the present study focuses on school level effects of SE context and policy, we sampled half the schools in one carefully selected district to control for district curriculum and policy interpretations. The sampling design thereby allowed for a clearer analysis of the relevant factors than a broad, but shallow, sampling design.

Valley district was selected because it was representative of California districts in the following ways: (a) the district means are quite close to California state means in API, percent English language learners, and percent FRL (Figure 1), and (b) it spans both urban and suburban areas of a mid-sized city, thus schools vary widely in FRL, ethnicity, and accountability measures (Appendix, Table A1). Due to the focus on one district, generalizability is a limitation. Nonetheless, the results of this study lay the foundation for additional studies, as well as providing veriÞcation for qualitative Þndings regarding accountability pressure.

To obtain a minimal sample size that would produce accuracy within ±5% of the district teacher population at a 95% conÞdence interval (Rea & Parker, 2005), we sampled 231 of 580 valley district elementary teachers (Grades KÐ5). A random stratiÞed sampling procedure was used to select 20 schools from the set of 42 elementary schools (Rea & Parker, 2005). Six schools were randomly selected from the lowest and highest API quartiles and four each from the middle quartiles. Selection was more heavily weighted at the ends of the spectrum to have adequate sampling representation for the Level 2 predictors, accountability pressure, and socioeconomic status. This represents a limitation in calculating instructional practice averages, but it has negligible effect on HLM statistics.

Sampling response bias examined on FRL and API were within a reasonable range (Table 1). Two schools out of the original sample that opted not to participate were replaced with the school with the most similar API. We asked all KÐ5th-grade teachers in each sampled school to complete the survey. Average teacher participation rate across schools was 71%, a total of 182 teachers. Of these teachers, 84% completed the entire survey and are represented in the full model. PearsonÕs chi-squared and z-tests were used to determine teacher response bias through differences in the numbers of teachers at each grade level per API quartile. Differences were statistically insigniÞcant.

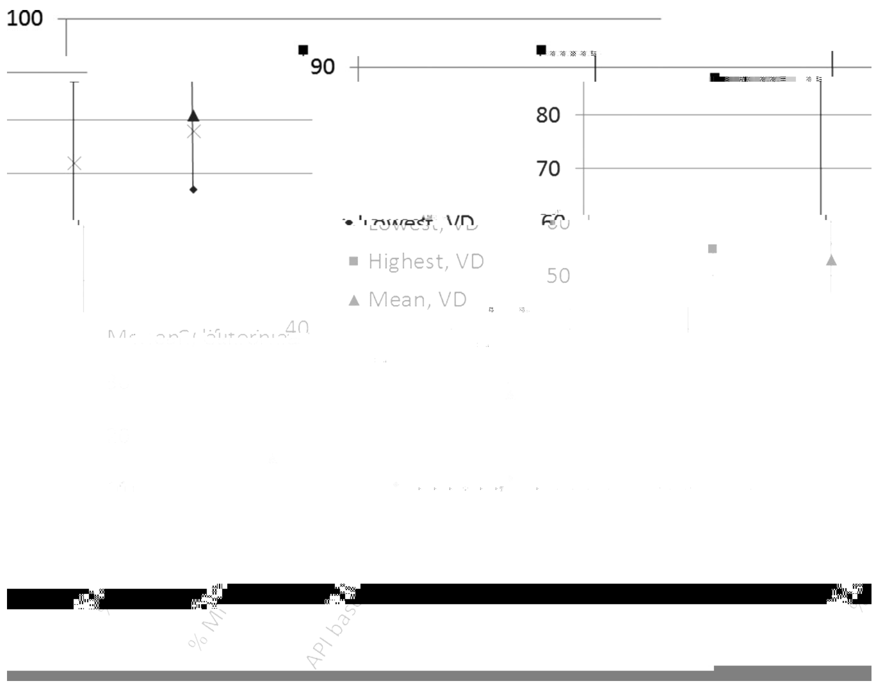

Figure 1. Mean and range of the percent of English language learners, free and reduced lunch, minority, and API base score (2011) for schools in valley district (VD), as compared to California state means (see Table A1 in the Appendix for numerical data).

TABLE 1
School Response by API Quartile, With Analysis of Sampling Bias in the Two Schools That Chose Not to Participate

				N [illegible]	
API Q[illegible]	Q[illegible]	P[illegible]	P[illegible]	FRL	API
F[illegible]	11	6	5	0.83 D	1 D
[illegible]	11	4	4	N/A	N/A
[illegible]	10	4	3	1 D	1.2 D
F[illegible]	10	6	6	N/A	N/A
	42	20	18		

Note: [illegible] (90%); 18 20.

Instrumentation

Survey. The Science Instructional Time and Pedagogy (SITP) survey consisted of seven sections, four of which were used for the models presented in this paper (see Table 2). Either the researchers or principal presented the survey to the teachers with an electronic link. In

TABLE 2
Description of Survey Sections

Section	Area	Question Types
1	Demographics	Three closed-response questions
2	Instructional time	Two numerical answer questions regarding science education time, divided into options
3	Pedagogy distribution	Two numerical answer questions asking for time attribution as a percentage
4	Thoughts on science education	Three Likert-type questions with a total of 10 statements Two demographic questions One question asking for details on hours of professional de0.1(e)-na8(5)n m e Tm .02d6One question asking forxplancation science t

TABLE 3
Description of Main Variables

[illegible]	[illegible]	[illegible]	[illegible]

TABLE 4
Reliability Measure (CronbachÕs Alpha) Between Percent Teachers Reported and Preferred for Each Pedagogy as well as Average Percent Reported

Variable (Reported and Preferred Percent)	Item: Percent of Time Teachers Report . . .	Cronbach's Alpha	Average Percent Reported (%)
Hands-on	Students doing hands-on or		

TABLE 5
Descriptive Statistics for Predictor and Dependent Variables

	N	M[illegible]	M[illegible]	M[illegible]/P[illegible]	D
[illegible]					
A P [illegible]	18	0.00	6.00	2.61	1.94
E [illegible]	18	11.30	89.95	42.21	24.07
[illegible]					
[illegible]	182	0.00	200	68.37	42.93
P [illegible]	161	0.00	100	42.19	24.70
P [illegible]	161	0.00	100	21.64	23.44
P [illegible]	159	0.00	100	54.58	20.58
P [illegible]	159	0.00	70.00	13.03	11.85
G [illegible] 2 3	161	0.00	1.00	0.33	0.47
G [illegible] 4	161	0.00	1.00	0.20	0.40
G [illegible] 5	161	0.00	1.00	0.16	0.37
PD [illegible] ([illegible])	162	0.00	17.00	4.86	2.86
[illegible] ([illegible])	162	2.00	5.00	4.56	0.80
D [illegible] ([illegible])	153	1.00	2.00	1.12	0.33
A [illegible] (L [illegible])	155	1.25	4.75	4.04	0.65

Note: F [illegible], N [illegible]. F [illegible], N [illegible].

TABLE 6
Intercorrelations Among Teacher-Level Variables (Level 1)

	G [illegible] (O [illegible])	D [illegible]	PD H [illegible]	A [illegible]	[illegible]
G [illegible] ([illegible], 1 5)	1	.163*	.052	Š.005	Š.012
D [illegible]		1	[illegible] .042	.147	.038
PD [illegible]			1	.200*	.101
A [illegible]				1	.076
[illegible]					1

*C [illegible] 0.05 [illegible]).

positively skewed due to a few teachers reporting many hours of PD and many reporting zero hours. However, transformations of the variable did not result in more accurate modeling or shifts in signiÞcance.

Years taughtallowed for Þve choices (1Ð3, 4Ð6, 7Ð9, 10Ð15, and 15 result 725.6.2226e0 -1

Level 2 Variables

SE Context. The SE context variable was a composite of school FRL percent and percent of students underrepresented in science (not White or Asian) (CronbachÕs alpha .923) (Ed-Data, 2013). We used percent underrepresented as part of SE context rather than percent minority because Asian students are highly represented in both science majors and careers, and thus percent minority would be misleading in terms of equity (PCAST, 2010).

AYP Pressure. Under NCLB, whether or not a school makes AYP each year for each subject (math and ELA) and subgroup of students is used to determine sanctions; in California these pressures accumulate; at Year 2 schools must notify parents of being out of complianceÑby Year 5 schools are subject to restructuring and alternative governance. They do not reset unless the school makes AYP 2 years in a row. Because Ònot making AYPÓ has been a key element of pressure on schools and teachers (Dorph et al., 2011; Hamilton et al., 2007; Judson, 2013; Penuel et al., 2008), and because sanctions accumulate up to Year 6 and are continual and iterative thereafter (California Department of Education, 2012), the measure for this construct was calculated by adding the cumulative years each school did not make AYP in either math or LA out of the last 6 years (Ed-Data, 2013). Following state policy, our measure of accumulated pressure was not reset unless the school made AYP in the given subject 2 years in a row. Because non-Title 1 schools still receive sanctions (although less speciÞed) for not making AYP or API growth in California, both Title 1 and non-Title 1 schools were included in this measure. Cumulative AYP pressure be interrelated with school-level factors, prior to centering, all variables were tested for interactions and whether they accounted for Level 2 variance. All interactions were in-signiÞcant. Only one variable, *attitude*, accounted signiÞcantly for Level 2 variance. When group centered, each measure of teacher traits represented the distance of that teacher trait from the school mean, with the school mean set at zero. The intercept ($_{0j}$) then became the mean for each school *j* at Level 2. Level 2 variables were left uncentered. Slopes of Level 1 variables were Þxed at Level 2 to maintain a focused model (Maerten-Rivera, Myers, Lee, & PenÞeld, 2010), and because tests of homogeneity of Level-1 variance (the variance of Level 1 slopes across Level 2) were insigniÞcant. Thus, this model portrayed how group means (Level 1 intercepts) varied across schools rather than variance in slope coefÞcients across schools (Raudenbush & Bryk, 2002). In other words, the model tests variance in the mean percentage of hands-on practices across schools rather than variance in the relationship (slope) between hands-on practices and teacher trait (Level 1) variables across schools. All models used restricted maximum likelihood; was set to .05.

As advocated by Raudenbush and Bryk (2002), after estimating the null model, this study compared Level 1 and Level 2 models separately, then added Level 2 predictors to the Þnal Level 1 model. Because there was little theoretical foundation for the order of adding variables to the model, Level 1 variables were each added to the model individually Þrst, then sequentially. The use of the deviance statistic to evaluate model Þt is inappropriate in this case due to sample size (Raudenbush & Bryk, 2002). Although Level 1 variables were insigniÞcant, they were retained in the Þnal model for theoretical purposes.

Level 1 model (each variable except grade centered on school mean):

$$\%\text{Hands-on}_{ij} = {}_{0j} + {}_{1j} \times \text{Yrs\_teach}_{ij} + {}_{2\check{S}4j} \times (\text{Gradedum}_{ij}) + {}_{5j} \times (\text{Degree}_{ij}) + {}_{6j} \times (\text{PD}_{ij}) + {}_{7j} \times (\text{Attitude}_{ij}) + r_{ij}$$

Akin to a basic linear regression, the Level 1 model speciÞes the predicted percent of hands-on (Y_{ij}) for individual teacher *i* in school *j*. $_{0j}$ is the intercept, or grand mean of all schools for kindergarten and Þrst-grade teachers (the omitted grade variable) when all others are centered around the school mean. $_{1j}$ through $_{7j}$ are Þxed coefÞcients identifying the vector of hands-on practices for each teacher at each school, *j*, based on *years taught*, *grade*, *degree*, *PD hours*, and *attitude*. In sum, the teacherÕs percent of time teaching hands-on is predicted as a function of their experience, grade level, BA degree, PD hours, and attitude, along with error unexplained by these variables. Interschool variation is represented by Level 2 models.

At Level 2, the constant from the level 1 model ($_{0j}$) is a function of the grand mean across schools ($_{00}$) plus a coefÞcient representing the effect of accountability pressure on the portion of variance in percent hands-on ($_{01}$) and a school-level error term (u_{0j}). In other words, the intercept from Level 1 (the mean of the school) is predicted as a function of AYP pressure and error. The Þxed coefÞcients vectors ($_{10}$ – $_{70}$) represent the constant coefÞcient for each Level 1 variable.

Level 2 model:

$$_{0j} =$$

TABLE 7
Variance Components of the Baseline Model (ICC)

	[illegible]	P[illegible]	X^2	p
[illegible] (L[illegible] 2)	153.77	24.9	67.91	< .001
[illegible] (L[illegible] 1)	464.21	75.1		
[illegible]	617.98	100		

Full model:

$$\%\text{Hands-on}_j = {}_{00} + {}_{01} \times (\text{AYPPressure}_j) + {}_{10} \times (\text{Yrs\_teach}_{ij}) + {}_{20\check{S}40j} \times (\text{Gradedum}_{ij}) + {}_{50} \times (\text{Degree}_{ij}) + {}_{60} \times (\text{PD}_{ij}) + {}_{70} \times (\text{Attitude}_{ij}) + u_{0j} + r_{ij}$$

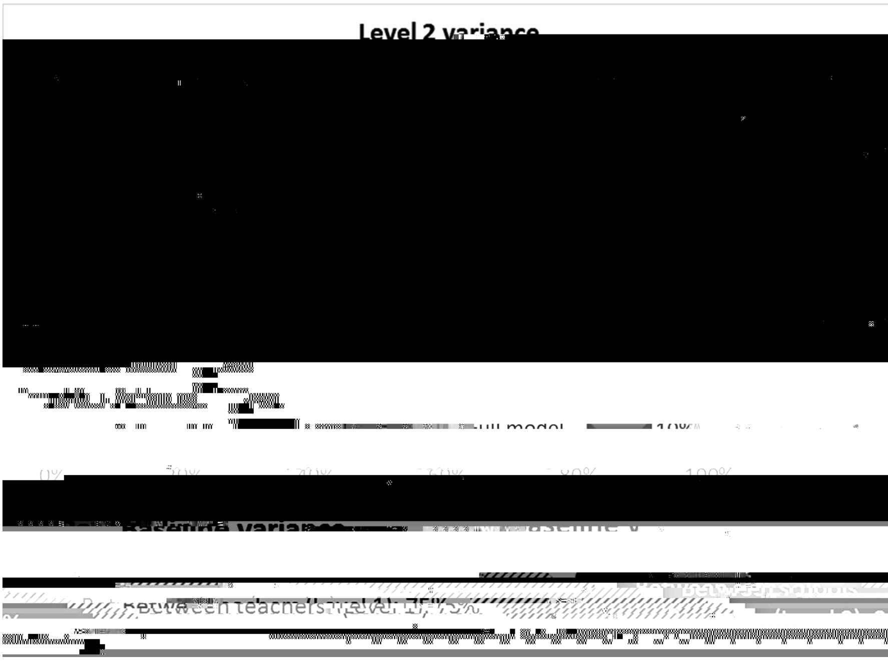

Figure 2.

Level 2

As a sole Level 2 predictor, *AYP pressure* was signiÞcant ($p < .05$) in predicting *percent hands-on*, and accounted for 32% of Level 2 variance and 8% of overall model variance compared to the baseline ICC (Table 8 and Figure 2). Similarly, with *percent textbook* as the dependent variable, *AYP pressure* (independently) accounted for 23% of the Level 2 variance and 5% of full model variance (not shown), signiÞcantly predicting 3.2 percentage points more text use for every consecutive year the school did not make AYP. Because *percent textbook* acted to some extent as a mirror to *percent hands-on*, the latter is the focus of the models and discussion.

In contrast to *AYP pressure*, *SE context* was insigniÞcant as a sole predictor in the model. This difference is noteworthy given the high correlation between *SE context* and AYP pressure ($r = .781$, $p < .001$). To reduce overspeciÞcation, *AYP pressure* was selected as the sole Level 2 variable in the full model. In the full model (Table 8, Model 4), *AYP pressure* predicts that for every consecutive year the school did not make AYP in either math or ELA, teachers on average reduced their use of hands-on and laboratory instruction 4.3 percentage points. Thus, schools that did not make AYP 0Ð4 years (i.e., made AYP most years) out of the last six averaged 47% hands-on and laboratory activities; students in schools that did not make AYP 5Ð6 consecutive years out of the last six (high accountability pressure) averaged 26%. We will use this measure of high accountability pressure descriptively throughout the results as an interpretive tool because the strongest sanctions are applied the Þfth consecutive year the school does not make AYP.

Hypotheses

Regarding the Þrst hypothesisÑcommunity socioeconomics and accountability pressure have a greater relationship to science instructional practices than teacher traitsÑthe results were positive. Although much of the variance was between teachers at Level 1 (75%), no tested teacher traits accounted for any of this variance. Training (*PD hours*), preparation (*degree*), *attitude*, and experience (*years taught*), often the focus of research and policy, were all insigniÞcant in the Þnal model (Table 8), and these variables explained little of the variation in instructional practice over the baseline decomposition of variance (less than 4%; Figure 2). Conversely, measures of *AYP pressure* were signiÞcant solely and in the full model and explained substantive variance. In addition, *attitude* when uncentered explained 9% of Level 2 variance, indicating the possibility of a relationship between school type and attitude.

The second hypothesisÑaccountability pressure has a greater relationship to elementary science education instructional practices than community SE contextÑwas also demonstrated in the model. As a sole Level 2 predictor as well as in the full model, *AYP pressure* was signiÞcant, whereas *SE context* was not. In addition, *AYP pressure* accounted for nearly double the Level 2 variance as *SE context*.

The third hypothesis was constructed to further clarify whether accountability pressure was predicting teacher *ability* to carry out particular practices rather than teacher *preference* for particular practices. For this hypothesis, we tested the relationships between the predictor variables and the difference between preferred and reported practices. The two variables of interest, *percent text* and *percent hands-on*, showed a substantial gap between predicted and reported (Figure 3).

For each respondent, the percent reported was subtracted from the percent preferred (Figure 4). On average, teachers in both high and low accountability pressure schools preferred to use hands-on or lab pedagogies around half of their science instruction time (49%

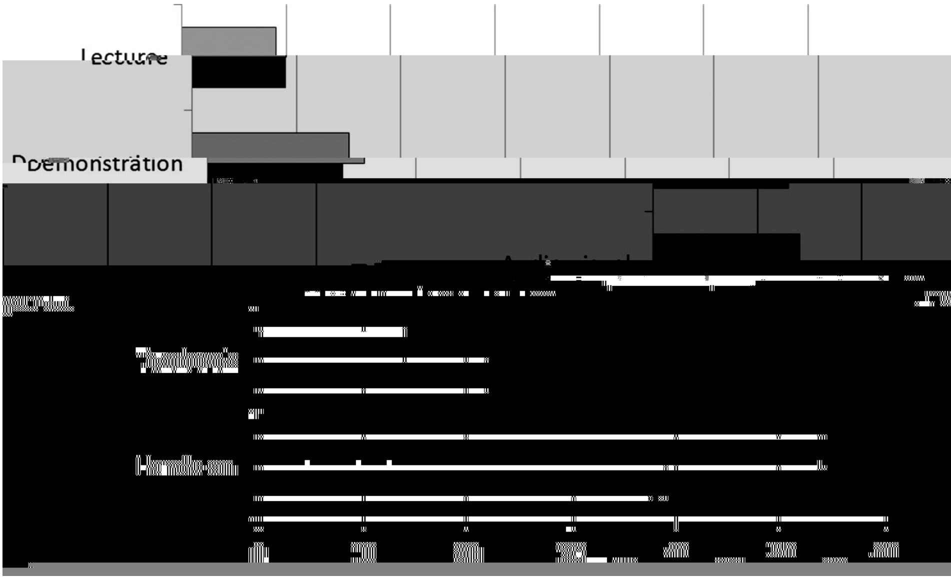

Figure 3. Comparison of teachersÕ average reported and preferred instructional practices (percent).

Figure 4. Difference between reported and preferred percent textbook and hands-on, high accountability pressure schools (AYP 56), and lower accountability pressure schools (AYP 0Ð4).

and 55%, respectively; the difference was insigniÞcant). However, there was a 23 percentage point difference between preferred and reported for teachers at high accountability-0.2(s)-25

TABLE 9
HLM CoefÞcients Reporting the Relationship of High Accountability Pressure to the Difference Between Preferred and Reported Hands-On and Textbook Pedagogies

Predictor	Hands-On	Textbook
AYP pressure	2.838*	Š2.301*
	(0.994)	(1.067)
Intercept	5.778	Š2.865
	(3.125)	(3.360)
Observations	159	159

Note: *p < .05.
Slope coef cient and standard errors in parentheses (xed effect).

the dependent variables were the difference between preferred and reported hands-on and textbook pedagogies. For both textbook and hands-on, all Level 1 variables were insigniÞcant and were omitted from the model. For percent hands-on, AYP pressure and SE context were each signiÞcant as the sole predictor, but AYP pressure accounted for the most underrepresented students had considerably less chance of being exposed to excellent and

IMPLICATIONS

These results have implications regarding science education reform efforts. Research and national priorities have up to this point been focused primarily on teacher development. Teachers are often the Òconvenient objects of criticism,Ó but within an institutional structure driven by external policies, their choices may be constrained (Cuban, 2004). As demonstrated in the PD literature, shifts in practice often require intensive, long-term PD, well integrated into schools and reliant on a shared vision (Desimone et al., 2002; Elmore, Peterson, & McCarthey, 1996). Leadership, school capacity, and resources also play a role (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; Goetz Shuler, Backman, & Olson, practices were unlikely to predict AYP pressure as science test scores account for less than 6% of school AYP calculation.

APPENDIX

TABLE A1
Range and Mean for Schools in Valley District, 2011 (Rounded For Confidentiality)

	Range	Mean
ELL	0–55%	16%
FRL	10–95%	50%
P[illegible]	10–85%	30%
API ([illegible] 1000)	665–950	790

REFERENCES

Abd-El-Khalick, F., Boujaoude, S., Duschl, R. A., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A., & Tuan, H.-L. (2004). Inquiry in science education: International perspectives. Science Education, 88, 397–419.
Allison, P. D. (1999). Multiple regression: A primer. Thousand Oaks, CA: Sage.
Anderson, K. J. B. (2012). Science education and test-based accountability: Reviewing their relationship and exploring implications for future policy. Science Education, 96, 104–129.

National Research Council (NRC). (2012b). Monitoring progress toward successful K-12 STEM education: A nation advancing? Washington, DC: The National Academies Press.
National Research Council (NRC). (2013). Next generation science standards: For states, by states. Washington, DC: The National Academies Press.
Penuel, W., Fishman, B., Gallagher, L., Korbak, C., & Lopez-Prado, B. (2008). Is alignment enough: Investigating the effects of state policies and professional development on science curriculum implementation. Science Education, 93, 656Ð677.
PresidentÕs Council of Advisors on Science and Technology (PCAST). (2010). Report to the President: Prepare and inspire: K-12 education in science, technology, engineering, and mathematics (STEM) for AmericaÕs future. Retrieved from http://www.whitehouse.gov/sites/default/Þles/microsites/ostp/pcast-stemed-report.pdf
Powers, J. M. (2003). Analysis of performance-based accountability: Factors shaping school performance in two school districts. Educational Policy, 17(5), 558Ð585.
Ramey-Gassert, L., Shroyer, M. G., & Staver, J. R. (1996). A qualitative study of factors inßuencing science teaching self-efÞcacy of elementary level teachers. Science Teacher Education, 80(3), 283Ð315.
Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage.
Rea, L. M., & Parker, R. A. (2005). Designing and conducting survey research: A comprehensive guide. San Francisco, CA: Jossey-Bass.
Reiser, B. (2013). What professional development strategies are needed for successful implementation of the Next Generation Science Standards? Washington, DC: Paper presented at the Invitational Research Symposium on Science Assessment. K-12 Center at ETS.
Schneider, R. M., & Plasman, K. (2011). Science teacher learning progressions: A review of science teachersÕ pedagogical content knowledge development. Review of Educational Research, 81(4), 530Ð565.
Settlage, J., & Meadows, L. (2002). Standards-based reform and its unintended consequences: Implications for science education within AmericaÕs urban schools. Journal of Research in Science Teaching, 39, 114Ð127.
Shaver, A., Cuevas, P., Lee, O., & Avalos, M. (2007). TeachersÕ perceptions of policy inßuences on science instruction with culturally and linguistically diverse elementary students. Journal of Research in Science Teaching, 44(5), 725Ð746.
Sirin, S. R. (2005). Socio-economic status and academic achievement: A meta-analytic review of research. Review of Educational Research, 75(3), 417Ð453.
Smarter Balanced Assessment Consortium. (2013). Achievement descriptors. Retrieved from http://www.smarterbalanced.org/achievement-level-descriptors-and-college-readiness/
Smith, L. K., & Southerland, S. A. (2007). Reforming practice or modifying reforms? Elementary teachersÕ response to the tools of reform. Journal of Research in Science Teaching, 44(3), 396Ð423.
Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. Journal of Research in Science Teaching, 37(9), 963Ð980.
Tabachnick, B. G., & Fidell, L. S. (2007). Experimental designs using ANOVA. Belmont, CA: Thompson.
U.S. Department of Education. (2007). Federal government press release. Retrieved from http://www2.ed.gov/news/pressreleases/2006/02/02062006.html
Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. American Educational Research Journal, 44(3), 519Ð558.
van Aalderen-Smeets, S., & Walma van der Molen, J. (2013). Measuring primary teachersÕ attitudes toward teaching science: Development of the dimensions of attitude toward science (DAS) instrument. International Journal of Science Education, 35(4), 577Ð600.
Warburton, E., & Torff, B. (2005). The effect of perceived learner advantages on teachersÕ beliefs about critical-