Factors Influencing Success in Introductory College Chemistry

Robert H. Tai,[1] Philip M. Sadler,[2] John F. Loehr[1]

[1]*Curry School of Education, University of Virginia, 405 Emmet Street South, Charlottesville, Virginia 22904*

[2]*Science Education Department, Philips Auditorium, Harvard–Smithsonian Center for Astrophysics, Cambridge, Massachusetts*

Received 23 March 2004; Accepted 13 December 2004

Abstract: Previous research has found a wide range of predictors of student performance in introductory college chemistry. These predictors are associated with both the students' backgrounds and their high school learning experiences. The purpose of this research study was to examine the link between high school chemistry pedagogical experiences and performance in introductory college chemistry while accounting for individual educational and demographic differences. The researchers surveyed 1531 students enrolled in first-semester introductory college chemistry courses for science and engineering majors at 12 different U.S. colleges and universities. Using multiple regression analysis, the researchers uncovered several interesting high school pedagogical experiences that appeared to be linked with varying levels of performance in college chemistry. Most notably, the researchers found that repeating chemistry labs for understanding was associated with higher student grades, whereas overemphasis on lab procedure in high school chemistry was associated with lower grades in college. These results suggest that high school teachers' pedagogical choices may have a link to future student performance.

Within the past decade, the quality of science education in the United States has come under increased scrutiny. Results from the Third International Mathematics and Science Study (TIMSS) placed U.S. high school students' science literacy in the lower third of countries included in the Final Year of Secondary School Survey (Mullis, Martin, Beaton, Gonzalez, Kelly, & Smith, 1998, p. 33). The subsequent uproar in response to these results has been further fueled by more recent findings from American College Testing (ACT, 2003), suggesting that, although standardized test scores have risen in the past few years, mathematics and science preparation in high school appears to be weak. As a result, many high school students founder when they undertake college

Contract grant sponsor: Interagency Educational Research Initiative; Contract grant number: National Science Foundation REC 0115649.
Correspondence to: R.H. Tai; E-mail: rht6h@virginia.edu
DOI 10.1002/tea.20082
Published online 9 May 2005 in Wiley InterScience (www.interscience.wiley.com).

studies. However, many other students do find great success in college science courses. The identification of factors that relate to science success or failure of students may provide useful clues for promoting science learning and achievement. Previous research suggests that influential factors fall within three categories: demographic background; general educational background; and previous science learning experiences. Demographic background and general educational background are indicators of social circumstance and prior academic achievement, respectively, and although of interest, are beyond the influence of high school science teachers. It is the high school science learning experiences over which science teachers have control, although some teachers more than others. It is these experiences that are closely linked to the pedagogical approaches chosen by these teachers. Identification of pedagogical practices in science classrooms that connect to subsequent success in science learning and that suggest influential links between science teaching in high school and science success in college may be beneficial in efforts to improve science achievement.

This study applies an epidemiological approach to the study of the question of success in introductory college science courses as it relates to high school science pedagogical experiences controlling for demographic and general educational background differences among the students. The purpose of this study is to uncover pedagogical practices used by high school science teachers that are associated with successful performance in college science of science students while accounting for differences among these students in terms of their demographic and general educational backgrounds. Whereas statistical models are used to uncover associations between the outcome and the predictors, the most critical step in this methodological approach is in the interpretation of the findings. The individual significant predictors are used to sketch a model of the high school chemistry experience and its association with success in college chemistry coursework. The approach is not merely to identify single variables and suggest that differences within these particular experiences define a model of success. Rather, this approach uses single variables as informants on the whole experiences of students in high school chemistry, and uses these single variables to construct a broader and more complex picture of student experience and college success. This approach has the potential to address critical questions: Could specific teaching and learning methods used in high school contribute to the success of students over and beyond the students' backgrounds? Could certain teaching and learning techniques experienced by some students give them an advantage over their peers with similar backgrounds? This study seeks to provide some clues in linking high school science learning to success in college science.

For this research study, we have chosen to focus on student preparation and performance in the area of chemistry. Chemistry is commonly viewed as the "central science," as mastery of its concepts regarding the structure of matter is essential to further coursework in all sciences. Evidence for this belief lies in the order of coursework required at many major universities in the United States. For example, the University of Illinois at Urbana–Champaign, requires that prospective majors in sciences other than chemistry take courses in chemistry prior to beginning coursework in their majors. Because chemistry courses are in many cases the first science courses taken by students at the college level, it is quite possible that prospective biology majors are deterred from taking biology courses because of their experiences in introductory college chemistry (ICC). In essence, chemistry performs the function of gatekeeper for future study in many sciences. Thus, it would appear that preparation for chemistry at the college level would be an important consideration and, in most cases, this preparation would take the form of high school (HS) chemistry coursework. The study described herein focuses directly on the connection between teaching practices in high school chemistry courses and the real-world measure of students' performance in ICC coursework.

The experiences of students in HS chemistry courses around the United States are many and varied, in large measure due to the decisions made by their HS chemistry teachers. These HS chemistry teachers must respond to several forces impinging upon their decisions. First among these forces is their own background. According to Weiss, Pasley, Smith, Banilower, and Heck (2003), who interviewed and observed 61 U.S. high school science teachers, the greatest influence on choices regarding instructional strategies is teacher knowledge, belief, and experience, influencing 90% of lessons taught by the teachers in their study. As one of their participants reported, the decision-making most often relies on ''. . .just trial and error over the years. Trying to get a feel for what works and what does not work'' (Weiss et al., 2003, p. 13). This confirms findings from earlier studies in which teachers selected learning experiences because these experiences were either helpful to them as students or corresponded with prior experiences (Aikenhead, 1984; Duschl & Wright, 1989).

Research studies directly addressing the issue of high school preparation in connection with college performance have a long history in chemistry and have spanned other content areas including biology and physics. Studies connecting high school chemistry with college chemistry have been published in long-running journals such as *School Science and Mathematics* and the *Journal of Chemical Education* since the 1920s (e.g., Brasted, 1957; Everhart & Ebaugh, 1925; West, 1932). Ogden (1976) reviewed 24 studies published prior to 1967 and concluded the following:

> There is some indication that the taking of high school chemistry may be used as an indicator of success in college chemistry. There are indications that a math/physics background, high placement scores, achievement tests scores, intelligence, and age may be a better, or at least as good, as indicators. There is also evidence that no indicator is all that good. (p. 125)

More recent studies on chemistry (e.g., Nordstrom, 1990; Yager, Snider, & Krajcik, 1988) and other fields (e.g., Alters, 1995; Gibson & Gibson, 1993; Harpole & Gifford, 1986; Hart & Cottle, 1993; Nordstrom, 1990) have yielded findings echoing Ogden's conclusions. These findings suggest that a more comprehensive comparative model is necessary to sort out the effects and influences of the various indicators. A comparative review of the various analytical methods used in each of these studies reveals that they fall into three categories: small-scale studies of unique educational programs, single institution survey studies, and multiple-institution survey studies.

An example of a small-scale study of a unique science program was done by Yager, Snider, and Krajcik (1986), who investigated the performance of high-ability HS students in ICC summer courses. The study compared students with and without prior HS chemistry coursework. After participating in the program, students from both groups showed similar performance on all assessments used. However, further investigation revealed differences in the amount of time and effort that students expended studying. In particular, students without HS chemistry science backgrounds sought out many more hours of tutoring than their peers with HS chemistry backgrounds. The importance of small-scale studies lies primarily in their ability to provide important insights into the mechanisms that may be at work in the transition between HS science experiences and college science performance.

The single-institution research surveys have also produced some interesting results. Hart and Cottle, and later Alters, studied physics students and concluded students who had taken HS physics performed better than their peers in introductory college coursework. Gifford and Harpole found similar results in their study of 248 physics students at Mississippi State University. Nordstrom (1990) carried out a 10-year study at Embry–Riddle Aeronautical University between

1980–1989 that included 980 chemistry students. He found students earning a C or higher in chemistry had higher SAT/ACT scores, HS grade point average (GPA), HS chemistry grade, HS math grade, and HS English grade than their peers.

Both these small-scale and single-institution survey studies have some advantages that allow for the identification of several factors that may play a role in the wider population. However, sampling numbers used in each of these studies limit the generalizability of the outcomes. A multiple-institution survey study provides the advantage of having a larger sampling of the overall population and therefore strengthens the argument for generalizability. The authors (Sadler & Tai, 2001; Tai & Sadler, 2001) performed such a study in 1994, looking at the connection between high school physics experiences and introductory college physics performance. The study included 1933 students from 18 college and universities. Not surprisingly, a strong connection between HS mathematics course-taking and college physics performance was discovered. The study revealed that students with a background in calculus fared better in both calculus-based and algebra-based introductory college physics courses, while accounting for other measures of academic performance. High school physics courses appeared to make a much smaller contribution on college physics course grades. Those students studying fewer topics in greater depth in high school physics had higher grades in college physics, on average, than students studying greater numbers of topics in less depth. A perplexing finding was that laboratory experiences appeared to be negatively associated with college physics performance. This result runs contrary to the conventional wisdom that laboratory experiences are beneficial to students in their subsequent performance in future coursework. The lessons learned from our prior study have been applied to this current study on chemistry. The preceding physics study included a survey questionnaire with 57 questions. Based on the results of this work, the current study used a survey questionnaire with 172 questions. This type of multiple-institutional study provides some needed scope and lends credibility to the generalization of the results to many more situations than would be possible with the other two approaches.

The purpose of the current study is to extend an epidemiological approach to an analysis of the connection between high school chemistry pedagogical experiences and introductory college chemistry performance in courses for science and engineering majors. The weaknesses in science preparation suggested by the results of TIMSS and the ACT study mentioned earlier in this discussion are directly related to high school science preparation. However, the measures used by both of these investigations reflect a reliance on the validity of the standardized exams as predictors of student performance and thus student preparation. In addition, neither of these studies includes the level of detail in their surveys that would allow researchers to distinguish variations in pedagogical practice that may very likely provide information useful to educators. It is the intention of this survey to explore the connection between high school chemistry learning experiences and introductory college chemistry performance. More than two-thirds of the 172 survey items included in the chemistry survey used in this study focus on high school chemistry pedagogy as it applies to the use of demonstrations, laboratory activities, classroom discussions, etc. This approach allows the researchers to paint a more detailed picture of the influence that particular HS pedagogical approaches chemistry teachers may favor in their classrooms have on their students' eventual performance in college. No previous research study has coupled the level of detail regarding chemistry instruction with the breadth of sampling used in this study. The research question guiding the analysis is as follows:

> What aspects of high school chemistry pedagogical experiences are most closely associated with successful performance in introductory college chemistry courses for science and engineering majors at 4-year colleges and universities in the United States when accounting for differences in students' demographic and general educational backgrounds?

Methods

This research study of chemistry instruction is an outgrowth of a larger study that includes biology, chemistry, and physics. Factors Influencing College Science Success (Project FICSS) is a grant-funded national study (grant no. NSF-REC 0115649) that includes interviews and surveys of college science students, high school science teachers, and professors in biology, chemistry, and physics programs from over 100 different 4-year colleges and universities selected through a stratified random sample from a comprehensive list of schools in the United States that vary in enrollment from under 5000 to over 20,000 students. To control for differences in college learning experiences, only schools using lecture-based classes were included in the study. These lecture-based courses typically followed a lecture/recitation/lab format where students would attend classes in large lecture halls led by a professor and then subsequently attend smaller recitation classes where the content of the lecture was reviewed and student questions were addressed. These sections were led by graduate students, advanced undergraduates, and, in some instances, professors. The labs were held in separate class meetings. Our initial impression was that small liberal arts colleges might have smaller class sizes and pedagogical approaches more focused on, for example, inquiry teaching practices. Despite our initial concerns regarding major differences in instructional approaches between large public universities and smaller liberal arts colleges, we found that the lecture/recitation/lab format was common at colleges and universities of all sizes. We discovered that the number of small liberal arts colleges in the United States was very large, and yet the teaching practices at the introductory level are not notably different from larger universities in the random sample of colleges participating in this study. Project FICSS is currently engaged in data collection at the time of the writing of this article. The data set analyzed herein is a subsample drawn from a pilot study that was carried out to gauge the challenges to be faced in the larger undertaking.

Sample and Survey

The data set was collected from 12 introductory college chemistry courses required for science and engineering majors at 12 different 4-year colleges and universities from across the United States in Fall 2002. The 12 schools are a subset of the schools chosen randomly from a comprehensive list of 4-year colleges in the United States. These 12 schools agreed to participate in a larger study and could accommodate the logistics of the chemistry pilot survey given differences in academic calendars. Other schools agreed to participate in the larger survey, but due to an earlier conclusion to their academic term could not participate in the chemistry pilot survey. Taking care to compare academic calendars, we found no systematic biasing connecting dates chosen to end the Fall 2002 term with the characteristics of schools we used to stratify the larger sample. Characteristics for each of the 12 schools included in the chemistry pilot survey are listed in Table 1. This sample included two private colleges and ten public institutions from ten states. The data set includes schools from the entire range of student enrollments, ranging in size from small liberal arts colleges to large state universities. The institutions also ranged in their selectivity of admissions. (Because the study is ongoing, we have chosen not to reveal the names of the participating institutions.) A total of 1531 student surveys were analyzed in this study. To control for differences in college learning experiences, only schools using the lecture/recitation/lab format were asked to participate. This college course format appears to be the most widespread. Table 2 shows some demographic and educational background details of the overall sample.

Survey Instrument

The survey instrument was designed based on previous research efforts, a review of relevant literature, and interviews of high school teachers and college professors. As mentioned earlier, the current survey was based on the survey used in a previous study on physics students. All the survey questions were developed and tested in earlier research and pre-pilot surveys. The question development also included information garnered from interviews of 22 classroom teachers from across the United States, which was carried out by Marc Schwartz of McGill University and 20 college professors from various colleges and universities carried out by Zahra Hazari of the University of Toronto, both collaborators in Project FICSS. Both studies were done in 2002. The current chemistry survey included questions addressing some 172 separate variables. Of these 172 variables, 87 survey questions dealt with issues associated with various aspects of high school chemistry classrooms. Details of these variables are discussed later in this investigation. The survey was in the form of a four-sheet booklet, the first sheet (a coversheet) included the title of the survey and information identifying the project and the university affiliation, blank on the reverse side. The remaining three sheets were all double-sided with questions formatted for use with an optical scanner for precise and quick recording of the data and identification numbers to allow sheets from the same survey to be matched should they become separated. The coversheet was detached and discarded upon return of the survey to the researchers. The survey also contained a space that allowed for follow-up contact of respondents through electronic mail, should they wish to provide this information. Of our sample, roughly 25% of the respondents volunteered their e-mail addresses. No specific analysis in this study is individually identifiable.

Limitations in Retrospective Self-Report Surveys

The design of this study relied on retrospection from the participants regarding their high school chemistry experiences. The data were collected through questionnaires distributed to the students in introductory college chemistry courses for science and engineering majors and completed during the class sessions. The responses to the surveys consisted of student self-reports of their demographic and educational background and high school pedagogical experiences. The authors acknowledge that this method poses challenges with respect to the accuracy of the high school chemistry learning experiences reported by the students. For this reason, we turned to work done by researchers who have studied this particular issue to address issues of reliability and validity. The existing research makes specific suggestions for improving the accuracy of these types of surveys and these suggestions were employed in the survey design and administration. Bradburn, Rips, and Shevell (1987) summarized the findings of several research articles regarding the answering of autobiographical questions and concluded that coalescing questions in surveys into sequences appears to have a positive effect on participant recall of events. Although errors in recall do persist, their work on the ‘‘forgetting’’ curve shows a nonlinear relationship over time where forgetting increases sharply for several weeks after the reference point before flattening out into a linear function over time. However, Bradburn et al. suggested that ‘‘recall usually improves if a respondent has appropriate cues, although different types of cues vary in effectiveness’’ (1987, p. 158). Groves (1989) in his review of work by several researchers, including Bradburn (Cannell, Fowler, & Marquis, 1965; Linton, 1982; Sudman & Bradburn, 1973), suggested that the shape of the ‘‘forgetting’’ curve is ‘‘almost linear, monotonically increasing over time until the period quite near the starting date of the reference point’’ (p. 426). In more recent work, Bradburn (2000) shifted his stance to a more hopeful one, concluding that memory and recall can be quite reliable, especially with contextual cues. Menon and Yorkston’s (2000) work on the ‘‘forgetting’’ curve, revealed findings suggesting that memory loss occurs at a rate of 3% each year in their 12-year study. Despite expected levels of inaccuracy in recall, researchers in public health routinely construct, administer, analyze, and draw conclusions from self-report surveys, many on sensitive topics such as drug use (e.g., Oetting & Beauvais, 1990) and less sensitive issues such as frequency of exercise, smoking, and alcohol consumption (e.g., Stengel, Tarver-Car, Powe, Eberhardt, & Brancati, 2003).

The authors accounted for the various suggestions made by survey researchers in both questionnaire design and survey implementation. The questionnaire format grouped related questions and paid particular attention to the sequencing of the questions within these relevant groupings; for instance, laboratory experiences, homework experiences, and teacher characteristics. The survey was administered during an introductory college chemistry course session taking advantage of contextual cues suggested by survey researchers as helpful in improving recall. The form and setting of the survey lays the groundwork for a ''best-case'' scenario for accurate recall of past experiences directly relevant to each individual's situation.

Although memory loss and inaccuracy in recall do present challenges, there are factors that work to the advantage of this study. It is very likely that students in a college chemistry class have strong impressions of their high school chemistry courses, because the intent of much high school chemistry instruction was preparation for college chemistry. It would also seem that a student's memory of past high school chemistry learning experiences would never be more clear than in an introductory college chemistry course in the present, especially since their future career plans likely hang in the balance given the intended or tacit ''gatekeeping'' function of this introductory chemistry course (Gainen, 1995). In short, enrollment in a college chemistry course would serve to improve the recall of past experiences related to chemistry learning. It should be noted that this study is primarily interested in articulating connections between high school chemistry instruction and introductory college chemistry instruction. As such, the instruction that students received in high school that made the strongest impressions is precisely the impressions this study is most interested in gauging and also would likely be the experiences students would recall most clearly.

One particular aspect of question and survey design the researchers assiduously avoided was asking students questions regarding attitudes. Many studies examining the relationship between attitudes and performance (Debaz, 1994; House, 1993; Lawrenz, 1976) have concluded that attitudes should generally be treated as outcomes measures. It appears that performance would better predict attitude, rather than attitude predicting performance.

Outcome

When selecting an outcome measure of student success, several factors come into consideration. Validity and overall acceptance of this outcome as a measure of success are two major considerations. In turning to the existing literature connecting high school science experiences and college science performance, the outcome measure that consistently arises is course grade. Studies as early as 1921 have relied on course grades as outcomes (e.g., Brasted, 1957; Hermann, 1931; McQuary, Williams, & Willard, 1952; Powers, 1921). More recent works have also used grades as a measure of performance (Alters, 1995; Hart & Cottle, 1993; Nordstrom, 1990; Sadler and Tai, 2001; Tai & Sadler, 2001). Apart from this approach having been taken by other researchers in the past, there are several reasons for choosing grades. One important reason behind the selection of this outcome lies in the fact that course grades represent the fulfillment of certain course standards listed in syllabi and known to all enrolled students. The course standards include various types of assessments such as passing examinations and writing lab reports; therefore, course grades serve as a summative evaluation of student performance. This conclusion supports the claim of validity. In addition, employers and admissions committees frequently use course grades to gauge student past performance and predict future success. This fact is well understood among students and supports the claim of overall acceptance.

An important step is reconciling the different grading schemes used by the professors participating in our study. The approach taken in this analysis was based the approach used in a previous study on introductory college physics performance (Sadler & Tai, 2001). Some professors used only whole-letter grades (A, B, C, etc.), others used pluses and minuses (A, A−, B+, B, etc.), and still others used point accumulation scales. All of these different final grading schemes were converted to a 100-point scale, where: A+ = 98; A = 95; A− = 91; B+ = 88; B = 85; B− = 81, etc. Figure 1 shows a graph of the ICCGRADE distribution. The average grade was 81.8 out of 100, with a standard deviation of 10. These results are nearly identical to the previous research (Tai & Sadler, 2001, p. 118). In addition, the expected variation in grades across institutions was accounted by using indicator variables for each college and university. This approach allows for adjustments due to grade inflation or deflation across different courses. The baseline school in this analysis contributed 513 surveys to this study.

Predictors

The predictors in this study fell into three categories: demographic predictors, educational background predictors, and HS chemistry pedagogical experience predictors. The demographic predictors included in the analysis were: parent's education, gender, race, year in college, HS type, HS graduating class size, and HS chemistry class size. The educational background predictors analyzed were: HS grades in science, math, and English courses, math courses taken, number of HS chemistry courses taken, and length and frequency of HS chemistry classes. These variables served as control predictors, accounting for demographic and educational differences that previous educational studies have reported to be significant (Brasted, 1957; Brickhouse, Lowery,

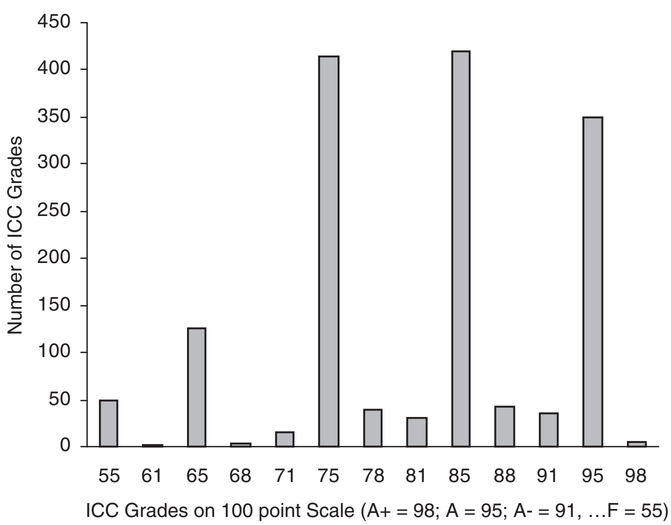

Figure 1. Distribution of introductory college chemistry course grades, ICCGRADE.

& Schultz, 2000; Burkam, Lee, & Smerdon, 1997; Catsambis, 1995; Deboer, 1987; Gibson & Gibson, 1993; Hendricks, Koelsche, & Bledsoe, 1963; Lamb, Waggoner, & Finley, 1967; Ogden, 1976; Reynolds & Walberg, 1992).

In considering the question predictors to be included in the survey the researchers turned to existing research and also carried out research that included interviews of practicing classroom teachers. Existing literature (e.g., Ausubel, Novak, & Hanesian, 1978; Hegarty-Hazel, 1990; Ramsden, 2003; Shumba & Glass, 1994; Yager, Snider, & Krajcik, 1988; Wellington, 1998; Woolnough, 1991), spanning decades, has frequently included studies and reviews discussing the impact of particular teaching practices on student learning. In addition to reviewing the literature, the researchers gathered data regarding teachers' impressions of effective college preparatory pedagogy in science through interviews with 22 science teachers from Florida, Massachusetts, and California (Schwartz, Hazari, & Sadler, unpublished manuscript). This study specifically focused on the pedagogy that these teachers used in their classrooms. We identified several aspects of teaching science that teachers frequently consider when planning and implementing their lessons: instructional techniques (e.g., lectures, class discussions, group work, etc.), demonstrations, laboratory experiences, student-designed projects and work, homework, textbook use, high school chemistry content, daily assignments, exams, course content, and characteristics of their high school chemistry teachers. The pedagogical questions included on the questionnaire asked students to provide two types of responses: fact recall and degree-of-experience ratings of a given pedagogy. Examples of fact recall questions and degree-of-experience rating questions are shown in Figure 2.

To address the issue of response reliability on these survey questions, a study was carried out with the questionnaire. Students from a first-semester (Fall) chemistry course for science and engineering majors from a large public university were asked to participate and 113 students volunteered. A small honorarium was paid to the participants for completing the survey in two separate administrations 2 weeks apart. The results of this reliability study revealed that, on average, for the questions analyzed in this study, 90.7% of the students responded with at least an adjacent choice and 60.0% responded with exactly the same response. An example of an adjacent choice would be if a particular student responded to Question A with "2" on the first administration and with "3" on the second (see Fig. 2). When distributions of the differences in response choices were considered, the results revealed a symmetric distribution across all survey

Examples of ***fact recall questions*** from the survey are shown below:

A) How many labs did you do each month?
None – 1 – 2 – 3 – 4 – 5 – More than 5

B) How many demonstrations did your teacher conduct each week?
None – 1 – 2 – 3 – 4 – 5 – More than 5

C) How much discussion about the lab did you have after it was over?
Not at all – 5 minutes – 10 minutes – Half of the class – A whole class or more

Examples of ***degree-of-experience rating questions*** are shown below:

C) Please rank the following for a particular lab:
i) Your understanding of concepts after lab
None 1 – 2 – 3 – 4 – 5 Complete
ii) Your understanding of lab procedure
None 1 – 2 – 3 – 4 – 5 Complete

D) In terms of learning the material your chemistry course required:
A lot of memorization 1 – 2 – 3 – 4 – 5 A Full Understanding of Topics

Figure 2. Selected example questions and corresponding response choices.

items. These results suggest that mismatches between first and second questionnaire responses were not biased and therefore the survey questions could be considered reliable.

Our approach will be to collectively consider the significant predictors, rather than draw conclusions from each predictor in isolation. It is our approach to view each significant predictor in relation to one another, constructing a model with the variables acting as multiple measures of the high school experiences of college students. We believe this approach will provide a more accurate rendering of the connection between high school chemistry learning experiences and college chemistry performance.

Statistical Approach

The researchers chose to use multiple linear regression controlling for college effects as the analytical method. We acknowledge that clustering of students within college classrooms would suggest hierarchical linear modeling as a more effective analytical method. However, an evaluation of the fully unconditional model obtained using ProcMixed (SAS, Inc.) revealed that the between-group differences accounted for only 6.7% of the variance in the outcome variable, course grade, whereas 93.3% of the variance was found to be within groups (Littell, Milliken, Stroup, & Wolfinger, 1996). In another study by Von Secker and Lissitz (1999), who did choose to apply HLM as an analytical method, the reported intraclass correlation was 0.36. This statistic indicates that differences between groups accounted for 36% of the overall variance in outcome. In two earlier studies by the authors, multiple regression was used in one study (Sadler & Tai, 2001), whereas HLM was used in the other (Tai & Sadler, 2001). The reasoning behind this difference in analytical approaches comes from the fact that the intraclass correlation of the sample was found to be 10%. The common perception among statisticians is that, in samples where greater 10% of the variance occurs between clusters, HLM would be a useful tool to account for additional variance in the fitted model. On the other hand, in samples where less than 10% of the variance occurs between clusters, multiple regression would provide ample rigor in forming a fitted model (X. Fan, September 4, 2003, personal communication). In our current analysis, an intraclass correlation of 0.067, which amounts to 6.7% of the variance being accounted for in between-group differences, suggests that employing HLM in this analysis would provide no clear advantage. In addition, the small number of different groups (i.e., chemistry courses) would also limit the statistical advantage of HLM. However, although the number of different schools does appear to be fairly small, the presence of wide-ranging differences among the institutions suggests that the results would have validity with respect to the generalizability of the findings. For this reason, we have included a variable for each college to account for systematic differences in stringency of grading and background of students. We choose to use multiple linear regression, because it provides the greatest analytical advantages with the least statistical complexity. Simply put, regression works best in this case.

In the Case of Missing Values

When analyzing survey data collected from many different sources, surveys with missing responses or multiple responses are common. For this analysis, missing and multiple responses are treated as missing. To deal with missing values, the first option is listwise deletion. In this instance, surveys with missing values for any of the variables being analyzed are not included in the analysis. For example, suppose a regression model being fitted included three different predictors, Predictors 1–3. Suppose ten surveys were missing values for Predictor 1. These ten surveys would be listwise deleted or not included in the regression analysis. Suppose, Predictor 1 was not found to be significant, and a new regression model that did not include Predictor 1 was being fitted, then these ten surveys would be included in the new analysis. The presumption when using listwise deletion is that the surveys with missing values are randomly distributed within the data set. Therefore non-inclusion of these surveys with missing values would maintain the randomness of the sample. Listwise deletion is the simplest and preferred option in the case of missing values.

However, there are instances when missing values may affect the randomness of the sample. A statistical test for this circumstance comes from an algorithm for analyzing missing values in a sample. This technique is called the expectation-maximization (EM) algorithm and is included in the SPSS statistical package. In addition to producing imputed values for missing data, the EM algorithm produces Little's MCAR test, a form of a χ^2 statistic (Little & Rubin, 2002; Scheffer, 2002). MCAR stands for "missing completely at random." In considering missing data, there are three possible instances that can occur: the data may be *missing completely at random* (MCAR); *missing at random* (MAR); and *not missing at random* (NMAR). In the case of MCAR data, the missing values do not appear to have a discernible pattern within the variables being used in an analysis and, therefore, the surveys with missing values may be excluded from the analysis (i.e., listwise deleted) without concern for any influence on the randomness of the sample. MCAR is considered a stringent standard for missing data (Allison, 2002; Little & Rubin, 2002; Scheffer, 2002). For MAR data, there appear to be patterns for missing values in the predictors, but these patterns do not appear to be related to the outcome. In this instance, the patterns among the predictors may be used to impute values for the missing data in cases where enough existing data can produce reliable and stable results. Therefore, because options exist in dealing with MCAR and MAR data, these cases are considered ignorable (Allison, 2002; Little & Rubin, 2002; Scheffer, 2002). For NMAR, the patterns of missing values are pervasive throughout the data including the outcome, and clear biasing would result from any analysis—this case is considered non-ignorable. No clear options exist for analysis of NMAR data.

Analytical Procedure

As stated previously, this study seeks to provide evidence for the connection between specific teaching and learning practices experienced by some students in high school chemistry that may give them advantage over their peers who share similar backgrounds. Therefore, the researchers began by developing a Demographic and General Educational Background Model. In discerning significant background characteristics, listwise deletion was applied in the analysis. Listwise deletion is a robust and widely used method for analyzing survey data that may contain a limited amount of missing values. Next, the researchers performed a missing value analysis using the SPSS statistical package with the EM algorithm option. This missing value analysis included the predictors in the Background Model and the outcome, ICCGRADE. No data were missing for the outcome. In the study sample, Little's MCAR test showed that the data were *not* missing completely at random (Little & Rubin, 2002). Thus, missing values were imputed and the imputed values included in subsequent analyses. The researchers decided to include the least amount of imputed data while meeting the standard for Little's MCAR test. This approach necessarily meant that, even with the inclusion of some imputed values, there would still be remaining variables with missing data. However, these missing data would be *missing completely at random* and allow for listwise deletion to be appropriate. In particular, the researchers found that including the imputed values for the variable SATVM in the data set resulted in Little's MCAR test meeting the standard that the remaining missing data were missing completely at random at the $\alpha = 0.05$ significance level; similar results were found when imputed values for SAT Mathematics scores and calculated SAT Verbal scores were analyzed. The next paragraph details the steps taken in this analytical step.

In addition to performing missing value analyses, the EM algorithm was used to calculate imputed values for the data missing from SATVM. The EM algorithm imputes the missing values in an iterative two-step procedure (an *expectation* step and a *maximization* step) using other closely associated variables such as academic achievement and gender. Prior to imputation, ACT scores were used to compute missing SAT values through an SAT–ACT concordance table (Dorans, Lyu, Pommerich, & Houston, 1997, p. 18; Dorans, 1999, p. 9). Published concordance tables allow for the conversion of ACT Composite scores to SAT Composite scores and ACT Mathematics scores to SAT Mathematics scores. In the analysis, the researchers replaced missing SAT Composite (SATVM) and SAT Mathematics (SAT-M) values with concorded ACT test scores, when reported. After this process, 7.3% or 112 missing cases were still missing for students who did not report either test score. Statistical simulation studies suggested that the EM algorithm would yield acceptable results for cases where up to 10% of the values for a variable were missing (Scheffer, 2002). The remaining missing values were imputed and retained in the data set under new variable names, SAT_IMP for the imputed SATVM scores and SATM_I for the imputed SAT-M scores. Missing SAT Verbal scores were then calculated by subtracting SATM_I from SAT_IMP. The resulting variables, SAT_IMP, SATM_I, and SATV_C, were reanalyzed using the SPSS missing value analysis procedure. The data set with imputed values for SATVM met the MCAR standard ($\chi^2 = 13.983$, $df = 21$, $p = 0.87$) at the $\alpha = 0.05$ level. For Little's MCAR test, the null hypothesis is that the data are missing completely at random. Thus, the value for p indicates the likelihood that the null hypothesis is true. In this particular instance, $p = 0.87$, suggesting that we accept that the null hypothesis is true and that the missing values are missing completely at random. The results for SATM_I and SATV_C were similar.

Once the Background Model was developed and the MCAR standard was met by the data set used in the analysis, the researchers moved on to consider the pedagogical predictors. This step of the analysis was done in two parts. First, a missing value analysis was performed with the pedagogical predictor, the other significant predictors, and the outcome. Second, pedagogical predictors meeting the MCAR standard were then included with the other significant predictors in a model fitted onto the outcome. Significant predictors were carried along in the analysis, whereas nonsignificant predictors were dropped from the fitted model. In instances where the MCAR standard was not met, the missing values for these predictors were imputed using the EM algorithm. This algorithm uses closely associated variables to produce imputed values for missing data using maximum likelihood estimation. The variable with imputed values was then reanalyzed to check if the MCAR standard was met. If not, other variables were imputed until the MCAR standard was met. The researchers did not proceed with fitting a regression model using listwise deletion until the MCAR standard was met. Throughout this process, the researchers made concerted efforts to include the least amount of imputed data while maintaining the MCAR standard.

Results and Discussion

This section includes descriptive statistics that aid in the characterization of the sample and a regression model of introductory college chemistry course grades fitted with predictors representing high school pedagogical experiences, demographic background factors, and general educational background factors.

Descriptive Statistics

The 1531 surveys in this data set came from 12 different introductory college chemistry courses for science and engineering majors at 12 different 4-year colleges and universities in ten

Table 1
Location and survey distribution among participating colleges and universities

School[a]	Location	Surveys	Percentage
School 1	ID	14	0.9
School 2	NY	15	1.0
School 3	GA	22	1.4
School 4	KY	29	1.9
School 5	SD	48	3.1
School 6	IN	83	5.4
School 7	CA	122	8.0
School 8	PA	123	8.0
School 9	AZ	160	10.5
School 10	MD	192	12.5
School 11	IN	210	13.7
School 12	KY	513	33.5

[a]Given the ongoing status of the study we have chosen to not reveal the participating institution names.

states. All of these courses followed a lecture–recitation–laboratory format. Schools included in this study were chosen because they followed this common course structure. Table 1 lists the locations and sample sizes from the 12 participating colleges and universities. The smallest sample of 14 was collected from School 1 in Idaho. The largest sample of 513 was collected from School 12 in Kentucky. The explanation for such large differences in sample sizes stems from the fact that some instructors surveyed their students during lecture sessions, whereas others surveyed one or more of their recitation sessions. Table 2 displays the sample distribution in

Table 2
Distribution of students surveyed with respect to institutional characteristics (N = 1531)

Admissions selectivity[a]	
High	911
Moderate	410
Nonselective	210
Affiliation	
Public	1325
Private	206
Size[b]	
Large	1019
Medium	243
Small	269
Minority enrollment[c]	
High	314
Moderate	407
Low	810

[a]Selectivity was based on average Composite SAT I Scores or Composite ACT Scores for Fall 2002. High = SAT I 1110+ or ACT 23+; medium = SAT I 1000–1100 or ACT 20–22; low = SAT I 990–, ACT 19–.
[b]Size indicates the number of full-time students enrolled. Large = 15,000+; medium = 5000–14,999; small = 4999–.
[c]Minority enrollment indicates the percentage of underrepresented minorities enrolled in the school. Underrepresented minorities include individuals who report themselves as being Native American, black, or Latino. High = 30%+; medium = 10–29%; low = 10%–.

terms of *selectivity in admissions*, *affiliation*, *size*, and *enrollment of minorities underrepresented in the sciences.* The greatest proportion of students appears to attend highly selective, large, public institutions with low minority enrollments. This description fits practically every state university. A review of federal education statistics shows that the largest group of the students attending 4-year colleges and universities enroll in state universities. These data suggest that our sample, although not representative of all undergraduate students in higher education, are representative of a significant portion of this population. Table 3 displays proportions of students in various demographic and general educational groupings. An interesting note is that the number of female students in this sample was 20% larger than the number of male students. Freshmen were by far the largest portion of students enrolled in introductory college chemistry for science, followed by sophomores, juniors, and seniors. Graduates and others (i.e., students from nontraditional backgrounds) made up 1.5% and 2.2% of the sample, respectively. Students who reported their background as black, Latino, or Native American made up 4.4%, 5.7%, and 0.8% of the sample, respectively. With regard to the general educational background of the students, half of the students reported taking calculus in high school. However, nearly all students (except for 28) reported taking high school chemistry. This very small number of students with no high school chemistry coursework suggests that the analytical approach, although powerful,

Table 3
Demographic and general educational background breakdown (N = 1531)

Gender	
Male	689
Female	816
Not reported	26
Year in college	
Freshman	968
Sophomore	302
Junior	133
Senior	50
Graduate	23
Other	33
Not reported	22
Racial background	
Asian	145
Black	68
Latino	87
Native American	13
White	1135
Multiracial	51
Not reported	32
HS chemistry	
None	28
Regular	832
Honors	408
Advanced placement	263
HS Calculus	
No Calculus	741
Calculus (Non-AP)	244
Advanced Placement AB	433
Advanced Placement BC	112
Not reported	1

would likely only have the ability to detect an effect if it were large. A comparative analysis of students with vs. without HS chemistry demonstrated no significant differences between students with no HS chemistry, students who took Honors Chemistry, and students who took regular Chemistry.

Regression Analysis

The analysis began with a consideration of the predictors from three distinct categories: demographic background, general educational background, and high school chemistry pedagogical experience. The outcome variable was introductory college chemistry grade, ICCGRADE. A detailed discussion of the analytical approach may be found in the Methods section. The researchers were careful in checking for threats to the robustness of the regression model, including interactions among predictors, considerations of possible sources of collinearity, influence of missing values on the statistical integrity of the sample, and the use of listwise deletion or missing value imputation. Steps were taken to carefully consider both the significance of each individual predictor included in a regression model as well as the change in explained variance as a result of the inclusion of each predictor. Predictors not meeting the level of significance ($\alpha = 0.05$) for the t-statistic or the ΔR^2 test F-statistic were excluded from the model. The final model is shown in Table 4 for a remaining sample of 1333 students. Of the original 1531 surveys included in the analysis, 198 were listwise deleted from the analysis for missing values in the variables included in the Final Model. The number of surveys listwise deleted from the analysis would have been substantially higher had imputed values not been used. This listwise deletion would have compromised the MCAR status reflected in the current sample of 1333 students in the regression analysis. The predictors in the Final Model that included imputed values are: *SAT Mathematics*, *Amount of Time on Stoichiometry*, and *Amount of Time on Nuclear Reactions*. Some of the details are included in Table 4. The remaining sample met Little's MCAR test standard ($\chi^2 = 501.880$, $df = 509$, $p = 0.58$), indicating that missing values were *missing completely at random*. This standard is discussed in greater detail in the Methods section. The sample used in the regression analysis constituted 87.0% of the original surveys. Interactions among the various predictors were analyzed, with none found to be significant.

The regression model accounted for 38.2% (36.3%, adjusted) of the variance in ICCGRADE. In the discussion that follows, we chose to use more descriptive variable names for the predictors to facilitate understanding. As such, readers should note that the names of the predictors appear with capitalization and in italics. The outcome variable remains as ICCGRADE.

Demographic and General Educational Background

Table 4 shows the Final Regression Model, which includes 11 significant demographic and general educational background predictors and predictor groups. Not surprisingly, students' academic achievement, as gauged through high school science and mathematics grades, course-taking patterns, and standardized tests, appear as significant predictors. Students reporting enrollment in Advanced Placement Chemistry and Calculus classes performed better than their peers who did not. Interestingly, the most influential predictor on ICCGRADE appears to be students' grade in their last mathematics course in high school. The difference in ICCGRADE between a student who earned an "A" in his or her last HS mathematics class and a student who earned a "B" is 2.79 points, or one-third of a letter grade.

In terms of the demographic predictors, gender did not appear as a significant predictor of performance. However, racial and ethnic groupings did prove to be weakly significant. Given that the largest proportion of students in the sample were white, the researchers chose this category to be the baseline. The results reveal some unexpected findings. *Asian* and *Black* students appeared to be at no significant disadvantage to white students, although *Latinos* were. Given the small number of *Native Americans* ($n = 13$) in the survey, the researchers refrain from drawing any conclusions regarding the regression results. This result suggests that a deeper examination into the mechanisms of sociocultural interaction is in order. Race and ethnicity have long been found to be predictive of achievement in many other studies (e.g., Hoffer, Moore, Quinn, & Suter, 1996; Hoffer, Racinski, & Moore, 1995; Oakes, 1990; Oakes, Ormseth, Bell, & Camp, 1990; Peng & Hill, 1995). In interpreting these results, small or insignificant β coefficients for race/ethnicity and gender were consistent with the view that SES and opportunity-to-learn underlie such differences between groups. The predictive power of race/ethnicity and gender shrink dramatically when other relevant variables are included in the model. Parental education, the availability of (and enrollment in) AP Chemistry and Calculus courses, and other variables account for much of the variance commonly attributed to race/ethnicity and gender. Controlling for backgrounds and opportunities, racial/ethnic and gender differences become nonsignificant or minor factors. Year in college revealed no significant differences among the levels of undergraduate students. The numbers of graduates and other (i.e., nontraditional) students were small, but large, significant differences were found. Graduates and nontraditional students earned higher grades than undergraduates.

The background analysis also revealed students who reported entering science as a means to a better career and who recalled having received no encouragement to take science classes both tended to have grades estimated to about one-tenth of a letter grade better than their peers. This result suggests that self-direction is associated with higher performance, although the advantage appears to be comparatively small.

The relative influence of the significant predictors may be gauged through the standardized β coefficients. The results show that *Last HS Math Grade* was the most influential predictor, with a coefficient of 0.20. In the same vein, SAT Mathematics is second, with a coefficient of 0.17. The next most influential predictors may be considered as a group with β coefficients ranging from 0.12 to 0.14 or having roughly 60–70% of the impact of *Last HS Math Grade*. The predictors include: *Last HS Science Grade*, *AP Calculus AB* enrollment, and *AP Calculus BC* enrollment. The third most influential group of background predictors had β coefficients ranging from 0.05 to 0.08, or 25–40% of *Last HS Math Grade's* impact. These predictors included *Highest Parental Educational Level*, *Race and Ethnicity*, *Year in College, Better Career*, *No Encouragement*, *Non-AP Calculus,* and *SAT Verbal*. The analysis included several preliminary models comparing the predictive effects of SAT Composite and separately entered SAT Mathematics and SAT Verbal. The researchers discovered that entering *SAT Mathematics* and *SAT Verbal* predictors produced a model accounting for 0.4% more variance than entering *SAT Composite*. Disentangling the effect of Mathematics and Verbal skills as measured through the SAT test, appears to improve the variance explained by the Final Model. A comparison of the β coefficients indicate that *SAT Mathematics* has three times the influence on the outcome than *SAT Verbal*, a result that has an intuitive appeal given the importance of quantitative skills in science coursework. The *SAT Mathematics* and *SAT Verbal* β coefficients along with last grades in science and mathematics may be thought of as measures for overall academic achievement. The large β coefficients for calculus while controlling for grades, *SAT Mathematics*, and *SAT Verbal* scores calls attention to the striking role of preparation in advanced mathematics on college chemistry success. AP Calculus has near double the inferred value of AP Chemistry on ICC. Why should mathematics, especially

Table 4
Final regression model of introductory college chemistry grade, ICCGRADE (N = 1333)

Predictor	B	SE B	β
College or university	Included		
(constant)	40.87***	2.64	
Demographic and general educational background			
Highest parental educational level	0.69**	0.23	0.07
College enrollment status			
Graduate student	8.26***	2.33	0.08
Nondegree student	4.77**	1.70	0.06
Race/Ethnicity			
Native American	−1.57	2.56	−0.01
Asian	−0.98	0.86	−0.03
Black	0.48	1.18	0.01
Latino	−3.04**	1.09	−0.07
Multiracial	−1.52	1.37	−0.03
Not reported	2.65	2.44	0.02
SAT Verbal[a]	0.006*	0.003	0.06
SAT Mathematics[b]	0.02***	0.003	0.17
Last HS mathematics grade	2.79***	0.35	0.20
Calculus (non–Advanced Placement)	1.58*	0.70	0.06
Advanced Placement Calculus AB	3.24***	0.61	0.14
Advanced Placement Calculus BC	5.18***	0.97	0.13
Last HS science grade	1.72***	0.37	0.12
Advanced Placement Chemistry	2.15***	0.63	0.08
Science as a means to a better career	1.48**	0.51	0.07
No encouragement to take science	0.99*	0.49	0.05
High school chemistry pedagogical experiences			
Learning requirements[c]	0.77***	0.23	0.08
Pedagogy frequency, individual work	−0.50**	0.19	−0.06
Lab preparation: read and discuss procedures day before	−1.03*	0.52	−0.05
Frequency of labs repeated for understanding	0.62*	0.27	0.05
Lab procedure understanding	−0.92***	0.25	−0.09
Number of own projects	−0.45*	0.20	−0.05
Number of assigned problems with calculations	0.45**	0.15	0.07
Test questions required memorization of terms/facts	1.75*	0.71	0.06
Amount of time on stoichiometry[d]	0.81***	0.20	0.10
Amount of time on nuclear reactions[e]	−0.72*	0.30	−0.06

$R^2 = 0.382$; adjusted $R^2 = 0.363$; 198 surveys listwise deleted for missing data from original sample.
[a]Calculated by subtracting SAT Mathematics scores from SAT Composite scores.
[b]Includes 112 imputed values, 7.3%.
[c]This particular variable used a Likert-type rating scale where Memorization = 1 to Full Understanding = 5.
[d]Includes 66 imputed values, 4.3%.
[e]Includes 71 imputed values, 4.6%.
$^*\le 0.05$.
$^{**}\le 0.01$.
$^{***}\le 0.001$.

calculus, which is not heavily drawn upon in ICC courses, be of such value? Proficiency in mathematics often lags well behind coursework. Students often must repeat their latest high school math course when entering college, and only 15% of students felt they were well prepared for college-level work (Mooney, 1994). Although the advanced topics of calculus are not directly utilized in ICC, calculus builds students' facility with algebraic functions, graph interpretation (including slope), mental computation, and calculation of rates of change. It appears that the virtually automatic facility with such mathematical skills that a successful calculus background bestows on students removes many impediments to understanding the quantitative aspects of chemistry that other students must endure.

High School Instruction and College Performance in Chemistry

Beyond the Background predictors, 11 high school chemistry pedagogical practices were found to be significant in predicting introductory college chemistry performance, as gauged by ICCGRADE. This result alone suggests that HS chemistry instruction does have an impact on college chemistry performance.

The predictor *Learning Requirements* asked the respondents to provide a general impression of their high school chemistry course, ranging from ‘‘A lot of memorization of facts’’ to ‘‘A full understanding of topics.’’ The regression model shows that students reporting greater emphases on understanding performed better than their peers reporting greater emphases on memorization. This result suggests that high school chemistry courses with an emphasis on greater depth of learning may be more helpful than superficial recall of facts and concepts. This is not a surprising result, but one that sets the tone for the results and discussion to follow.

Three predictors describing laboratory experiences were found to be significant: *Labs Repeated for Understanding*, *Lab Preparation* (*Read and Discuss Day Before*), and *Understanding Lab Procedure*. The parameter estimate for *Repeated for Understanding* shows that students reporting more instances of repeating labs to enhance their understanding earned higher college chemistry grades than their peers who reported few or no instances of repeating labs for understanding. We note that the number of high school laboratory sessions is not a significant factor. This finding supports repeating laboratory experiences, but not increasing the number of different labs. This result is in contrast to *Lab Preparation* and *Lab Procedure*. For both of these predictors, the parameter estimates were negative, suggesting that students reporting more instances of reading and discussing their labs the day before doing them and/or reporting that they were expected to have a complete understanding of their lab procedures had lower college chemistry grades than their peers. These three results considered individually may appear perplexing, but taken as a group and considered carefully, a picture emerges regarding the importance of understanding the concepts of chemistry rather than the mechanics of completing assigned tasks. These results appear to suggest that students reporting greater emphasis on learning for understanding rather than merely learning to successfully complete assignments have greater success in future study. Although it is true that students learn from doing labs properly, it may also be said that a great deal of learning occurs when students are allowed to return to the bench and redo the lab activity for purposes of enhancing their understanding (and not as punishment). Redoing labs can provide second chances when more and better observations may be made, mistakes may be corrected, and further explorations may be carried out. It would seem that students engaging in lab activities that resemble assembly lines where the emphasis lies in proper procedure may be disadvantaged in future coursework.

These findings should not be a surprise to science educators given the work of Woolnough (1998) and Hodson (1996), both of whom discussed the characteristics of effective and ineffective ‘‘practical work’’ (i.e., high school laboratory assignments). Woolnough distinguished between two primary aims in teaching science: ‘‘to know *what* and to know *how* (author’s emphasis)’’ (p. 113). He explained that students should come to understand the principles and theories of science and also understand ‘‘the way scientists work’’ (p. 113). Hodson identified three kinds of learning through scientific inquiry: conceptual understanding, procedural knowledge, and investigative expertise (pp. 131–132). Woolnough’s and Hodson’s ideas emphasize that the focus of laboratory experiences should be on the development of students' ideas and not on the procedure aspects of carrying out experiments.

Our results and the conclusions of Woolnough and Hodson call into question the value of extensive preparation for labs especially focused on procedure, which includes pondering the "proper" outcomes, common practices favored by many teachers. Such preparation may simply be at the expense of sufficient time to carry out the experiment and to acquire the tacit knowledge necessary to coax useful results from a laboratory set-up. As physicist Samuel Devons of Barnard College reflected:

> . . .experiment is a craft. . .craft is knowledge you have in your fingertips, little tricks you learn from doing things, and when they don't work and you do them again. You have little setbacks and you think, how can I overcome them? And then you find a way. (Crease, 2003, pp. 184–185)

Four predictors associated with tests and assignments appeared as significant in the regression analysis: *Test Questions Requiring Memorization*, *Number of Problems with Calculations*, *Number of Own Projects*, and *Frequency of Individual Work*. *Test Questions Requiring Memorization* and *Number of Problems with Calculations* both had positive parameter estimates, suggesting that students reporting greater frequency of test questions requiring memorization and class assignments with problems requiring calculations earned higher college chemistry grades. Again, these results seem to initially run counter to the findings discussed earlier, especially the first result suggesting that understanding is more beneficial than memorization. However, a more careful consideration of the learning process might lead one to conclude that memorization does have its place in developing a greater degree of understanding as students with facts easily recalled can move more quickly than students needing to search for these essential facts. Students engaging in solving stoichiometric problems might do well to memorize Avogadro's number (6.022×10^{23} atoms/mole—just in case you were wondering) as well as the atomic weights of particular elements such as carbon, oxygen, nitrogen, and hydrogen. Note that analysis suggests that factual information should be tested, but there is no support that it should be explicitly taught. Class time appears to be better spent on concept development rather than fact memorization. This relationship between memorization and deeper learning was discussed by Marton and Booth (1997). They talked about memorization playing a role in deepening understanding in that the act of memorization can accentuate different aspects or segments of the information being memorized. Marton and Booth explained that memorization need not be limited to mere "mnemonic tricks," but that "successive repetition and review" may help learners develop a deeper sense of meaning and make the information their own (p. 44). With this in mind, the positive relationship between *Test Questions Requiring Memorization* and ICCGRADE does not contradict the first finding. The positive relationship between *Problems with Calculations* and ICCGRADE appears to speak to academic rigor, because correct calculation often relies on a combination of clear conceptual understanding and accurate recall of necessary information. In fact, in a review of learning research and theory, Bransford, Brown, and Cocking (2000) noted:

> To develop competence in an area of inquiry, students must (a) have a deep foundation of factual knowledge, (b) understand facts and ideas in the context of a conceptual framework, and (c) organize knowledge in a way that facilitates retrieval and application. (p. 16)

The findings suggest a connection exists linking learning in high school chemistry classes and performance in college chemistry, providing an example of the long-term impact of formal learning on performance (Conway, Cohen, & Stanhope, 1991).

The negative relationships of *Own Projects* and *Individual Work* with ICCGRADE have a fairly clear common thread that suggests that individualized student learning in high school may be less beneficial to students in college chemistry. This finding is consistent with research discussed by Bransford et al. (2000) regarding young children. In their review of research concerning learning in children, they discussed a wide range of research studies highlighting the importance of interaction and guidance in learning.

These results may initially appear to some to contradict the earlier finding suggesting student self-direction as gauged by the *No Encouragement* predictor is a positive influence on ICCGRADE. Although self-direction appears to be an important positive influence, this characteristic must be distinguished from self-directed learning. For instance, some students may possess the desire to learn chemistry, but without the proper guidance they would be at a distinct disadvantage to their peers who have access to knowledgeable instructors and helpful peers. The distinction may be summarized in the difference between the desire to learn and the actual practice of learning. Thus, self-directed learning appears to be a negative influence on ICCGRADE, suggesting that student–teacher and student–student interactions are likely an important part of learning. This conclusion finds support in the work of Brown and Campione (1994), regarding the role of teachers in student learning, and Halloun and Hestenes (1985), regarding the management of discourse in a classroom. Brown and Campione indicated that responsibility for learning in their ‘‘ideal classroom’’ is shared between students and teacher and emphasized the importance of the teacher in guiding student inquiry. Halloun and Hestenes reported that managing classroom discourse may be the single most important pedagogical feature for enhancing student learning. These conclusions support our findings and suggest that pedagogy in high school chemistry courses should be guided by an attentive, but not heavy hand.

Two content-related predictors were found to be significant: *Amount of Time on Stoichiometry* and *Amount of Time on Nuclear Reactions*. However, the relationships point in opposite directions, with more *Time on Stoichiometry* predicting higher ICCGRADE and more *Time on Nuclear Reactions* predicting lower ICCGRADE. Considering stoichiometry and nuclear reactions as topics of study, clearly, stoichiometry plays a central and recurring role in introductory chemistry courses, whereas nuclear reactions are commonly relegated to the end of most textbooks. The concepts and calculations related to stoichiometry involve ‘‘the quantitative relations between elements and compounds in chemical reactions’’ (Mortimer, 1979, p. 149)—clearly a fundamental topic in introductory college chemistry. Thus, students reporting greater amounts of time spent on nuclear reactions will most likely have encountered topics in the intervening chapters as well, suggesting that their high school chemistry class focused on breadth rather than depth. The concentration on stoichiometry may reflect the same approach to the development of sound fundamental understandings of chemistry concepts as suggested by the earlier result showing *Labs Repeated for Understanding* as a positive predictor of ICCGRADE. A strong foundation in stoichiometry appears to be the cornerstone to introductory college chemistry success. The positive connection between a fundamental topic and college performance was also discovered in a previous study connecting high school physics experiences and college physics success (Sadler & Tai, 2001).

Finally, this study considered the relationship among the various characteristics of HS chemistry teachers and ICCGRADE. The list of teacher characteristics included: organizational ability, ability to explain problems in several different ways, ability to handle discipline, pleasantness, enthusiasm for chemistry, and fairness. An analysis revealed very high correlations among the predictors associated with these characteristics (ρ ranged from 0.7 to nearly 0.9); it appeared that students did not differentiate among the characteristics represented by these variables. The variables were combined into composite predictors, but were not found to be significant. One can interpret this result in two different ways, either these teacher attributes have little bearing on the quality of teaching or students are unable to assign reliable values that represent such teacher qualities.

The significant predictors discussed previously represent only a fraction of the high school physics experience predictors included in this analysis. Many other predictors included in the analysis did not appear to be significant. These include: (a) block scheduling (including extended lab periods); (b) number of labs per month; (c) high school physics class size; (d) frequency of demonstrations; (e) studying qualitative problems; and (f) amount of homework assigned. Although the lack of significance in a single study does not preclude the existence of a correlation between these predictors and college chemistry performance, it does suggest that the effect size of these predictors likely does not rise above the effect size of the predictors found to be significant.

In summary, the results from the regression analysis appear to paint a complex, but intuitively appealing picture of students' HS chemistry experiences and ICCGRADE. Careful consideration of the predictors led to some interesting findings suggesting that an emphasis on understanding in HS chemistry courses may contribute to the future success of students in introductory college chemistry.

Conclusions

At this time, we revisit the theory behind our approach in this research study. The use of epidemiological methods to inform our understanding of the connection between high school chemistry instruction and college chemistry performance necessitates the collective analysis of several measures of the high school chemistry experience. The purpose is not to identify and interpret these measures singly, in fact this approach would likely lead to a misinterpretation. Rather, the purpose is to identify measures of an experience that show significant associations with an outcome and then to collectively consider these measures in a more holistic interpretation. Critics of this approach may rightly point to the instability of the significance of any given single predictor in our models and indeed this may prove to be correct in future analyses. As such, we do not profess that every single predictor found to be significant in our models would be found significant in all future studies of this type. We caution the reader that the interpretation of the predictors included in our models in isolation is unsound. For example, the parameter estimate for *Number of Own Projects* was found to be negative. This result should *not* be interpreted as an indication that having high school chemistry students do their own projects lowers their college chemistry grades. This explanation would be ill-founded. Instead, we urge the reader to consider the 11 high school chemistry pedagogical predictors as measures of parts of a whole. Continuing with this example of *Number of Own Projects*, its negative parameter estimate interpreted in accordance with the negative parameter estimate of *Frequency of Individual Work* suggests that working in isolation in high school chemistry appears to be associated with lower college performance. (This point was discussed in greater detail in the Results section.) The trends uncovered in this research were found through a collective interpretation of the results and, based on this methodological approach, are likely robust.

This study has sought to find evidence for or against the beliefs of teachers, professors, and researchers concerning the factors contributing to success in introductory college chemistry courses. The picture that emerges reveals complex, yet intuitively reasonable factors at play. We acknowledge that using regression models to relate variables with specific outcomes is problematic; first, correlation does not imply causation and, second, significance of predictors in one study does not insure significance in all subsequent studies. Yet, we feel our regression model is helpful and will serve to:

- Separate out and account for factors over which interventions by teachers or institutions will have little impact from those that provide opportunities for considerable leverage.
- Refute, or at least, equivocate, some strongly held beliefs by identifying patterns among variables that are not statistically significant or that have coefficients of the opposite sign of what was expected (lack of correlation is evidence for lack of causation).
- Identify variables that account for substantial amounts of variance that will benefit from additional analysis or research.
- Assist in supporting the formulation of explanatory frameworks that tie together diverse factors including the roles of decisions made by students and by their teachers.

Demographic predictors do account for substantial variance in how well students perform in introductory college chemistry courses. Prime among these is the level of parent's education, which is characteristically used as a proxy for family income and the affluence of the community. Gender of students does not appear to make a difference between students if their course-taking and testing history is similar. Race does not appear to be significant, with the exception of Latino students. The researchers acknowledge the relatively small proportion of Latino students as well as all other groups of students of color in this sample. Care must be taken in such studies to account for factors that mask the impact of different educational approaches, such as: graduate and special students performing exceedingly well in introductory courses; overall high school performance and standardized test scores; and motivation to prepare for careers requiring science.

One characteristic that distinguishes the sciences in high schools from all other content areas is the role of the laboratory. It is clear that laboratory work holds great promise in helping to prepare students for college-level studies (e.g., Hegarty-Hazel, 1990; Wellington, 1998). For many, discovering how nature behaves is both fascinating and instructive. Yet, when discovery is drained from the experiment, when there is insufficient time to truly explore and ponder physical phenomena, and when the ‘‘correct result’’ is what matters most, laboratory experiments appear to be counterproductive. A few select labs that may be repeated to enhance deeper understanding should be considered over a broad-stroke approach that would necessarily entail constantly changing equipment and procedures. This caveat appears true in the coverage of topics as well; students are better prepared if they study fewer topics in greater depth, while being required to recall relevant facts. Also, experience with quantitative problems is important, because chemistry becomes more quantitative in college. Toward that end, it appears that the more mathematics taken in high school, the better.

Overall, we find support for high school chemistry courses that: (1) value understanding over coverage; (2) involve students in collaborative activities; (3) emphasize the quantitative over the qualitative; and (4) allow room for personal discovery and wonder without teachers abdicating the responsibility for direction and guidance as well as academic rigor in their courses. Clearly, these choices come with a cost to maintaining a faster-paced content-exposure approach to teaching. Certainly, high-stakes testing and comprehensive lists of content ‘‘standards’’ further exacerbate the tension between insuring a deep understanding of topics versus simply covering the material. It would seem from the results of the regression analysis that students engaging in assembly-line lab activities focusing on procedure and manufacturing required results, fast-paced content coverage, and unguided learning activities are at a disadvantage to their peers who engage in the learning of chemistry for understanding, doing calculations, memorizing when necessary, and revisiting topics and laboratory experiments. In the end, the results of this study suggest that HS chemistry pedagogical experiences do appear to play significant roles in the future success of students in introductory college chemistry courses. Although conventional wisdom has focused primarily on achievement tests and content standards, it appears that students' success in the ‘‘real-world’’ circumstance of college coursework may be influenced by their high school chemistry teachers and the decisions these teachers make in designing the segments of their chemistry courses that are not mandated through high-stakes testing.

Notes

The authors would like to acknowledge the people who helped make this research project possible: Janice M. Earle, Finbarr C. Sloane, and Larry E. Suter of the National Science Foundation for their insight and vision of the ‘‘big picture’’; James H. Wandersee, Joel J. Mintzes, Lillian C. McDermott, Eric Mazur, Dudley R. Herschbach, and Brian Alters of the FICSS Advisory Panel for their invaluable guidance; Nancy Cianchetta, Susan Matthews, Dan Record, and Tim Reed of our FICSS High School Science Teachers Advisory Board for their time and wisdom, Xitao Fan of the Curry School of Education at the University of Virginia for his guidance on statistical analysis; Matthew H. Schnepps, Nancy Finkelstein, Alex Griswold, Tobias McElheny, Yael Bowman, and Alexia Prichard of the Science Media Group for their work to get the meassage out to school teachers; and Michael Filisky, Hal Coyle, Cynthia Crockett, Bruce Ward, Judith Peritz, Annette Tranga, Freeman Deutsch, Zahra Hazari, and Marc Schwartz of the Project FICSS Team for their work to take ideas and make them happen. (Special thanks go to Zahra H. for making the innumerable telephone calls to potential participants.) Most of all, we wish to thank the college chemistry professors and their students for their willingness to help in this investigation. It was only through their participation that this study was even a possibility.

References

Aikenhead, G. (1984). Teacher decision making: The case of Prairie High. Journal of Research in Science Teaching, 21, 167–186.

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

Alters, B.J. (1995). Counseling physics students: A research basis. The Physics Teacher, 33, 413–415.

Ausubel, D.P., Novak, J.D., & Hanesian, H. (1978). Educational psychology: A cognitive view (2nd ed.). New York: Holt, Rinehart and Winston.

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.

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.

Bransford, J.D., Brown, A.L., & Cocking, R.R. (Eds.). (2000). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.

Brasted, R.C. (1957). Achievement in first year college chemistry related to high school preparation. Journal of Chemical Education, 34, 562–565.

Brickhouse, N.W., Lowery, P., & Schultz, K. (2000). What kinds of a girl does science: The construction of school science identities. Journal of Research in Science Teaching, 37, 441–458.

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

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, 297–331.

Cannell, C.F., Fowler, F.J., & Marquis, K.H. (1965). A report on respondents’ reading of the brochure and letter and an analysis of respondents’ level of information. Ann Arbor, MI: Institute for Social Research, University of Michigan.

Catsambis, S. (1995). Gender, race, ethnicity, and science education in the middle grades. Journal of Research in Science Teaching, 32, 243–258.

Conway, M.A., Cohen, G., & Stanhope, N. (1991). On the very long-term retention of knowledge through formal education: Twelve years of cognitive psychology. Journal of Experimental Psychology, 120, 395–409.

Crease, R.P. (2003). The prism and the pendulum. New York: Random House.

Debaz, T.P. (1994). A meta-analysis of the relationship between students' characteristics and achievement and attitudes towards science. Unpublished doctoral dissertation, Ohio State University.

Deboer, G.E. (1987). Predicting continued participation in college chemistry for men and women. Journal of Research in Science Teaching, 24, 527–538.

Dorans, N.J. (1999). Correspondence between ACT and SAT I scores (College Board Research Report 99-1). New York: The College Board.

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, 24–35.

Duschl, R. & Wright, E. (1989). A case study of high school teacher's decision making models for planning and teaching science. Journal of Research in Science Teaching, 26, 467–501.

Everhart, W.A. & Ebaugh, W.C. (1925). A comparison of grades in general chemistry earned by students who (a) have had and (b) have not had high-school chemistry. Journal of Chemical Education, 2, 770–775.

Gainen, J. (1995). Barriers to success in quantitative gatekeeper courses. In J. Gainen & E.W. Williamson (Eds.), Fostering student success in quantitative gateway courses, No. 61 (pp. 5–14). San Francisco, CA: Jossey-Bass.

Gibson, D.J. & Gibson, L.S. (1993). College students' perceptions on adequacy of high school science curriculum as preparation for college level biology. The American Biology Teacher, 55, 8–12.

Gifford, V.D. & Harpole, S. (1986). Factors contributing to freshmen physics achievement. Memphis, TN: Mid-South Educational Research Association. (ERIC Document Reproduction Service No. ED 278 554.)

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

Halloun, I.A. & Hestenes, D. (1985). The initial knowledge state of college physics students. American Journal of Physics, 53, 1043–1048.

Hart, G.E. & Cottle, P.D. (1993). Academic backgrounds and achievement in college physics. The Physics Teacher, 31, 470–475.

Hegarty-Hazel, E. (Ed.) (1990). The student laboratory and the science curriculum. London: Routledge.

Hendricks, B.O., Koelsche, C.L., & Bledsoe, J.C. (1963). Selected high school courses as related to first quarter chemistry marks. Journal of Research in Science Teaching, 1, 81–84.

Hermann, G.A. (1931). An analysis of freshman college chemistry grades with reference to previous study in chemistry. Journal of Chemical Education, 8, 1376–1385.

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., Moore, W., Quinn, P., & Suter, L. (1996). High school seniors' instructional experiences in science and mathematics (NCES 95-278). Washington, DC: National Center for Educational Statistics.

Hoffer, T.B., Racinski, K.A., & Moore, W. (1995). Social background differences in high school mathematics and science coursetaking and achievement (NCES 95–206). Washington, DC: National Center for Educational Statistics.

House, J.D. (1993, May). Noncognitive predictors of achievement in introductory college chemistry. Paper presented at the Association for Institutional Research Annual Forum, Chicago, IL (ERIC Document Reproduction Service No. ED360944).

Lamb, D.P., Waggoner, W.H., & Findley, W.G. (1967). Student achievement in high school chemistry. School Science and Mathematics, 47, 221–227.

Lawrenz, F. (1976). The prediction of student attitude toward science from student perception of the classroom learning environment. Journal of Research in Science Teaching, 13, 509–515.

Linton, M. (1982). Transformations of memory in everyday life. In U. Neisser (Ed.), Memory observed (pp. 77–91). New York: W.H. Freeman & Co.

Littell, R.C., Milliken, G.A., Stroup, W.W., & Wolfinger, R.D. (1996). SAS systems for mixed models. Cary, NC: SAS Institute.

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

Marton, F. & Booth, S. (1997). Learning and awareness. Mahwah, NJ: Lawrence Erlbaum.

McQuary, J.P., Williams, H.V., & Willard, J.E. (1952). What factors determine student achievement in first-year college chemistry? Journal of Chemical Education, 29, 460–464.

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: Lawrence Erlbaum.

Mooney, C.J. (1994, June 22). The shared concerns of scholars. The Chronicle of Higher Education, A34–A37.

Mortimer, C.E. (1979). Chemistry: A conceptual approach (4th ed.). New York: Van Nostrand.

Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., & Smith, T.A. (1998). Mathematics and science achievement in the final year of secondary school: IEA's Third International Mathematics and Science Study (TIMSS). Boston: Center for the Study of Testing, Evaluation, and Educational Policy.

Nordstrom, B.H. (1990). Predicting performance in freshman chemistry (ERIC Document Reproduction Service No. ED347065).

Oakes, J. (1990). Lost talent: The underparticipation of women, minorities, and disabled persons in science. Santa Monica, CA: RAND.

Oakes, J., Ormseth, T., Bell, R., & Camp, P. (1990). Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science. Santa Monica, CA: RAND.

Oetting, E.R. & Beauvais, F. (1990). Adolescent drug use: Findings of national and local surveys. Journal of Consulting and Clinical Psychology, 58, 385–394.

Ogden, W.R. (1976). The effect of high school chemistry on college chemistry achievement: A summary. School Science and Mathematics, 76, 122–126.

Peng, S.S. & Hill, S.T. (1995). Understanding racial-ethnic differences in secondary school science and mathematics achievement (NCES 95-710). Washington, DC: National Center for Educational Statistics.

Powers, S.R. (1921). The achievement of high school and freshman college students in chemistry. School Science and Mathematics, 21, 366–377.

Ramsden, P. (2003). Learning to teach in higher education (2nd ed.). London: Routledge Falmer.

Reynolds, A.J. & Walberg, H.J. (1992). At structural model of science achievement and attitude: An extension to high school. Journal of Educational Psychology, 84, 371–382.

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

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

Shumba, O. & Glass, L.W. (1994). Perceptions of coordinators of college freshman chemistry regarding selected goals and outcomes of high school chemistry. Journal of Research in Science Teaching, 31, 381–392.

Stengel, B., Tarver-Car, M.E., Powe, N.R., Eberhardt, M.S., & Brancati, F.L. (2003). Lifestyle factors, obesity, and the risk of chronic kidney disease. Epidemiology, 14, 479–487.

Sudman, S. & Bradburn, N. (1973). Effects of time and memory factors on response in surveys. Journal of the American Statistical Association, 68, 805–815.

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

Von Secker, C.E. & Lissitz, R.W. (1999). Estimating the impact of instructional practices on student achievement in science. Journal of Research in Science Teaching, 36, 1110–1126.

Wellington, J. (Ed.) (1998). Practical work in school science: Which way now? London: Routledge.

Weiss, I.R., Pasley, J.D., Smith, P.S., Banilower, E.R., & Heck, D.J. (2003). Looking inside the classroom: A study of K-12 mathematics and science education in the United States—highlights report. Chapel Hill, NC: Horizon Research.

West, G.A. (1932). Influence of high school science on grades in college chemistry. School Science and Mathematics, 32, 911–913.

Woolnough, B.E. (Ed.) (1991). Practical science: The role and reality of practical work in school science. Philadelphia: Open University Press.

Woolnough, B.E. (1998). Authentic science in schools, to develop personal knowledge. In J. Wellington (Ed.), Practical work in school science: Which way now? (pp. 109–125). London: Routledge.

Yager, R.E., Snider, B., & Krajcik, J. (1988). Relative success in college chemistry for students who experienced a high school course in chemistry and those who had not. Journal of Research in Science Teaching, 25, 387–396.