International Journal of Science Education
Vol. 31, No. 5, 15 March 2009, pp. 675–696

RESEARCH REPORT

Same Science for All? Interactive association of structure in learning activities and academic attainment background on college science performance in the USA

Robert H. Tai[a]* and Philip M. Sadler[b]
[a]*University of Virginia, USA;* [b]*Harvard-Smithsonian Center for Astrophysics, USA*

This US study investigates interactive associations between structure in inquiry-type learning activities and academic attainment in high school science with introductory college science performance as the outcome. Past studies of this type have tended to use smaller samples and shorter-range methods of assessing the influence of interactions. This study used a large-scale nationally representative sampling of science students and investigated the existence of long-range associations between high school and college. Replicated across three different disciplinary data-sets (biology, chemistry, and physics) totaling over 8,000 surveys, the analysis discovered an interaction between structure and attainment associated with differences in long-range student performance. The implications for these findings in terms of theory and practise are discussed in the conclusions.

Introduction

Inquiry-based instructional practises are a mainstay of the *National Science Education Standards* (National Research Council, 1996) and *Benchmarks of Science Literacy* (AAAS, 1993) in the USA. An important consideration of inquiry is the degree of teacher guidance/instructional structure; the National Research Council teachers' guide noted that 'Inquiry-based teaching can also vary in the amount of detailed guidance that the teacher provides' (2000, p. 28). In fact, the guide asked the question 'How does a teacher decide how much guidance to provide in an inquiry?' (National Research Council 2000, p. 30). Of primary concern is the quality of student work

*Corresponding author. Curry School of Education, University of Virginia, 405 Emmet Street South, PO Box 400273, Charlottesville, VA 22904-4273, USA. Email: rht6h@virginia.edu

ISSN 0950-0693 (print)/ISSN 1464-5289 (online)/09/050675–22

DOI: 10.1080/09500690701750600

produced in these activities. For many teachers who assign inquiry activities relying on students to design and conduct them, the reality is that while some students may do good work, others languish (O'Neill & Polman, 2004; Polman, 2000).

One possible source of difference in students' performance is their background. Associations linking learning activities and students' backgrounds to attainment are called aptitude treatment interactions (Cronbach & Snow, 1977). Although a popular research topic from the 1960s to about the 1980s, findings were difficult to replicate (Tobias, 1981). Since the 1980s, aptitude treatment interactions research into inquiry-based instructional practises has been less common. Two exceptions are Eysenck (1996) who reported on interactions involving students' personality, and Windschitl and Andre (1998) who reported on interactions involving students' epistemological sophistication. Neither study considered students' prior academic attainment.

In fact, studies on inquiry-based learning in science do not typically distinguish students' performance in light of their academic backgrounds. For example, Tytler (1992) studied students selected from Science Talent Search prizewinners in Australia. While noting students' supportive home environments, he does not specifically discuss academic backgrounds, leaving the impression that these students were all top science achievers, which in all likelihood they were. In a review of their 'Community of Learners' project, Brown and Campione (1994) discussed pedagogical practise in terms of 'guided discovery', and specifically addressed the teachers' role in structuring inquiry-type learning activities. Their discussion, however, did not include academic attainment. Wallace, Tsoi, Calkin, and Darley's (2003) qualitative study of five non-major biology students found differences based on students' constructivist versus positivist learning beliefs, but, again, prior attainment was not included. In a pair of studies aimed at the efficacy of project-based science to prepare students for standardized examinations, Schnieder, Krajcik, Marx, and Soloway (2002) and Rivet and Krajcik (2004) considered student performance as an outcome measure. While prior academic attainment was discussed in general, it was not included in the standardized examination score analysis. A study by Hofstein, Shore, and Kipnis (2004) investigated Grade 11 and Grade 12 students participating in inquiry-type laboratory activities and found improvements in students' ability to ask scientific questions. The authors compared 'more experienced' and 'less experienced' students and found differences in students' experimental designs, but made no comparisons based on academic attainment.

A major finding of aptitude treatment interactions research was that higher achievers responded better in less-structured learning environments, while lower achievers responded better to more-structured environments. Therefore, accounting for students' attainment in research on inquiry-based learning activities would seem to be important. If instructional methods were chosen to more closely match students' backgrounds, more optimal levels of performance may be expected. However, does matching student prior attainment with pedagogy have an impact on student performance beyond the immediate learning activity? Investigations of long-range interactive associations are not common. In fact, a look at both Cronbach and Snow (1977) and Tobias (1981) show no investigations of long-range outcomes. Is matching student attainment with level of structure linked with differences in student performance beyond immediate measures of attainment? In light of increasing numbers of students entering college (Bureau of Labor Statistics, 2005) and self-directed pressure by many teachers themselves to both prepare their students for college and ignite their interests in science (Hoffer, Quinn, & Suter, 1996), one option for a long-range measure is performance in introductory college science.

To address the question above, the present study investigated the existence of an interactive association linking students' academic background (as measured by high school grades, standardized examinations, and advanced high school course-taking) and students' experience with two types of inquiry-related pedagogies (number of student-designed projects and degree of freedom in laboratory exercises) with introductory college science course performance as the outcome. The investigation included analyses replicated across three data-sets in three science disciplines. We hope that the results of this investigation will provide both science teachers and researchers with some insight into how matching students' attainment backgrounds with learning activities may have a long-range influence on science performance.

Methodology

About the Research Project

The Factors Influencing College Science Success Project, a four-year study funded through the Interagency Educational Research Initiative and administered through the National Science Foundation (NSF-REC 0115649), has collected a wide array of survey data from college students, high school teachers, and college professors. For this study, we analysed a subset of the existing data: student survey data from the first course of the introductory biology, chemistry, and physics sequences for science or engineering majors, collected in fall 2002 and fall 2003. In the case of physics, algebra-based courses were also surveyed since this course could be used by non-physics science majors and professional studies majors to fulfill their major requirements (e.g., pre-medicine, pre-dentistry, pre-pharmacy, and pre-veterinary majors).

The survey questionnaires (three versions for biology, chemistry, and physics) were based on an earlier national survey of physics students (Sadler & Tai, 2001; Tai & Sadler, 2001). Interviews with high school teachers and college professors carried out by the Factors Influencing College Science Success Project researchers M. Schwartz and Z. Hazari were used to development the instruments.

About the Data and Data Collection

The three data-sets analysed in this particular investigation were collected from college students in 128 different first-semester introductory college science courses taught at 55 four-year US colleges and universities, a subset of 67 selected from a comprehensive list of nearly 1700 institutions[1]. Although small private schools make up a majority of the individual institutions, nearly 50% of all students in 4-year colleges and universities attend only 5% of the institutions, typically large public universities. Given this unequal distribution of students in colleges, a stratified random sample based on school size was used to ensure proper representation.

Faculty in 29 biology departments, 31 chemistry departments, and 37 physics departments in the 55 schools from 31 states participated. Institutional descriptive data are presented in Tables 1–3. Concerned about institutional 'self-selection' bias, we compared participating and non-participating schools across measures such as

Table 1. Summary of institutional characteristics of participating schools in biology

College/ university	Participants	Affiliation	Average ACT	Average SAT	School size[a]	US State
School 1	16	Private	22	1,050	Small	AZ
School 2	16	Private	20	890	Small	KY
School 3	16	Public	24	1,120	Large	TN
School 4	19	Public	20	970	Small	GA
School 5	22	Private	25	1,130	Small	NC
School 6	23	Public	25	1,140	Large	FL
School 7	28	Public	23	1,080	Medium	NY
School 8	34	Public	17	840	Small	PA
School 9	41	Public	21	1,000	Medium	TX
School 10	45	Public	22	1,050	Small	NE
School 11	45	Public	20	970	Medium	KY
School 12	68	Public	21	990	Medium	AK
School 13	74	Private	28	1,240	Small	TX
School 14	80	Public	22	1,040	Small	MI
School 15	86	Public	22	1,050	Large	FL
School 16	91	Private	23	1,170	Small	OR
School 17	92	Private	25	1,160	Small	MI
School 18	95	Private	26	1,180	Medium	NY
School 19	99	Private	23	1,080	Small	NJ
School 20	103	Public	24	1,100	Large	IN
School 21	103	Public	22	1,050	Medium	WV
School 22	104	Private	24	1,120	Small	IA
School 23	116	Public	19	930	Medium	LA
School 24	129	Public	23	1,080	Medium	WI
School 25	132	Private	30	1,320	Small	IL
School 26	189	Public	24	1,090	Large	CA
School 27	240	Public	24	1,120	Large	LA
School 28	251	Public	24	1,120	Large	OH
School 29	392	Public	23	1,080	Medium	UT
Category totals	2,749	19 public, 10 private			13 small, 9 medium, 8 large	23

Note: [a]Small schools <5,000 full-time equivalents (FTE), medium-size schools between 5,000 and 15,000 FTE, and large schools >15,000 FTE.

Table 2. Summary of institutional characteristics of participating schools in chemistry

College/ university	Participants	Affiliation	Average ACT	Average SAT	School size[a]	US State
School 1	12	Public	23	1,080	Small	NY
School 2	13	Public	22	1,050	Medium	WV
School 3	17	Public	22	1,050	Medium	TN
School 4	21	Public	22	1,020	Medium	GA
School 5	22	Private	21	1,010	Small	PA
School 6	23	Private	20	975	Small	SC
School 7	30	Public	20	970	Medium	KY
School 8	34	Private	19	930	Small	KY
School 9	36	Public	23	1,080	Medium	NY
School 10	38	Private	22	1,050	Small	AZ
School 11	39	Public	17	830	Small	CA
School 12	43	Private	25	1,160	Small	MI
School 13	43	Private	22	1,050	Small	IL
School 14	48	Private	23	1,170	Small	OR
School 15	57	Public	21	990	Small	NH
School 16	59	Public	21	1,000	Medium	TX
School 17	60	Public	23	1,060	Small	ME
School 18	68	Private	21	1,010	Small	MI
School 19	84	Public	24	1,100	Medium	WA
School 20	88	Private	24	1,120	Small	AL
School 21	94	Public	22	1,050	Small	SD
School 22	118	Public	21	1,010	Large	CA
School 23	120	Private	26	1,180	Small	PA
School 24	134	Public	19	930	Medium	LA
School 25	155	Public	24	1,110	Large	AZ
School 26	177	Private	26	1,200	Small	IN
School 27	256	Public	23	1,080	Medium	ID
School 28	271	Public	24	1,120	Large	LA
School 29	411	Public	27	1,210	Medium	MD
School 30	434	Public	21	990	Medium	IN
School 31	516	Public	24	1,120	Large	KY
Category totals	3,521	20 public, 11 private			16 small, 11 large, 4 medium	22

Note: [a]Small schools <5,000 full-time equivalents (FTE), medium-size schools between 5,000 and 15,000 FTE, and large schools >15,000 FTE.

school size, admissions selectivity, and geographic location. Our analysis found no indications of bias.

For purposes of comparability, this study surveyed included only courses using the lecture/recitation/laboratory formats. This format is by far the most common in US institutions. Institutions with novel formats were not included in the study; for

Table 3. Summary of institutional characteristics of participating schools in physics

College/ university	Participants	Affiliation	Average ACT	Average SAT	School size[a]	US State
School 1	7	Public	17	830	Small	CA
School 2	7	Private	23	1,080	Small	NJ
School 3	8	Private	20	975	Small	SC
School 4	8	Public	24	1,100	Large	IN
School 5	8	Public	17	840	Small	PA
School 6	8	Private	19	930	Small	KY
School 7	10	Public	22	1,050	Small	NE
School 8	10	Public	22	1,050	Medium	WV
School 9	11	Private	25	1,130	Small	NC
School 10	11	Private	22	1,050	Small	IL
School 11	13	Private	22	1,050	Small	AZ
School 12	13	Public	24	1,120	Large	TN
School 13	19	Public	23	1,080	Medium	NY
School 14	19	Private	21	1,010	Small	PA
School 15	19	Private	25	1,160	Small	MI
School 16	20	Public	22	1,050	Medium	TN
School 17	22	Private	22	1,050	Small	IA
School 18	26	Public	20	970	Small	GA
School 19	28	Public	21	990	Medium	IN
School 20	33	Public	23	1,080	Medium	CO
School 21	33	Private	26	1,200	Small	IN
School 22	36	Public	20	970	Medium	KY
School 23	36	Public	21	990	Medium	AK
School 24	38	Public	23	1,080	Medium	ID
School 25	44	Private	28	1,240	Small	TX
School 26	45	Public	24	1,090	Large	CA
School 27	50	Public	19	930	Medium	LA
School 28	51	Public	24	1,100	Small	GA
School 29	57	Private	23	1,170	Small	NY
School 30	58	Private	30	1,320	Small	IL
School 31	64	Public	23	1,080	Medium	UT
School 32	75	Public	24	1,100	Medium	WA
School 33	183	Public	24	1,120	Large	KY
School 34	185	Public	24	1,110	Large	AZ
School 35	202	Public	27	1,210	Medium	MD
School 36	216	Public	24	1,120	Large	LA
School 37	230	Public	24	1,120	Large	KS
Category totals	1,903	24 public, 13 private			18 small, 12 medium, 7 large	26

Note: [a]Small schools <5,000 full-time equivalents (FTE), medium-size schools between 5,000 and 15,000 FTE, and large schools >15,000 FTE.

example, Studio Physics at Rensselaer Polytechnic Institute (Kummings, Marx, Thornton, & Kuhl, 1999).

The surveys were administered by professors or teaching assistants during class meetings. The college professors later entered the students' final course grades on the surveys and returned them. The sample sizes totaled: 2,754 biology surveys, 3,521 chemistry surveys, and 1,903 physics surveys. For these sample sizes, statistical power analysis indicated a greater than 90% chance of detecting a small effect with a correlation of 0.20 (Light, Singer, & Willett, 1990, p. 201). We were careful to analyse sample clustering through the use of zip code data reported by the students. For biology, 1,820 zip codes had one zip code cluster, representing 1.3% of the sample. For chemistry, 1,983 zip codes had four zip code clusters, each at about 0.5% of the sample. For physics, 1,281 zip codes had one zip code cluster, representing 2.0% of the sample. The clusters were not found to be of sufficient size to influence our analysis.

About Retrospective Self-reports

Limitations in retrospective self-report surveys are an important consideration. Although retrospective self-report surveys are common (e.g., Longitudinal Survey of American Youth and the National Educational Longitudinal Survey of 1988) and play instrumental roles in education and public health research, still accuracy and reliability of self-reports are essential to consider.

Early findings questioning accuracy and reliability (Bradburn, Rips, & Shevell, 1987) have shifted to suggest that memory and recall can be reasonably accurate and reliable. Some studies have identified factors to improve accuracy (e.g., Niemi & Smith, 2003; Sawyer, Laing, & Houston, 1988; Schiel & Noble, 1991; Valiga, 1987), while others have indicated that reliability can also be improved to reasonable levels even over extended periods of time (Bradburn, 2000; Groves, 1989; Menon & Yorkston, 2000). In a review study, Kuncel, Credé, and Thomas (2005) noted that self-reported grades and other academic information were most accurate when the respondents considered this information relevant to their experiences. Our survey format and administration technique employed several accuracy-enhancing and reliability-enhancing measures: proper wording of questions, grouping questions into conceptually related sequences, providing contextual cues in the questionnaire, surveying students in associated situations and surroundings (such as classrooms and lecture halls during science classes), and making the survey relevant to the students. The questionnaires were developed through the use of pilot studies, field tests, and student focus groups.

We also performed a reliability study with 113 introductory college chemistry students who completed the questionnaire on two separate occasions, 2 weeks apart. The students were paid a small fee for participation. Their responses were, on average, exact in 60% of the cases, and within one choice of their original selection in 90% of the cases. The reliability correlations ranged from 0.46 to 0.69, which represent an acceptable reliability for this form of large-scale inferential analysis (Thorndike, 1997, p. 117). We took the approach of searching for trends within large samples of a population. While applying these techniques for individuals would be inaccurate, in large groups the use of such measures is quite powerful.

About Missing Data

As is common in large-scale surveys, not every participant answered every question without leaving blanks or marking multiple responses. The most typical method for dealing with missing data in regression analysis is list-wise deletion, excluding an entire survey from the analysis based on missing data. This tactic raises some concerns. First, a case might have all other responses present and still be list-wise deleted. Second, missing data may not be missing at random, and list-wise deleting surveys with missing responses would introduce bias. To mitigate the loss of data, the authors chose to employ the Expectation–Maximization Algorithm to impute data for the following predictors: scholastic aptitude test (SAT)-Mathematics, SAT-Verbal, Last High School Mathematics Grade, Last High School Science Grade, and Last High School English Grade (Allison, 2002; Little & Rubin, 2002; Peugh & Enders, 2004; Scheffer, 2002). Variables associated with one another through prior research are entered into the Expectation–Maximization Algorithm and missing values are calculated based on the non-missing values. This technique has been shown to be effective in statistical simulations where up to 10% of the data are missing (Scheffer, 2002). For this sample, missing value rates do not exceed 9.4% and they average 7.6% for the imputed variables. The use of the Expectation–Maximization Algorithm for imputed data has been shown to be effective in eliminating problems associated with biasing due to list-wise deletion (Scheffer, 2002).

In addition, data imputation allows for the application of $\Delta\chi^2$ significance testing to determine the effect size of each predictor, a technique not possible with missing values. Three standards for missing data exist: not missing at random (i.e., a biased sample), missing at random, and missing completely at random (i.e., the highest standard for missing values). We applied Little's Missing Completely at Random Test (Allison, 2002; Little & Rubin, 2002) in this analysis with the continuous control predictors included in the regression models. Overall, 88% of the surveys were retained in the regression analysis.

About the Approach to Data Analysis

The first step in data analysis was a descriptive comparison across different demographic and general educational background variables. Next, multiple linear regression models were fitted to the outcome variable, GRADES. This analysis included students' academic background measures (high school grades, standardized examinations, and patterns of advanced course taking), and students' experience with inquiry-related pedagogies (student-designed projects and degree of freedom in laboratory exercises) controlling for differences in demographic and general educational backgrounds. Even when sample sizes are large, care must be taken to avoid capitalizing on chance. To this end, this study included three replicate analyses.

College science grades are a natural choice to gauge college science performance (Tai, 2001; Gainen & Willemsen, 1995; Ozsogomonoyan & Loftus, 1979; Spencer, 1996). Given the importance of this attainment measure on students' career aspirations, final grades are highly relevant to students. Still, how is an introductory college science grade determined? A review of course syllabi shows that final grades are not a single measure, but a composite of several different measures including tests, quizzes, homework sets, and a comprehensive final examination, collected over months. This characteristic suggests that final grades are likely more indicative of student performance than a single attainment test.

Another issue to consider is over-reliance on knowledge and comprehension in assessing college grades to the detriment of higher-order thinking such as application, analysis, synthesis, and evaluation (Bloom, Hasting, & Madaus, 1971). Students learning science through inquiry-based activities and then under-performing in knowledge and comprehension activities would be no surprise, although this has been shown to not always be the case (Rivet & Krajcik, 2004). However, this study investigates the association of high school learning activities with college science performance. College-level content subsumes high school content, often in a matter of weeks. Therefore, students with a grasp of the fundamental topics at a deeper level when entering college science would certainly have an advantage over their peers. Students engaging in higher-order thinking through inquiry-based instructional methods in high school would be expected to possess a deeper understanding of concepts (Ausubel, 2000; Ramsden, 2003), and consequently outperform peers without this experience in college courses. Among these inquiry-based instructional methods are student-designed projects and laboratory experiences. Still, some skepticism regarding the use of college grades may remain about their comparability across institutions. To address this issue, college effects variables were included in the regression analysis to account for these differences (Pike & Saupe, 2002).

The inquiry-type learning activities predictors included Number of Student-designed Projects and Degree of Freedom in Designing/Conducting Labs. (See Figure 1 for question mock-ups.) The interactions included in the regression analysis were products of an attainment variable and an inquiry-type activity variable.

The academic attainment predictors included SAT-Quantitative and Verbal scores, Last High School grade in Mathematics, Science, and English, high school enrollment in calculus (regular, Advanced Placement® Calculus A/B, and Advanced Placement® Calculus B/C), and enrollment in Advanced Placement® science courses. Students reported the highest level of mathematics they took in high school; students selecting non-Calculus responses were coalesced into a single group, while students responding with any of the three calculus class choices were assigned to one of the three high school calculus enrollment categories. Students were also grouped according to whether they reported taking Advanced Placement® courses in the corresponding sciences. This predictor was used as a dummy variable. Students were also asked to report their SAT scores within particular groupings (200–300, 310–400, 410–500, etc.). For students with only American College Testing (ACT) scores, concordance tables were used to match ACT and SAT scores, both overall

Question 17:

How many projects of your own design did you carry out per year in your science* class?

○ None ○ 1 ○ 2 ○ 3 ○ More than 3

Question 25e:

Concerning the labs you conducted in science* class, please rank the following for a typical lab:

	None				Complete
Your freedom in conducting/designing the lab	①	②	③	④	⑤

* Actual surveys used discipline specific terms, biology, chemistry, and physics, rather than the general term

Figure 1. Mock-up of projects and laboratory questions

and mathematics (Dorans, Lyu, Pommerich, & Houston, 1997). Since ACT does not have a corresponding score to the SAT Verbal score, we calculated a total ACT score and matched it with an overall SAT score and then subtracted the matching SAT Mathematics score to produce an estimated SAT Verbal score. Last High School Grades in Science, Mathematics, and English were included in the models as numerical values assigned to letter grades (i.e., Grade A = 4, Grade B = 3, etc.).

Demographic background predictors included Race/Ethnicity, Gender, Highest Parent Educational Level, Year in College and Median Household Income by Zip Code. Past studies have shown the importance of these types of predictors (e.g., Bryk, Lee, & Holland, 1993; Burkam, Lee, & Smerdon, 1997). For Race/Ethnicity, six categories were used: not-reported, black, Hispanic, Asian, white, and multi-racial. Two survey questions were used to create this dummy variable set: racial grouping (black, white, etc.) and ethnicity (Hispanic, non-Hispanic). Gender was recorded as Male or Female; non-respondents were coded as a separate category and included in the analysis. Highest Parent Educational Level was a Likert-type scale variable with numerical values assigned to categories: Did not finish high school (= 1), High school, Some college, four years of college, and Graduate school. The highest value between the two parents was assigned to the variable. For a missing parent, the remaining parent's educational level was used. Year in college included four categories: Freshman, Sophomore, Junior, and Senior. The respondents selecting Graduate or Other were deleted. Missing values for categorical variables were also list-wise deleted from the analysis. Students were asked to provide home zip codes, and national data obtained from *Melissa Data*® were assigned for corresponding zip codes.

Results and Discussion

In this section, we begin with a descriptive analysis of the samples. Next, we discuss the results of the final multiple regression models, comparing similarities and differences. Finally, we calculate and discuss the predicted introductory course grades for prototypical students.

Descriptive Analysis

The variables included in this investigation were treated as either continuous or categorical. Descriptive statistics for the continuous variables are shown in Table 4 and included: Number of Student-designed Projects, Degree of Freedom in Designing/Conducting Labs (a Likert-scale item), SAT Score-Quantitative, SAT Score-Verbal, Last High School Grade in Mathematics, Science, English, and Highest Parental Education Level. The categorical predictor groups (Table 5) included Race/Ethnicity, Year in College, Advanced Placement® science enrollment, and High School calculus enrollment. Note that Native American and Multi-Racial categories had low representations; however, other categorical predictors had fairly large representations.

The descriptive statistics for the predictor, Number of Own Projects, are very similar across all three data-sets, with the average number of laboratories reported by students in all three disciplines at about one per high school course. However,

Table 4. Statistics for continuous predictors

	Minimum	Maximum	Biology, *M* (*SD*)	Chemistry, *M* (*SD*)	Physics, *M* (*SD*)
Number of Student-designed Projects (None = 0, More than 3 = 4)	0	4	1.2 (1.2)	1.0 (1.2)	1.2 (1.3)
Degree of Laboratory Freedom (None = 0, Complete = 4)	0	4	1.2 (1.2)	1.6 (1.3)	1.7 (1.2)
SAT Score					
Quantitative	220	790	580 (100)	590 (100)	620 (90)
Verbal	200	800	570 (100)	570 (100)	580 (100)
Last High School Grade in … (Grade A = 5, Grade F = 1)					
Mathematics	2	5	4.3 (0.8)	4.3 (0.8)	4.5 (0.7)
Science	1	5	4.4 (0.8)	4.4 (0.8)	4.5 (0.7)
English	1	5	4.6 (0.6)	4.6 (0.6)	4.6 (0.6)
Highest Parent Education Level (Attended High School = 0, Graduate School = 4)	0	4	2.8 (1.1)	2.7 (1.1)	2.9 (1.0)

Table 5. Statistics for categorical predictors

Predictor	Biology: Number of students	Biology: %	Chemistry: Number of students	Chemistry: %	Physics: Number of students	Physics: %
AP® science						
Did not enroll	2,428	88	3,155	90	1,747	92
Enrolled	326	12	366	10	156	8
High school calculus						
No high school calculus	1,775	65	2,000	57	842	44
Regular	373	14	518	15	351	18
AP® A/B	473	17	792	22	526	28
AP® B/C	133	5	211	6	184	10
Race/ethnicity						
Native American	33	1	36	1	27	1
Black	207	8	202	6	102	5
Hispanic	134	5	177	5	89	5
Asian	167	6	293	8	151	8
Multi-racial	73	3	93	3	38	2
Not Reported	53	2	95	3	106	6
White	2,087	76	2,625	75	1,390	73
Year in college						
Freshman	1,496	54	2,103	60	234	12
Sophomore	691	25	811	23	733	39
Junior	404	15	393	11	579	30
Senior	127	5	143	4	273	14
No Response	34	1	64	2	82	4

Note: AP®, Advanced Placement®.

Degree of Freedom in Designing/Conducting Labs appears to vary widely between biology students, who report an average rating of 1.2 (*SD* = 1.2), and chemistry and physics students, who report average ratings of 1.6 (*SD* =1.3) and 1.7 (*SD* =1.2), respectively. The biology ratings appear to be nearly one-half of a standard deviation below the other averages. This result suggests that chemistry and physics students appeared to have had more freedom in high school chemistry and physics courses than college biology students had in high school biology courses.

A look at the mathematics attainment measures shows some differences across the three samples. With respect to calculus background, 35.4% in biology, 43.2% in chemistry, and 55.8% in physics reported that they had enrolled in high school. The average SAT-Mathematics scores for biology, chemistry, and physics students were 580, 590, and 620, respectively. The corresponding percentile rankings of these three averages were 69 (biology), 72 (chemistry), and 79 (physics). For high school mathematics grades, biology and chemistry students' average high school grades were both 4.3 (Grade A = 5, Grade B = 4), while the average for physics students was 4.5, roughly one-quarter of a standard deviation higher. In general, introductory college physics students were higher mathematics achievers than introductory college biology and chemistry students.

A comparison of the averages for SAT-Verbal, Last High School Grades in Science and English, and Highest Parental Educational Level found them all to be very similar. For Advanced Placement® enrollment, 11.8% in biology, 10.4% in chemistry, and 8.2% in physics enrolled in the corresponding Advanced Placement® science course in high school. An analysis of the proportion of males versus females across the three disciplines revealed no surprises, with more females in biology and chemistry and more males in physics. The frequency distribution of the race/ethnicity predictors reveals students enrolled in introductory sciences are overwhelmingly white, with three-quarters of the sample in each discipline (see Table 5). For Year in College, biology and chemistry students were primarily freshman, while physics students were primarily sophomores. This difference may be due to the calculus co-requisite/pre-requisite for some physics courses.

Inferential Analysis

The final regression models are presented in Table 6. Our main focus was the investigation of interactive associations between operationalized attainment variables and measures of degree of structure in inquiry-type learning activities. A primary concern of this study is the replication of findings across the samples. In this analysis, the variables representing inquiry-type activities were treated as continuous for all three samples. Given this approach, identical procedures were used to analyse the samples. Our analytical approach called for each predictor of the inquiry-type learning activities to be entered singly into the models after the attainment variables were already entered. When entered separately, each inquiry-type activity predictor reported a significant negative association with college grades. When entered simultaneously, only one predictor remained significant. This result suggests that each predictor probably explained the same portion of variance in each sample, with one predictor subsuming the variance initially explained by the other. The analysis also found a significant interaction between Last High School Mathematics Grade and Degree of Lab Freedom for biology and chemistry, although not for physics.

For the biology and chemistry models, the parameter estimates (*B* in Table 6) are very similar in size and value. In addition, the standardized parameter estimates, β, for both models were also found to be very similar in proportion to the other predictors in the two models. The results present strikingly similar interactive associations. However, the results for the physics model appear to be in conflict. Here, no interactions were found to be significant, and the inquiry-type activity with greater significance was Number of Own Projects rather than Degree of Laboratory Freedom. However, recall that the average physics student appears to have higher overall mathematical attainment. This artifact suggests that fewer students with low or very low Last High School Mathematics Grades were in the physics sample. It seems that a reasonable argument for differences in average mathematics attainment between physics and the other two disciplines may be the result of students' self-selection. Students with weaker

Table 6. Final regression models

	Biology			Chemistry			Physics		
	B	*SE*	β	*B*	*SE*	β	*B*	*SE*	β
College effects		Included			Included			Included	
Race/ethnicity effects		Included			Included			Included	
Highest Parental Education Level		Included			Included			Included	
Year in college effects		Included			Included			Included	
Constant	34.46***	2.64		42.97***	2.22		44.73***	2.96	
SAT Score									
Quantitative	0.01***	0.00	0.12	0.02***	0.00	0.16	0.01***	0.00	0.12
Verbal	0.01***	0.00	0.11	0.004*	0.00	0.04	0.01*	0.00	0.06
AP® science Enrollment	2.71***	0.59	0.08	3.47***	0.55	0.10	3.30***	0.89	0.08
HS Calculus									
Regular	2.18***	0.58	0.07	2.14***	0.50	0.07	2.01**	0.66	0.07
AP® A/B	3.46***	0.57	0.12	3.26***	0.46	0.12	2.74***	0.65	0.11
AP® B/C	4.75***	0.93	0.09	4.37***	0.75	0.10	3.59***	0.94	0.09
Last HS Grade in									
Science	1.98***	0.29	0.13	1.72***	0.27	0.11	1.83***	0.42	0.11
English	2.15***	0.37	0.11	1.00**	0.31	0.05	0.96*	0.42	0.05
Mathematics	1.39***	0.41	0.09	2.07***	0.36	0.14	3.04***	0.37	0.19
Freedom In Lab Design	−2.59**	0.87	−0.29	−2.64***	0.76	−0.30			
Number of Student Projects							−0.47**	0.18	−0.05
Last High School Mathematics Grade x Freedom in Lab Design	0.42*	0.20	0.21	0.51**	0.17	0.26			

Table 6. *Continued*

	Biology			Chemistry			Physics		
	B	*SE*	β	*B*	*SE*	β	*B*	*SE*	β
R^2	0.355				0.337			0.363	
Adjusted R^2		0.340			0.325			0.335	
ΔR^2 (= $R^2_{\text{Final}} - R^2_{\text{Inquiry Not Included}}$)		0.021			0.014			0.003	
Number of Student Projects							−0.47**	0.18	−0.05
Last High School Mathematics Grade x Freedom in Lab Design	0.42*	0.20	0.21	0.51**	0.17	0.26			
n		2430			3187			1577	
Missing		319			334			326	

Notes: Dependent variable: GRADES, final college science grade. *B*, parameter estimates; β, standardized parameter estimates. $***p < 0.001$, $**p < 0.01$, $*p < 0.05$.

high school mathematics attainment may not choose college physics. With overall higher mathematics attainment, the lack of an interactive association is not surprising in the physics sample.

Next, the predicted GRADES for prototypical students were calculated and graphed. Multiple regression analyses produce linear equations (i.e., GRADES = $B_0 + B_1X_1 + B_2X_2 + \ldots + B_iX_i + B_jX_j + B_{ij}X_iX_j$), where the parameter estimates are the coefficients (B_0, B_1, …, B_i, B_j, B_{ij}) for variables (X_1 X_2, X_i, X_j, X_{ij}) whose ranges are presented in Table 4. The interactions are represented by the term $B_{ij}X_iX_j$. The calculation of predicted values for the outcome, GRADES, from prototypical values for the predictors are graphed in Figures 2–4. In these graphs, the following classifications for Last High School Mathematics Grade were used: A = *High*; B = *Moderate*; C = *Low*; D or F = *Very Low*. The lines in the graphs compared predicted GRADES for Degree of Laboratory Freedom from 'None' to 'Complete'. Grand means were entered for the remaining predictors. Although it appears that those reporting complete freedom in designing and conducting laboratories are at a disadvantage in

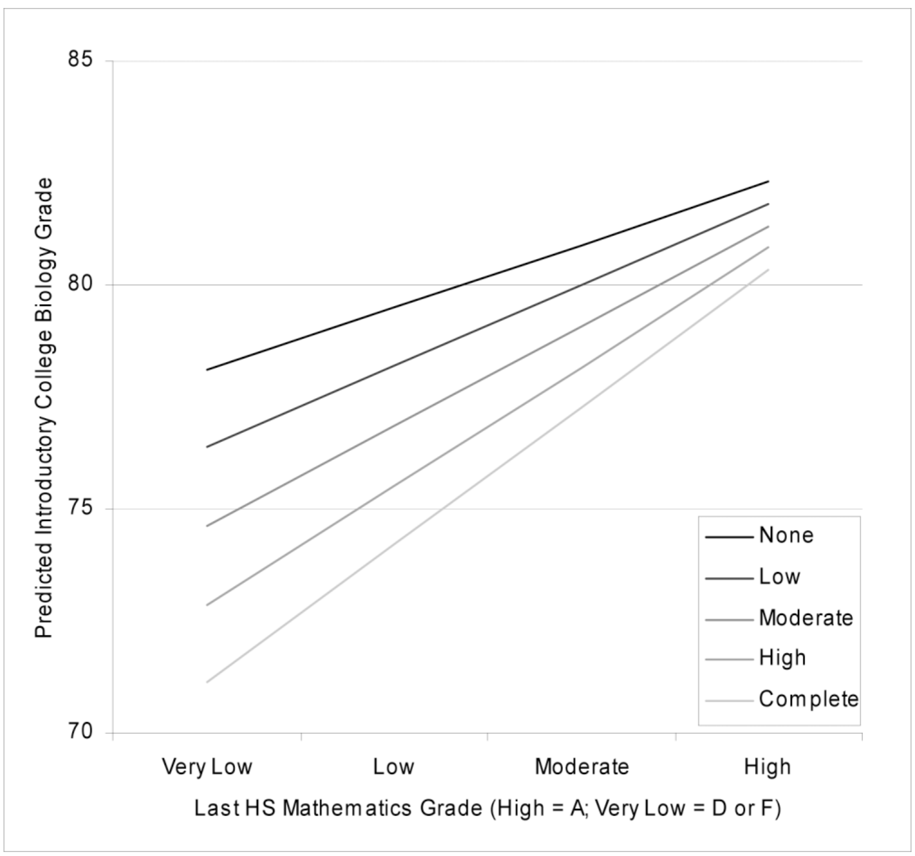

Figure 2. Interactive association of Degree of Lab Freedom and Last HS Mathematics Grade on Predicted Introductory College Biology Grade

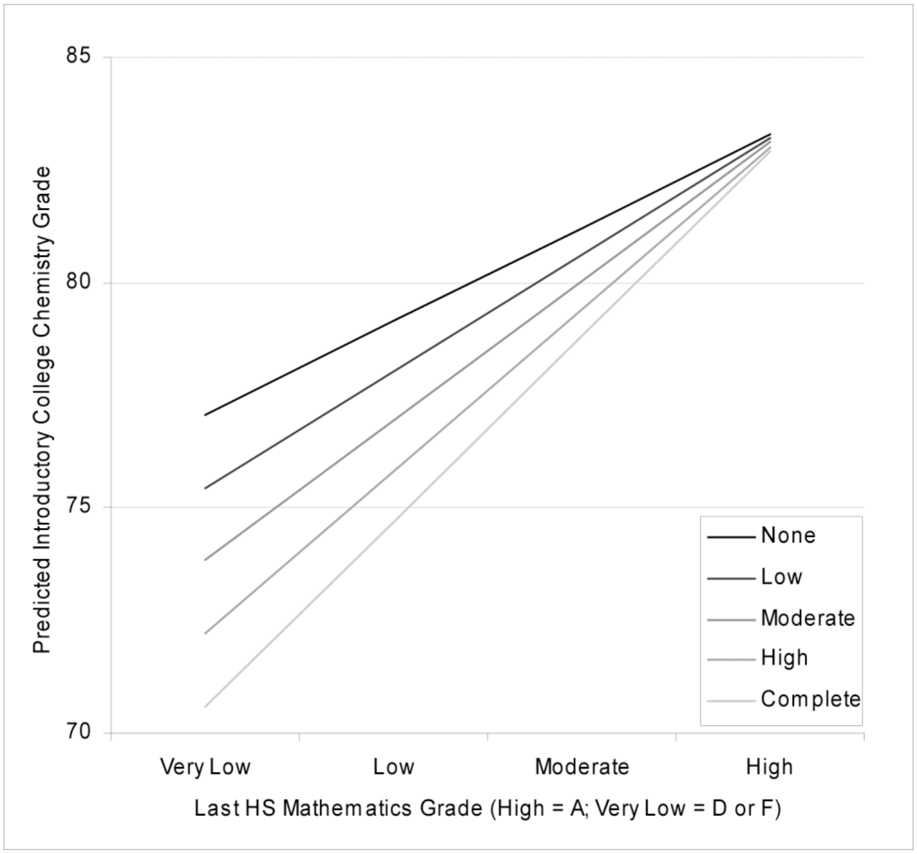

Figure 3. Interactive association of Degree of Lab Freedom and Last HS Mathematics Grade on Predicted Introductory College Chemistry Grade

all disciplines, a closer look showed that large differences appear mainly in the *Low* and *Very Low* Last High School Mathematics grade groups, ranging from 5.3 to 7.0 for biology students and from 4.4 to 6.5 for chemistry students (10 points represents an entire letter grade). The differences for the *High* and *Moderate* groups were modest, ranging from 2.0 to 3.6 for biology and from 0.4 to 2.4 for chemistry. *Low* and *Very Low* mathematics grade earners who reported experiencing Complete Lab Freedom are predicted to be at a substantial disadvantage later in college biology and chemistry. The results for the physics analysis show only a small 1.4 point difference, regardless of the Last High School Mathematics grade.

Conclusions, Limitations, and Directions for Further Study

These findings raise an important issue, if freedom in laboratory design is associated with lower college science performance, should this approach to instruction be abandoned in high school science courses? Abandoning freedom for students to

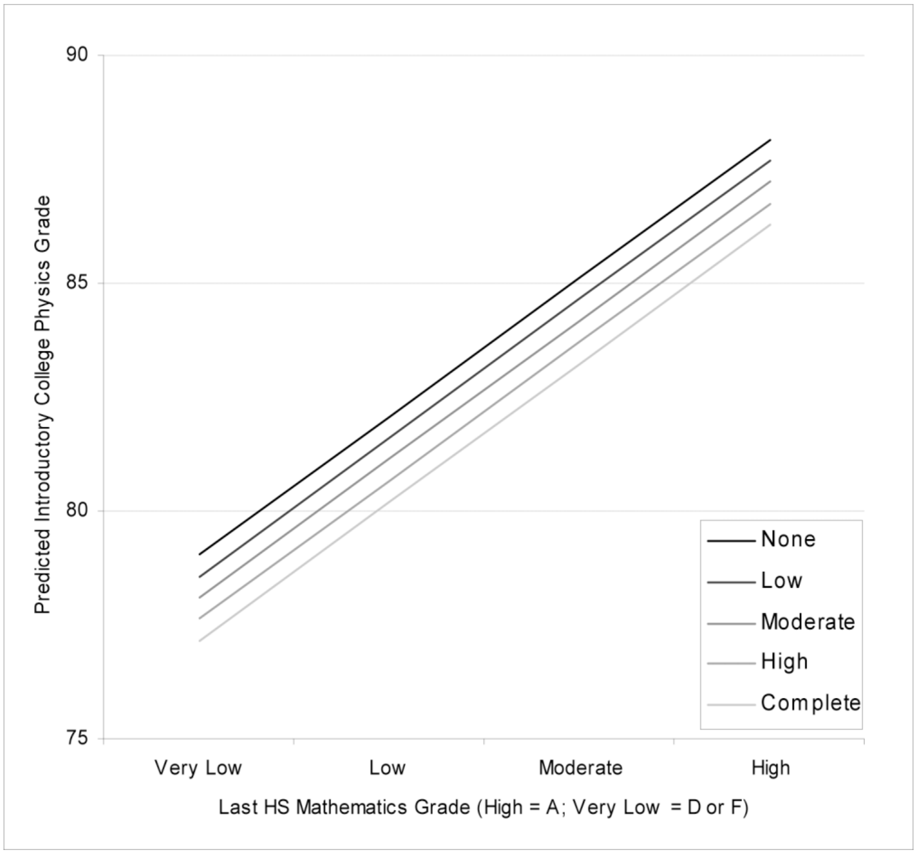

Figure 4. Degree of Lab Freedom and Last HS Mathematics Grade vs. Predicted Introductory College Physics Grade (no interactive association)

design and conduct their own experiments in science laboratoris is very short sighted. Autonomous learning is the seed of scientific research, a point expressed by Dudley Herschbach, Nobel Laureate in Chemistry:

> [W]hen it comes to research … no one knows the right answer at the outset, and often don't know the right approach. You don't have any body telling you what you need to know. So it's very different than what kids have been conditioned to…. [K]ids are handicapped if they're worried at the outset 'Am I doing this right? Am I going to get the right answer?' …. I often say to students, 'The neophyte and the veteran researcher are very much alike in one important respect, namely they're both confused most of the time. But the difference is the neophyte is uncomfortable … whereas the veteran researcher is happy…. If you're confused there's a chance that when you finally figure out what's going on, you will have learned something that's of interest to other researchers. (Interview with R.H. Tai, 31 October 2003)

Scientists must achieve some level of research independence on order to make contributions to human knowledge. Certainly, there have been questions raised about the authenticity of laboratory work as a reflection of scientific practise (Hodson, 1996).

However, school laboratory work is a tool used to teach both content and practise in scientific inquiry, thus a balance must be struck between structure and autonomy in inquiry-type learning activities. These results suggest that decisions on the degree of instructional structure should include student attainment.

While structured learning may be essential for building the knowledge-base necessary for understanding more advanced scientific concepts, autonomous learning forms the foundation of scientific inquiry. The findings from this study suggest that students with lower levels of high school mathematics attainment had greater success in college science when they reported more structured laboratory experiences. Students with higher high school mathematics attainment did not show much variation with differences in laboratory structure. These results corroborate and extend earlier aptitude treatment interaction research, suggesting that instructional experiences may have interactive associations with long-range impacts.

This study compelled us to think more deeply about teaching and learning science and the outcomes. Grades in introductory college sciences are but one possible outcome. Chosen for this study primarily because of its clear impact on students' career paths, other outcomes are also important to consider—enhancing students' science interest is one. While some forms of teaching may be highly effective in building students' background knowledge and enhancing science performance, other forms of teaching seem to raise students' interest and spark imagination enhancing their continuance in the study of science.[2] Performance and continuance may be two different dimensions of science paedagogy (Figure 5).

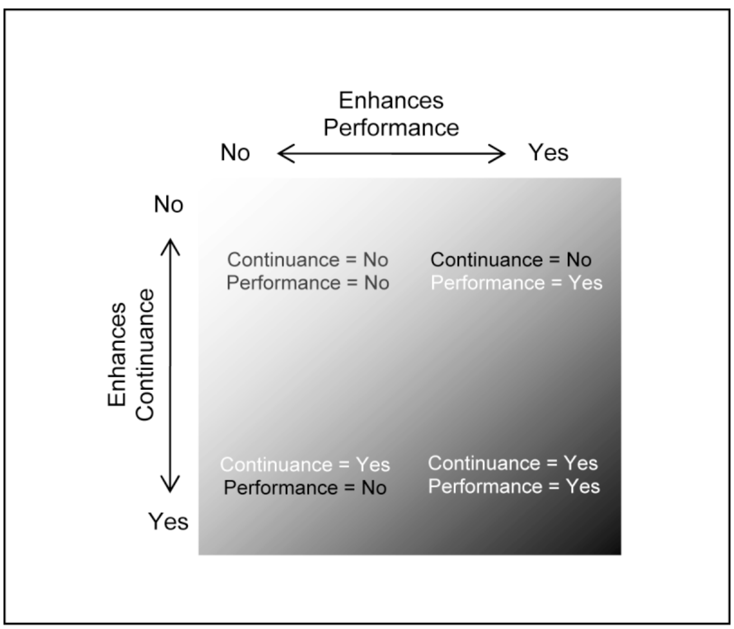

Figure 5. Diagram of paedagogical impact on performance and continuance in science learning

Some teaching methods may fall within the No/No region, being non-effective for either enhancing student performance or continuance. Others may fall in the Yes/No (enhancing performance, but not continuance) or No/Yes (enhancing continuance, but not performance) regions. Most interesting would be the identification of pedagogies within the Yes/Yes region, enhancing both performance and continuance. Using this theoretical construct, it is seems that neither dimension may be entirely ignored. Other studies have found important positive impacts of instructional methods on student interest and attitudes (e.g., Hofstein et al., 2004; O'Neill & Polman, 2004). Given the foundational nature of autonomous learning in relation to scientific research, the role of projects and laboratory freedom in furthering student interest in science study cannot be ignored. In light of current trends in the USA that focus on student achievement, further research into continuance and experiences with instructional methods is important.

Note

1. The comprehensive list of institutions was generated using the following website supported by the National Center for Educational Statistics of the US Department of Education, http://necs.ed.gov/collegenavigator/, accessed 4 December 2007.
2. The term 'continuance' is used rather than 'persistence' since this study captures only shorter-term course-taking rather than a long-term pursuit of an educational goal (e.g., Seymour & Hewitt, 1997).

References

Allison. P.D. (2002). *Missing data: Quantitative applications in the social sciences.* Thousand Oaks, CA: Sage Publications.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy.* Oxford, UK: Oxford University Press.

Ausubel, D.P. (2000). *The acquisition and retention of knowledge: A cognitive view.* Dordrecht, The Netherlands: Kluwer Academic.

Bloom, B.S., Hasting, J.T., & Madaus, G.F. (1971). *Handbook on formative and summative evaluation of student learning.* New York: McGraw-Hill.

Bradburn, N.M. (2000). Temporal representation and event dating. In A.A. Stone, J.S. Turkkan, C.A. Bachrach, J.B. Jobe, H.S. Kurtzman, & V.S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 49–61). Mahwah, NJ: Lawrence Erlbaum Associates.

Bradburn, N.M., Rips, L.J., & Shevell, S.K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science, 236,* 157–161.

Brown, A.L., & Campione, J.C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229–270). Cambridge, MA: Massachusetts Institute of Technology.

Bryk, A.S., Lee, V.E., & Holland, P.B. (1993). *Catholic schools and the common good.* Cambridge, MA: Harvard University Press.

Bureau of Labor Statistics. (2005, March). *News release: College enrollment and work activity of 2004 high school graduates* (USDL 05-487). Retrieved July 25, 2005, from http://www.bnls.gov/cps.

Burkam, D.T., Lee, V.E., & Smerdon, B.A. (1997). Gender and science learning early in high school: Subject matter and laboratory experiences. *American Educational Research Journal, 34*(2), 297–331.

Cronbach, L.J., & Snow, R.E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions.* New York: Irvington Publishers.

Dorans, N.J., Lyu, C.F., Pommerich, M., & Houston, W.M. (1997). Concordance between ACT Assessment and recentered SAT I sum scores. *College and University, 73*(2), 24–35.

Eysenck, H.J. (1996). Personality and the experimental study of education. *European Journal of Personality, 10,* 427–439.

Gainen, J. & Willemsen, E.W. (Eds.). (1995). *Fostering student success in quantitative gateway courses* (No. 61). San Francisco: Jossey-Bass.

Groves, R.M. (1989). *Survey errors and survey costs.* New York: J. Wiley and Sons.

Hodson, D. (1996). Laboratory work as scientific method: Three decades of confusion and distortion. *Journal of Curriculum Studies, 28*(2), 115–135.

Hoffer, T.B., Quinn, P., & Suter, L.E. (1996). *High school seniors' instructional experiences in science and mathematics: National Educational Longitudinal Study of 1988* (NCES 95-278). Washington, DC: Government Printing Office.

Hofstein, A., Shore, R., & Kipnis, M. (2004). Providing high school chemistry students with opportunities to develop learning skills in an inquiry-type laboratory: A case study. *International Journal of Science Education, 26*(1), 47–62.

Kummings, K., Marx, J., Thornton, R., & Kuhl, D. (1999). Evaluating innovation in studio physics. *American Journal of Physics, 67*(7), S38–S44.

Kuncel, N.R., Credé, M., & Thomas, L.L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75*(1), 63–82.

Light, R.J., Singer, J.D., & Willett, J.B. (1990). *By design: Planning research on higher education.* Cambridge, MA: Harvard University Press.

Little, R.J.A., & Rubin, D.B. (2002). *Statistical analysis with missing data.* New York: Wiley.

Menon, G., & Yorkston, E.A. (2000). The use of memory and contextual cues in the formation of behavioral frequency judgments. In A.A. Stone, J.S. Turkkan, C.A. Bachrach, J.B. Jobe, H.S. Kurtzman, & V.S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 63–79). Mahwah, NJ: Earlbaum.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning.* Washington, DC: National Academy Press.

Niemi, R.G., & Smith, J. (2003). The accuracy of students' reports of course taking in the 1994 National Assessment of Educational Progress. *Educational Measurement: Issues and Practice, 22*(1), 15–21.

O'Neill, D.K., & Polman, J.L. (2004). Why educate 'little scientists?' Examining the potential of practice-based scientific literacy. *Journal of Research in Science Teaching, 41*(3), 234–266.

Ozsogomonoyan, A., & Loftus, D. (1979). Predictors of general chemistry grades. *Journal of Chemical Education, 55*(3), 173–175.

Pike, G.R., & Saupe, J.L. (2002). Does high school matter? An analysis of three methods of predicting first-year grades. *Research in Higher Education, 43*(2), 187–207.

Polman, J.L. (2000). *Designing project-based science: Connecting learners through guided inquiry.* New York: Teachers College Press.

Peugh, J.L., & Enders, C.K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74*(4), 525–556.

Ramsden, P. (2003). *Learning to teach in higher education.* London: Routledge Falmer.

Rivet, A.E., & Krajcik, J.S. (2004). Achieving standard in urban systemic reform: An example of a sixth grade project-based science curriculum. *Journal of Research in Science Teaching, 37*(2), 139–159.

Sadler, P.M., & Tai, R.H. (2001). Success in college physics: The role of high school preparation. *Science Education, 85*(2), 111–136.

Sawyer, R., Laing, J., & Houston, M. (1988). *Accuracy of self-reported high school courses and grades of college-bound students.* ACT research Report Series 88-1. Iowa City, IA: American College Testing Program.

Scheffer, J. (2002). Dealing with missing data. *Research Letters in the Information and Mathematical Sciences, 3,* 153–160.

Schiel, J., & Noble, J. (1991). *Accuracy of self-reported course work and grade information of high school sophomores.* ACT Research Report Series 91–6. Iowa City, IA: American College Testing Program.

Schneider, R.M., Krajcik, J., Marx, R.W., & Soloway, E. (2002). Peformance of students in project-based science classrooms on a national measure of science assessment. *Journal of Research of Science Teaching, 39*(5), 410–422.

Seymour, E. & Hewitt, N.M. (1997). *Talking about leaving: Why undergraduates leave the sciences.* Boulder, CO: Westview Press.

Spencer, H.E. (1996). Mathematical SAT test scores and college chemistry grades. *Journal of Chemical Education, 73*(12), 1150–1153.

Tai, R.H., & Sadler, P.M. (2001) Gender differences in introductory undergraduate physics performance: University physics versus college physics in the United States. *International Journal of Science Education, 23*(10), 1017–1037.

Thorndike, R.M. (1997). *Measurement and evaluation in psychology and education* (6th ed.) Upper Saddle River, NJ: Merrill.

Tobias, S. (1981). Adapting instruction to individual difference among students. *Educational Psychologist, 61*(2), 111–120.

Tytler, R. (1992). Independent research projects in school science: Case studies of autonomous behaviour. *International Journal of Science Education, 14*(4), 393–411.

Valiga, M.J. (1987). *The accuracy of self-reported high school course and grade information.* ACT Research Report Series 87-1. Iowa City, IA: American College Testing Program.

Wallace, C.S., Tsoi, M.Y., Calkin, J., & Darley, M. (2003). Learning from inquiry-based laboratories in nonmajor biology: An interpretive study of the relationships among inquiry experience, epistemologies, and conceptual growth. *Journal of Research in Science Teaching, 40*(10), 986–1024.

Windschitl, M., & Andre, T. (1998). Using computer simulations to enhance conceptual change: The roles of constructivist instruction and student epistemological beliefs. *Journal of Research in Science Teaching, 35*(2), 145–160.