Evaluation of an Authentic Research-based Curriculum's Effects on Undergraduate Biology Student Achievement

Brandon S. Diamond
David Posnack Jewish Day School

Abstract

An authentic research-based science curriculum for undergraduate science majors was implemented at a major private university in order to improve student learning and interest in science. Because most involved students majored in life sciences, cumulative GPA, biology GPA, and upper-level biology GPA were used as outcome measures of student achievement. Participation in the program, college admissions test scores, high school GPA, gender, cohort, and major were used as predictors in the multiple linear regression models. The overall model accounted for 44% of the between-student variance. Authentic research was found to improve biology student achievement compared to a control group when introductory coursework was included in the GPA calculation, but not when upper-level biology courses were examined without introductory courses. College admission test scores and high school GPA were found to be significant predictors of student achievement. The advantages of accounting for prior student achievement when evaluating program success are discussed.

Keywords: authentic research, SAT, ACT, Grade Point Average, undergraduate biology education.

College student success is a national concern in the United States, with a university dropout rate around 25% (Tuckman & Kennedy, 2011). In science majors, this rate is even worse for women and minorities, although the gender gap is beginning to show signs of closing (Griffith, 2010; Whalen & Shelley, 2010). Science programs in universities around the world are understandably working to find more effective methods to educate students in an attempt to both retain them until graduation and to matriculate higher quality graduates.

This study examines the effectiveness of an authentic research-based science curriculum housed in a biology department towards increasing student achievement. Additionally, the study examines the effectiveness of commonly used metrics in predicting student achievement to determine whether differences in student achievement were caused by the curriculum or by preexisting student ability. Using three measures of student achievement, (a) cumulative college grade point average (GPA), (b) biology GPA, and (c) upper-level biology GPA, the following two research questions (R.Q.) were examined:

1. What effect does an authentic research-based undergraduate science curriculum have on student achievement?
2. What student variables were successful predictors of student achievement?

This study evaluates the effectiveness of an authentic research-based science curriculum using a large data set with a comparison group. Additionally, it examines the effectiveness of student variables such as college admissions test scores and high school GPA in predicting student achievement in biology.

Literature Review

Inquiry in undergraduate laboratories

Both the National Research Council and the American Association for the Advancement of Science recommend that undergraduate students participate in student-centered authentic (or inquiry-based) research, in both early and advanced courses (American Association for the Advancement of Science, 2011; National Research Council (U.S.), 2003; National Research Council (U.S.), 2009; National Research Council (U.S.), 2012). This recommendation is in response to the long history of literature demonstrating that inquiry-based pedagogies promote deeper understanding of content and/or greater scientific reasoning skills (Timmerman, Strickland, & Carstensen, 2008). A meta-analysis of 12 studies examining inquiry demonstrated that inquiry had a large effect on student achievement (Schroeder, Scott, Tolson, Huang, & Lee, 2007).

Defining inquiry

The development of quality inquiry-based curricula has been complicated by the fact that use of the word "inquiry" has become ambiguous in regard to classroom labs (Brownell, Kloser, Fukami, & Shavelson, 2012; Buck, Bretz, & Towns, 2008). It has also been noted that there is even less agreement about the definition of inquiry between the research community and the teaching community (Buck et al., 2008). This ambiguity has led researchers (including those involved in the AAAS report (2011)) to begin using the term "authentic research" rather than "inquiry-based research" (Brownell et al., 2012). Authentic research consists of students creating their own research questions to which the answers are not yet known and designing and performing experiments to answer those original questions. However, the modifiers used when describing types of inquiry (e.g., "authentic") often also have disparate definitions in the literature.

Buck, Bretz, and Towns (2008) proposed a quantitative rubric to characterize the level of inquiry occurring in laboratory activities based on six characteristics: problem/question, theory/background, procedures/design, results analysis, results communication, and conclusion. A higher level of inquiry indicated that more of these characteristics were provided by the student rather than by the instructor. Level 0, or confirmation inquiry, consisted of all levels being provided by the instructor. Level ½, or structured inquiry, requires students to communicate their results and provide their own conclusions. Level 1, or guided inquiry, requires students to also analyze results. Level 2, or open inquiry, requires students to also develop their own procedures. Level 3, or authentic research, requires students to develop all of the characteristics themselves. While it might be expected (but not preferred) that Level 0 and Level ½ are the most common types of inquiry found in elementary schools, Buck et al.'s (2008) analysis of 386 undergraduate laboratory manuals found that the majority of activities were at Levels 0 and ½, while none were at Level 3. In fact, the two biology manuals they examined did not include any activities at Level 2 or 3.

Buck et al. were not the only researchers to suggest standardized levels of inquiry. Unfortunately, other rating schemes use similar terminology to describe different levels of inquiry. Bell, Smentana, and Binns (2005) describe four levels of inquiry: confirmation (Level 1), structured (Level 2), guided (Level 3), and open (Level 4). Open inquiry in this scheme is the equivalent of authentic research in Buck et al. Xu and Talanquer (2013) used a similar rating scheme to that of Bell et al., claiming they based it on Inquiry and the National Science Education Standards (National Research Council (U.S.), 2000), which describes inquiry as ranging from guided to open. All of these rating schemes agree that the highest level of inquiry requires students to develop the entire experiment from the beginning, and the instructor's role is more of support than of leadership.

Effectiveness of inquiry

There are many studies that have demonstrated the effectiveness of inquiry when it is carried out consistently with researcher recommendations. For example, Brownell et al. (2012) used surveys to compare students participating in a traditional “cookbook” lab (in which students are given step-by-step instructions) to students participating in authentic research. The results showed that students participating in authentic research increased their self-confidence and self-efficacy more than students participating in cookbook labs. Students in the experimental group also increased their interest in conducting future research more than those in the comparison group. This study had some flaws. Students in the experimental group were volunteers, so while the researchers planned for a randomized experiment, there were likely preexisting differences between the experimental group students and the comparison group students. The issues caused by this nonrandom assignment are clear when examining the pre-course results of the survey questions asking about student preferences in regards to the format of lab experiments. Students in the experimental group were significantly more likely to prefer to design their own experiments before the class began. Additionally, the experimental group’s classes were facilitated by PhD biology graduate students, while the comparison classes were mostly facilitated by undergraduate students. These issues were mentioned in the article, but were not controlled for in the analysis.

Other researchers have explored the effects of inquiry on student outcomes without including comparison groups. One study examined a freshman biology laboratory course, included as part of an accelerated introductory program, which engaged students in authentic or open inquiry (both terms were used) (Kazempour, Amirshokoohi, & Harwood, 2012). Participants in this study self-reported a better understanding of scientific processes (e.g., collaboration and failure are both common in scientific experiments) and stronger self-perception as a scientist. Another study examined an introductory biology lab newly designed to include inquiry-based sections by examining pretest and posttest scores from a content knowledge test (Timmerman et al., 2008). This study found that many of the didactic sections of the course showed greater gains than the inquiry-based sections of the course, but the researchers attributed this result to differences in time spent on and difficulty of the topics covered using the different teaching methods. The study also found that students were better able to generalize inquiry-based topics, reflecting a better level of understanding. There clearly seem to be advantages to inquiry-based curricula, at least in the short term. However, none of these studies accounted for differences in baseline ability of the students being taught.

Predicting student success

For many years, the literature has examined predictive models for collegiate outcomes such as freshman grade point average (GPA) (Rothstein, 2004). However, fewer studies have examined longer-term outcomes such as cumulative college GPA (Atkinson & Geiser, 2009). While college admissions test scores, primarily in the form of SAT and ACT scores, are widely used as predictors of college success, much of the recent literature seems to consist of a debate between the College Board (which administers the SAT) and the faculty at the University of California, which are justifying their decision to make standardized testing optional for college admission (Atkinson, 2002; Atkinson & Geiser, 2009; Geiser, 2009; Mattern, Shaw, & Kobrin, 2010).

The SAT and ACT have repeatedly been shown to predict freshman GPA (Bridgeman, McCamley-Jenkins, & Ervin, 2000; Coyle & Pillow, 2008; Patterson & Mattern, 2013; Robinson & Monks, 2005; Rothstein, 2004; Sawyer, 2010). However, the SAT (Bridgeman et al., 2000) and ACT (Noble & Sawyer, 2002) have both been found to be more predictive of freshman GPA in high-ability students than in low-ability students. Some argue that high school grades are a better indicator of college readiness than college admissions test scores, so test scores should be reduced or eliminated in the college admissions process (Atkinson, 2002; Atkinson & Geiser, 2009;

Geiser, 2009). In higher achieving students, composite ACT scores have been found to be more predictive of college GPA than high school GPA (Noble & Sawyer, 2002). It has also been noted that low socioeconomic status is a major contributor to low SAT scores, but that is because students from a lower socioeconomic status are actually less likely to be well prepared for college success (Rothstein, 2004). A more recent report released by The College Board suggests that this situation has been corrected, with low socioeconomic status and minority students' SAT scores overpredicting freshman GPA (Patterson & Mattern, 2013). Similarly, the ACT has been found to overpredict the achievement of minority students (Radunzel & Noble, 2013), a group statistically more likely to be from a lower socioeconomic status.

Combining the results of several older studies suggests that the best predictive model of student achievement would include SAT or ACT scores, high-school GPA, and class rank (Cohn, Cohn, Balch, & Bradley Jr., 2004). Cohn et al. confirmed this suggestion in their study of 521 economics students.

The College Board has examined the ability of the SAT to predict first year, second year, and third year cumulative GPA (Mattern & Patterson, 2011), and the studies found that the SAT is strongly correlated with GPA at all three time points. The ability of the SAT to predict second year cumulative GPA in biology majors is similar to other majors (Shaw, Kobrin, Patterson, & Mattern, 2012). However, the same study found that the SAT is more highly correlated with second year cumulative GPA in male biology students than in female biology students. Both the SAT and ACT have also been found to be predictive of graduation rates, but they are better at predicting between-school graduation rates than predicting within-school graduation rates (Stumpf & Stanley, 2002).

The current study

The current study uses measures that are traditionally used to predict student success to control for preexisting ability levels of students in treatment group participating in an authentic research-based science program and a comparison group participating in a more traditional program. As is mentioned above, few studies examining authentic research-based learning use a comparison group, and no studies found used a comparison group and accounted for differences in preexisting ability levels of the students. This analysis will attempt to separate the advantages of learning science in an authentic research-based science program from learning students would have achieved in a more traditional program.

Authentic Research-based Science Program

The Advanced Program for Integrated Science and Math (PRISM) was initiated at the university level and funded by the vice provost. The National Research Council suggests integration of the various biological disciplines and of the broader scientific disciplines, especially when teaching undergraduate students (National Research Council (U.S.), 2009). The purpose of PRISM is to create a community of learners that is able to integrate scientific disciplines while engaging in authentic research by having students complete all introductory science courses in the first two years of undergraduate study. PRISM is primarily meant for majors in the natural sciences, although any student who meets the minimum qualifications may apply. Most PRISM students are life sciences majors, while very few are not majoring in any science (see Table 1). Admission requirements for PRISM include demonstrated achievement in high school and either an SAT math score over 720, an SAT Math II score of 620, or an ACT Math score of 31.

Table 1. *Student Demographics (N = 3296).*

		PRISM		Control	
Variables	Demographic Groups	*n*	%	*n*	%
Gender	Male	88	59.5	1309	41.6
	Female	60	40.5	1839	58.4
Ethnicity	Asian	23	15.5	368	11.7
	Black non-Hispanic	2	1.4	267	8.5
	Hispanic	18	12.2	675	21.4
	White non-Hispanic	89	60.1	1542	49.0
	Not specified	16	10.8	296	9.4
COHORT	2008	9	6.1	134	4.3
(entering	2009	15	10.1	473	15.0
	2010	20	13.5	594	18.9
	2011	30	20.3	628	19.9
	2012	34	23.0	639	20.3
	2013	40	27.0	680	21.6
MAJOR	Biology	43	29.1	747	23.7
	Biochemistry	21	14.2	122	3.9
	Marine Science	8	5.4	310	9.8
	Exercise Physiology	5	3.4	214	6.8
	Microbiology & Immu-	15	10.1	207	6.6
	Neuroscience	31	20.9	305	9.7
	Psychology	2	1.4	261	8.3
	Other Science	16	10.8	144	4.6
	Undeclared	7	4.7	838	26.6

The PRISM curriculum has students complete the basic introductory sciences and calculus in the first two years of study. It is the same curriculum as for the regular science track, except that classes are smaller and somewhat more rigorous, each of which could account for achievement differences. PRISM students take Biology with labs, Inorganic Chemistry with labs, and Calculus during their first year. Introductory courses are PRISM only, primarily lecture-based, courses with approximately 60 to 75 students (compared to the general population lectures with approximately 230 students per section). The second year includes Organic Chemistry with labs, Calculus-based Physics with labs, and a semester of Scientific Computing. PRISM students must also choose a second-year biology course to take with the general biology population. Students must maintain a 3.5 cumulative GPA and a 3.0 GPA in PRISM courses to remain in PRISM.

The biggest differences between the PRISM curriculum and the general curriculum are that PRISM introductory labs engage students in authentic research and the integrated PRISM courses are more quantitative and research-based than their traditional counterparts. Kloser, Brownell, Chiariello, and Fukami (2011) offered six recommendations for creating research-based student courses that also contribute to the instructor's research: 1) small need for technical expertise, 2) checks and balances to ensure quality, 3) variables that offer choices to students without overwhelming the instructional team, 4) a central standardized database into which students can upload data, 5) authentic assessments to match the structure of the course, and 6) involvement of instructors with expertise in the study system. Following a similar system, first-year students in

PRISM labs are overseen by a biology professor (instead of by a graduate student) as they create their own experiments based on the research currently occurring in the professor's lab. For example, students whose professor was an ecologist designed experiments about biodiversity, while students whose professor was a geneticist designed experiments that involved gene sequencing. As the students engage in authentic original research, they must prove to their instructors that the answers to their research questions do not already exist in the literature. At the end of the semester, students give conference-style poster presentations of their research.

The general biology labs taken by non-PRISM students are more traditional inquiry-based, taught at a lower level of inquiry. Students were given an online lab manual, which they read before each class and about which they took quizzes each week. Graduate-student instructors then instructed students in how to conduct a sample experiment in which the students learned the skills necessary for the investigation through hands-on practice. Students were then given a limited number of materials they could request in order to conduct a similar experiment with one or two independent variables changed from the sample. The general biology laboratory course covered two major topics per semester, with the students having three to four weeks to conduct and analyze their experiments.

Method

The study used a retrospective case-comparison research design. Records from Fall 2008 to Spring 2014 were collected for 5866 students whose majors matched those of PRISM students. Variables indicating success prior to beginning college and after beginning college were included in models as described below. Due to existing correlations between some of the variables and the experimental condition (i.e. PRISM students had, on average, higher admission test scores and high school GPA), multiple linear regression was used rather than ANCOVA (Field, 2009).

Research Setting

The research was conducted in a large private non-profit university in the southeastern United States. In the fall of 2013, 51% of the undergraduate students in the university were female, 12% were Asian/Pacific Islander, 8% were Black, 27% were Hispanic, 50% were White non-Hispanic, and 3% were two or more races. Of the 12,733 undergraduate students in the university, 729 were biology majors, making biology the largest major in the university. The mean average combined SAT score for students entering the university in 2013 was 1325 and the mean composite ACT score was 30.1. The mean high school percentile for the same students was 90.7, and 72% of students were in the top 10% of their high school class.

Participants

Out of the 5866 students studied, 264 were participants in the PRISM program. Students had to specifically apply to the PRISM program, so these participants were self-selected and selected by the faculty based on their college admissions test scores and high school GPA. Students who had removed themselves from the PRISM program at the time of data collection were included in the comparison group, while students who had graduated as PRISM students were retained in the PRISM group. 2570 students were removed from the analysis because they were missing admission test scores and/or high school GPA, leaving 3296 students in the analysis. The students removed were similar to the students retained according to an Analysis of Covariance conducted on the available data, except that students in the 2008 cohort (the earliest cohort examined) were more likely to be removed due to missing data. This data loss may be attributed to a data migration that occurred in 2013. Demographics for students in PRISM and in the comparison group are shown in Table 1 and college admissions test scores and GPA data are shown in Table 2.

Data Collection

Data from before college admission (e.g., admissions test scores and gender) were collected from student admissions records. Data created while the student was at the university (e.g., major and grades) were collected from student registrar records with institutional review board approval. The data were collected and de-identified by a university data specialist and then forwarded to the researcher to ensure anonymity. In cases of discrepancy, the latest available data were used for each student.

Table 2. *Student Scores on Admission Tests and GPAs (N=3296).*

		PRISM			Control		
Measure	Component	*n*	*M*	*SD*	*n*	*M*	*SD*
SAT	Math	106	717.74	49.57	2483	653.58	75.59
	Reading	106	680.57	55.94	2483	640.67	74.63
	Total	106	1398.30	84.74	2483	1294.24	132.21
	z-score of total	106	0.64	0.64	2483	-0.03	1.00
ACT	Math	148	32.76	1.95	1993	28.68	4.13
	Reading	148	32.26	2.95	1993	29.75	4.49
	English	148	32.24	2.81	1993	29.63	4.46
	Science	148	31.18	3.12	1993	28.00	4.23
	Composite	148	32.01	1.86	1993	28.84	3.67
	z-score of composite	148	0.75	0.48	1993	-0.02	0.95
TEST	Highest of SAT or ACT *z*-score	148	0.98	0.48	3148	0.12	0.94
GPA	HSGPA	148	4.57	0.71	3148	4.34	0.88
	CGPA	148	3.66	0.33	3148	3.32	0.57
	BGPA	148	3.59	0.43	2515	3.04	0.85
	UGPA	93	3.53	0.67	1543	3.26	0.83

Data Analysis

Variables

The data were analyzed three times, with outcome variables representing more specific achievement outcomes in each analysis. Each outcome was calculated based on the grades available at the end of the Spring 2014 semester. The first analysis used cumulative college GPA (CGPA) as the most general outcome measure. CGPA was obtained directly from the registrar. Because PRISM is composed mostly of life sciences majors, the second analysis used the total college biology GPA (BGPA). BGPA was calculated by the researcher from student grades, assigning 0 points to an F, 1 point to a D, 2 points to a C, 3 points to a B, and 4 points to an A. Intermediate grades, i.e., B+, were given intermediate points according to the university’s grading system. The grades were further weighted to account for the number of credits each course was worth. The third analysis used the college upper-level biology GPA (UGPA; i.e., introductory courses were not included) in order to determine if the established predictive value of college admissions tests (Bridgeman et al., 2000; Coyle & Pillow, 2008; Patterson & Mattern, 2013; Robinson & Monks, 2005; Rothstein, 2004; Sawyer, 2010) and high school GPA (Rothstein, 2004) on achievement in the first year of college extended past the first year. UGPA was also meant to account for the fact that PRISM students all took the first year biology courses together, but took the upper-level courses with the general science-major population. UGPA was calculated using the same method as BGPA, except that entry-level courses were excluded from the calculation. Due to the increasing specificity of the analyses, CGPA has the highest N, followed by BGPA, with UGPA having the smallest N.

Participation in the PRISM program (PRISM), gender (FEMALE), entering class (COHORT), and major area of study (MAJOR) were entered into each model as background predictor variables. COHORT was examined to determine if the changes that occurred in both the treatment and control programs over the years had an effect on student achievement. MAJOR was examined out of concern that biology majors would be more likely to excel in biology courses. The covariate TEST was created by calculating z-scores (which estimates the standard deviation of test scores) for each the total SAT score and the composite ACT score for each student. For students with only an SAT or an ACT score, the score provided was included in TEST. For students with both an SAT score and an ACT score, the highest z-score was included in TEST. Students' high school GPA (HSGPA) was also included as a covariate. COHORT was dummy coded into each year of matriculation, with 2008 excluded as the reference group because it was the first cohort to participate in PRISM. MAJOR was dummy coded into each major of study, with undeclared as the reference group, because those students did not have a major.

Differences in student achievement

Three separate multiple linear regression models were examined to determine if student achievement, as measured by CGPA, BGPA, and UGPA, differed by PRISM. TEST, HSGPA, FEMALE, COHORT, and MAJOR were entered into the models as predictors to determine if student characteristics other than participation in PRISM affected student achievement. TEST and HSGPA were included because they are often used as predictors of student achievement and as proxies of student potential. Additionally, students in the PRISM group were selected to have higher TEST and HSGPA, so these variables must be controlled for in order to determine the actual effect of PRISM on student achievement. FEMALE was included in the analysis to account for the historical underrepresentation of women in the sciences (Szelényi, Denson, & Inkelas, 2013). COHORT was included to account for changes in the curriculum and faculty over the course of the intervention. MAJOR was included to account for potential differences in interest and motivation in biology classes held by students who major in fields more closely related to biology compared to those majoring in fields less closely related to biology. The descriptive statistics of the variables used in the regression analysis are shown in Tables 1 and 2.

Because the comparison group was much larger than the treatment group, the data was analyzed both using all available data and by case-matching by TEST and HSGPA (Stuart, 2010). While TEST and HSGPA decreased in their predictive power as expected, there was almost no change in the effect of PRISM in any of the models, so the larger data set, which gave a more comprehensive picture of the effects of the predictor variables, was retained for the analyses.

Results

Effect of PRISM on Cumulative GPA

Table 3 presents the results from the linear regression analysis predicting CGPA. The model explained a statistically significant proportion of the variance in CGPA, $R^2 = .44$, $F(17, 3278) =$

45.35, $p < .001$. The predictors combined to account for 44% of the variance in CGPA, which is a large effect size (Cohen, 1988). PRISM, TEST, HSGPA, FEMALE, and some MAJORS were significant predictors.

Table 3. *Results of Regression Analysis for CGPA.*

Source	Coefficient	*SE*	*t*	*p*
Intercept	2.71	0.06	42.17	<.001
PRISM	0.14	0.04	3.25	.001
TEST	0.19	0.01	18.27	<.001
HSGPA	0.11	0.01	9.95	<.001
FEMALE	0.10	0.02	5.43	<.001
COHORT				
2009	0.05	0.05	0.94	.345
2010	0.00	0.05	-0.08	.936
2011	0.00	0.05	-0.02	.981
2012	-0.04	0.05	-0.93	.354
2013	0.01	0.05	0.14	.888
MAJOR				
Biology	0.05	0.03	1.85	.064
Biochemistry	0.09	0.05	1.85	.065
Marine Science	-0.02	0.03	-0.58	.561
Exercise Physiology	0.19	0.04	4.95	<.001
Microbiology/Immunology	0.14	0.04	3.58	<.001
Neuroscience	0.16	0.03	4.71	<.001
Psychology	0.11	0.04	3.06	.002
Other science	0.04	0.04	0.91	.361

R.Q. 1a. The unstandardized coefficient for PRISM was 0.14, $t(3278) = 3.25$, $p = .001$. The resulting coefficient of 0.14 indicates that participation in PRISM was associated with an increase in CGPA of 0.14 points, after accounting for the other variables in the model.

R.Q. 1b. The unstandardized coefficient for TEST was 0.19, $t(3278) = 18.27$, $p < .001$. The resulting coefficient of 0.19 indicates that each one point increase in the *z*-score of a student's admission test was associated with an increase in CGPA of 0.19 points, after accounting for the other variables in the model.

The unstandardized coefficient for HSGPA was 0.11, $t(3278) = 9.95$, $p < .001$. The resulting coefficient of 0.11 indicates that each one point increase in high school GPA was associated with an increase in CGPA of 0.11 points, after accounting for the other variables in the model.

The unstandardized coefficient for FEMALE was 0.10, $t(3278) = 5.43$, $p < .001$. The resulting coefficient of 0.10 indicates that female students were expected to have a CGPA that was 0.10 points higher than males, after accounting for the other variables in the model.

Each COHORT was examined individually as a dichotomous predictor of CGPA, but no COHORT was a significant predictor of CGPA. This result indicates that all students had similar CGPAs regardless of COHORT, after accounting for the other variables in the model.

Each MAJOR was examined individually as a dichotomous predictor of CGPA. Biology, Biochemistry, Marine Science, and Other Science were not found to lead to significantly different CGPAs after accounting for the other variables in the model. The unstandardized coefficient for Exercise Physiology was 0.19, $t(3278) = 4.95$, $p < .001$. The resulting coefficient of 0.19 indicates that Exercise Physiology majors were expected to have a CGPA that was 0.19 points higher than undeclared majors, after accounting for the other variables in the model. The unstandardized coefficient for Microbiology/Immunology was 0.14, $t(3278) = 3.58$, $p < .001$. The resulting coefficient of 0.14 indicates that Microbiology/Immunology majors were expected to have a CGPA that was 0.14 points higher than undeclared majors, after accounting for the other variables in the model. The unstandardized coefficient for Neuroscience was 0.16, $t(3278) = 4.71$, $p < .001$. The resulting coefficient of 0.16 indicates that Neuroscience majors were expected to have a CGPA that was 0.16 points higher than undeclared majors, after accounting for the other variables in the model. The unstandardized coefficient for Psychology was 0.11, $t(3278) = 3.06$, $p < .001$. The resulting coefficient of 0.11 indicates that Psychology majors were expected to have a CGPA that was 0.11 points higher than undeclared majors, after accounting for the other variables in the model.

Table 4. *Results of Regression Analysis for BGPA.*

Source	Coefficient	*SE*	*t*	*p*
Intercept	2.12	0.11	19.84	<.001
PRISM	0.22	0.07	3.31	.001
TEST	0.31	0.02	16.89	<.001
HSGPA	0.14	0.02	8.04	<.001
FEMALE	-0.04	0.03	-1.31	.190
COHORT				
2009	0.03	0.08	0.45	.655
2010	0.05	0.07	0.67	.505
2011	0.06	0.07	0.75	.454
2012	-0.09	0.07	-1.26	.208
2013	0.01	0.07	0.18	.861
MAJOR				
Biology	0.32	0.05	6.92	<.001
Biochemistry	0.33	0.07	4.37	<.001
Marine Science	0.22	0.06	3.77	<.001
Exercise Physiology	0.14	0.07	2.10	.036
Microbiology/Immunology	0.46	0.06	7.28	<.001
Neuroscience	0.39	0.06	6.88	<.001
Psychology	0.15	0.06	2.35	.019
Other science	0.25	0.08	2.99	.003

Effect of PRISM on Biology GPA

Table 4 presents the results from the linear regression analysis predicting BGPA. The model explained a statistically significant proportion of the variance in BGPA, $R^2 = .21$, $F(17, 2645) = 22.80$, $p < .001$, which is a medium effect size (Cohen, 1988). The predictors combined to account for 21% of the variance in BGPA. PRISM, TEST, HSGPA, and some MAJORS were significant predictors.

R.Q. 1b. The unstandardized coefficient for PRISM was 0.22, $t(2645) = 3.31$, $p = .001$. The resulting coefficient of 0.22 indicates that participation in PRISM was associated with an increase in BGPA of 0.22 points, after accounting for the other variables in the model.

R.Q. 2b. The unstandardized coefficient for TEST was 0.31, $t(2645) = 16.89$, $p < .001$. The resulting coefficient of 0.31 indicates that each one point increase in the z-score of a student's admission test was associated with an increase in BGPA of 0.31 points, after accounting for the other variables in the model.

The unstandardized coefficient for HSGPA was 0.14, $t(2645) = 8.04$, $p < .001$. The resulting coefficient of 0.14 indicates that each one point increase in high school GPA was associated with an increase in BGPA of 0.14 points, after accounting for the other variables in the model.

The unstandardized coefficient for FEMALE was not significant, $t(2645) = -0.04$, $p = .190$. This result indicates that all students had similar BGPAs regardless of gender, after accounting for the other variables in the model.

Each COHORT was examined individually as a dichotomous predictor of BGPA, but no COHORT was a significant predictor of BGPA. This result indicates that all students had similar BGPAs regardless of COHORT, after accounting for the other variables in the model.

Each MAJOR was examined individually as a dichotomous predictor of BGPA. All declared MAJORS were found to lead to significantly different BGPAs after accounting for the other variables in the model. The unstandardized coefficient for Biology was 0.32, $t(2645) = 6.92$, $p < .001$. The resulting coefficient of 0.32 indicates that Biology majors were expected to have a BGPA that was 0.32 points higher than undeclared majors, after accounting for the other variables in the model. The unstandardized coefficient for Biochemistry was 0.33, $t(2645) = 6.92$, $p < .001$. The resulting coefficient of 0.33 indicates that Biochemistry majors were expected to have a BGPA that was 0.33 points higher than undeclared majors, after accounting for the other variables in the model. The unstandardized coefficient for Marine Science was 0.22, $t(2645) = 3.77$, $p < .001$. The resulting coefficient of 0.22 indicates that Marine Science majors were expected to have a BGPA that was 0.22 points higher than undeclared majors, after accounting for the other variables in the model. The unstandardized coefficient for Exercise Physiology was 0.14, $t(2645) = 2.10$, $p = .036$. The resulting coefficient of 0.14 indicates that Exercise Physiology majors were expected to have a BGPA that was 0.14 points higher than undeclared majors, after accounting for the other variables in the model. The unstandardized coefficient for Microbiology/Immunology was 0.46, $t(2645) = 7.28$, $p < .001$. The resulting coefficient of 0.46 indicates that Microbiology/Immunology majors were expected to have a BGPA that was 0.46 points higher than undeclared majors, after accounting for the other variables in the model. The unstandardized coefficient for Neuroscience was 0.39, $t(2645) = 6.88$, $p < .001$. The resulting coefficient of 0.39 indicates that Neuroscience majors were expected to have a BGPA that was 0.39 points higher than undeclared majors, after accounting for the other variables in the model. The unstandardized coefficient for Psychology was 0.15, $t(2645) = 2.35$, $p = .019$. The resulting coefficient of 0.15 indicates that Psychology majors were expected to have a BGPA that was 0.15 points higher than undeclared majors, after accounting for the other variables in the model. The unstandardized coefficient for Other Science was 0.25, $t(2645) = 2.99$, $p = .019$. The resulting coefficient of 0.25 indicates that Other science majors were expected to have a BGPA that was 0.25 points higher than undeclared majors, after accounting for the other variables in the model.

Effect of PRISM on Upper-level Biology GPA

Table 5 presents the results from the linear regression analysis predicting UGPA. The model explained a statistically significant proportion of the variance in UGPA, $R^2 = .32$, $F(17, 1618) = 10.84$, $p < .001$. The predictors combined to account for 32% of the variance in UGPA, which is a large effect size (Cohen, 1988). PRISM, TEST, HSGPA, and some MAJORS were significant predictors.

Table 5. *Results of Regression Analysis for UGPA.*

Source	Coefficient	*SE*	*t*	*P*
Intercept	2.37	0.16	15.23	<.001
PRISM	0.06	0.09	0.70	.484
TEST	0.23	0.02	9.27	<.001
HSGPA	0.12	0.02	4.94	<.001
FEMALE	-0.04	0.04	-1.00	.317
COHORT				
2009	0.10	0.09	1.11	.268
2010	0.08	0.09	0.97	.330
2011	0.08	0.09	0.90	.367
2012	-0.01	0.09	-0.07	.941
2013	0.24	0.10	2.56	.011
MAJOR				
Biology	0.29	0.10	3.08	.002
Biochemistry	0.18	0.12	1.53	.127
Marine Science	0.15	0.11	1.38	.167
Exercise Physiology	0.15	0.13	1.09	.278
Microbiology/Immunology	0.35	0.11	3.24	.001
Neuroscience	0.28	0.10	2.67	.008
Psychology	0.14	0.12	1.19	.233
Other science	0.46	0.14	3.23	.001

R.Q. 1c. The unstandardized coefficient for PRISM was not significant, $t(1618) = 0.70$, $p = .317$. This result indicates that all students had similar UGPAs regardless of participation in PRISM, after accounting for the other variables in the model.

R.Q. 2c. The unstandardized coefficient for TEST was 0.23, $t(1618) = 9.27$, $p < .001$. The resulting coefficient of 0.23 indicates that each one point increase in the z-score of a student's admission test was associated with an increase in UGPA of 0.23 points, after accounting for the other variables in the model.

The unstandardized coefficient for HSGPA was 0.12, $t(1618) = 4.94$, $p < .001$. The resulting coefficient of 0.12 indicates that each one point increase in high school GPA was associated with an increase in UGPA of 0.12 points, after accounting for the other variables in the model.

The unstandardized coefficient for FEMALE was not significant, $t(1618) = -0.04$, $p = .317$. This result indicates that all students had similar UGPAs regardless of gender, after accounting for the other variables in the model.

Each COHORT was examined individually as a dichotomous predictor of UGPA, and only 2013 was significant. The unstandardized coefficient for 2013 was 0.24, $t(1618) = 2.56$, $p = .011$. The resulting coefficient of 0.24 indicates that each one point increase in high school GPA was associated with an increase in UGPA of 0.24 points, after accounting for the other variables in the model. However, it should be noted that the only 2013 students included in this model are expected to be high achievers, because they are freshmen taking upper-level biology courses.

Each MAJOR was examined individually as a dichotomous predictor of UGPA. Biochemistry, Marine Science, Exercise Physiology, and Psychology were not found to lead to significantly different UGPAs from undeclared majors after accounting for the other variables in the model. The unstandardized coefficient for Biology was 0.29, $t(1618) = 3.08$, $p = .002$. The resulting coefficient of 0.29 indicates that Biology majors were expected to have a UGPA that was 0.29 points higher than undeclared majors, after accounting for the other variables in the model. The unstandardized coefficient for Microbiology/Immunology was 0.35, $t(1618) = 3.24$, $p < .001$. The resulting coefficient of 0.35 indicates that Microbiology/Immunology majors were expected to have a UGPA that was 0.35 points higher than undeclared majors, after accounting for the other variables in the model. The unstandardized coefficient for Neuroscience was 0.28, $t(1618) = 2.67$, $p = .008$. The resulting coefficient of 0.28 indicates that Neuroscience majors were expected to have a UGPA that was 0.28 points higher than undeclared majors, after accounting for the other variables in the model. The unstandardized coefficient for Other Science was 0.46, $t(1618) = 3.23$, $p = .001$. The resulting coefficient of 0.46 indicates that Other Science majors were expected to have a UGPA that was 0.46 points higher than undeclared majors, after accounting for the other variables in the model.

Discussion and Implications

This study examined the effect of an authentic research-based curriculum on undergraduate biology student achievement. It further examined the predictive value of college admissions test scores and high school GPA on biology student achievement.

Discussion

R.Q. 1.

The results indicate that the PRISM curriculum, which was rich in authentic research, had a significant effect on student achievement as measured by cumulative undergraduate GPA and biology GPA, but that effect was reduced to insignificance when introductory classes that were exclusive to PRISM students were excluded from the GPA calculation. This result is consistent with the literature on instruction through authentic research (Brownell et al., 2012; Kazempour et al., 2012; Schroeder et al., 2007), which usually examines short-term outcomes. This result may be due to the higher and less variable admission test scores and high school GPA found in PRISM students compared to comparison students representing a higher expected achievement in PRISM students on average regardless of curriculum. The lack of significance in the UGPA model may also indicate that the grade benefits of PRISM exist more in the short-term than in the long-term, emphasizing the importance of researchers using long-term measures of achievement. The sample population for both the PRISM and comparison groups also changed over time, as struggling students were more likely to leave both PRISM and the challenging science majors. This differential attrition based on ability level may have made the PRISM and comparison groups more similar after freshman year. Finally, it is possible that the non-PRISM introductory biology labs were at a high enough level of inquiry to allow those students similar benefits to those provided by PRISM.

R.Q. 2.

FEMALE was included in the analysis to account for the historical underrepresentation of women in the sciences (Szelényi et al, 2013). While gender was clearly not a predictor of achievement in the biology department studied, female students apparently did better in their classes overall than male students with similar characteristics. COHORT was examined out of concern that teaching methods and instructors tend to change over time, so it was possible that students in each cohort would have had different exposure to the material. However, COHORT did not have an effect in any model, except that the few freshmen advanced enough to have already taken upper-level biology courses were likely to do better in those courses than non-freshmen taking upper-level courses.

MAJOR was examined out of concern that biology majors would be more likely to excel in biology courses. Students in all of the majors studied had higher expected BGPAs than undeclared students, which may be an indication of focus on the biology content shown by students with a strong interest in biology-related fields. The UGPAs varied among majors, with Biology, Microbiology/Immunology, Neuroscience, and Other Science majors having higher grades in upper-level biology courses than students with other majors examined, which may be for similar reasons as the pattern seen in BGPA. The CGPAs of the majors varied differently, with Exercise Physiology, Microbiology/Immunology, Neuroscience, and Psychology students having higher CGPAs than the other majors studied. This may be an indication of differences in the difficulty levels of the programs, or it may be due to unmeasured attributes of those students.

The results further indicate that TEST and HSGPA both had significant effects on all outcome measures. The fact that they were entered into the model together allows us to infer that they are independently predicting student success, which is consistent with the idea that the two measures should be used together (Cohn et al., 2004). However, contrary to assertions of some researchers (Atkinson, 2002; Atkinson & Geiser, 2009; Geiser, 2009; Rothstein, 2004), but consistent with others (Mattern et al., 2010; Noble & Sawyer, 2002), college admissions test scores predicted more of the between-student variance than did HSGPA. While the composite z-score used here combined SAT and ACT scores in order to increase sample size and therefore power, exploratory models using the SAT total z-score or the ACT composite z-score showed similar results. It is interesting to note that TEST also predicted all three outcome measures better than PRISM did, which suggests that the abilities with which students came into college had a bigger effect on student GPA than did the authentic research-based curriculum.

Implications

Contributions and limitations

This study contributes to a growing body of literature demonstrating the advantages of authentic research in improving undergraduate student achievement. First, while there are many possible measures of student achievement, three different GPAs were used in this analysis to determine student achievement overall, in biology, and in advanced biology classes. Second, this study included college admissions test scores and HSGPA as proxies for preexisting student ability. Third, this study demonstrated that student achievement varied across majors. Fourth, this study provides evidence that efforts to decrease the gender achievement gap in biology have been successful. Finally, the large sample size of this study allows the results to be powerful and reliable.

This study had several limitations. First, the retrospective nature of the study meant only the data available could be used. There was a lot of missing data, which led to almost half of the available students being removed from the analysis. However, the students retained had similar characteristics to those removed. Second, the inclusion of class rank as a predictor would have been beneficial (Cohn et al., 2004), but there was too much missing class rank data to make its use viable for this study. Third, because the program is in progress and multiple cohorts and majors were used, there was variety in the students included in each analysis, e.g., freshmen were unlikely to have a UGPA and some students were still in the program at the time of the analysis. Finally, during the last couple of years of the study, the comparison group switched to a more authentic-inquiry style lab in introductory courses. However, PRISM students still had much more autonomy in the labs than comparison students (i.e., PRISM used a higher level of inquiry), and COHORT was not a significant predictor in any of the models, with the exception of freshmen taking upper-level courses.

Implications for future research

The results of this study can have an impact on future research. First, it is important for researchers examining the potential of authentic research to improve student achievement to include predictors such as TEST and HSGPA in the model. These variables allow the researcher to rule out the students' native abilities as the entire cause of increased academic success. Second, the differences seen between the UGPA model and the CGPA and BGPA models suggest that freshman GPA is not necessarily the best measure of student achievement (Rothstein, 2004), especially used alone. Factors that are traditionally used to predict achievement in the first year of college, college admissions test scores and HSGPA, are useful predictors of longer-term biology success.

Implications for practice

The results of this study can have a potential impact on undergraduate biology education. First, the study supports the literature that authentic research is beneficial for students (Brownell et al., 2012; Kazempour et al., 2012), in this case biology students, specifically. However, since the benefit seems to be somewhat short-lived, the authentic research should be included throughout the curriculum to provide reinforcement. Second, the study demonstrates the effectiveness of both admission test scores and HSGPA at predicting student achievement as students advance through a biology program, which is information that can be used in college admissions.

In conclusion, authentic research is beneficial to biology student achievement, but college admission test scores and HSGPA are even better predictors of achievement after freshman year, and must be taken into account when evaluating the effectiveness of undergraduate biology programs.

References

American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action*. Washington D.C.: American Association for the Advancement of Science.

Atkinson, R. C. (2002). Achievement versus aptitude in college admissions. *Issues in Science and Technology*, 18(2), 31-36.

Atkinson, R. C. & Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher*, 38(9), 665-676.

Bell, R. L., Smetana, L., & Binns, I. (2005). Simplifying inquiry instruction. *The Science Teacher*, 72(7), 30-33.

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning test.* (No. 2000-1). New York: College Entrance Examination Board.

Brownell, S. E., Kloser, M. J., Fukami, T., & Shavelson, R. (2012). Undergraduate biology lab courses: Comparing the impact of traditionally based "cookbook" and authentic research-based courses on student lab experiences. *Journal of College Science Teaching*, 41(4), 36-45.

Buck, L. B., Bretz, S. L., & Towns, M. H. (2008). Characterizing the level of inquiry in the undergraduate laboratory. *Journal of College Science Teaching*, 38(1), 52-58.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohn, E., Cohn, S., Balch, D. C., & Bradley Jr., J. (2004). Determinants of undergraduate GPAs: SAT scores, high-school GPA and high-school rank. *Economics of Education Review*, 23(6), 577-586. doi:10.1016/j.econedurev.2004.01.001

Coyle, T. R., & Pillow, D. R. (2008). SAT and ACT predict college GPA after removing g. *Intelligence*, 36(6), 719-729. doi:10.1016/j.intell.2008.05.001

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles: SAGE Publications.

Geiser, S. (2009). Back to the basics: In defense of achievement (and achievement tests) in college admissions. *Change*, 41(1), 16-23.

Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: Is it the school that matters? *Economics of Education Review*, 29(6), 911-922. doi:http://iiiprxy.library.miami.edu:4357/10.1016/j.econedurev.2010.06.010

Kazempour, M., Amirshokoohi, A., & Harwood, W. (2012). Exploring students' perceptions of science and inquiry in a reform-based undergraduate biology course. *Journal of College Science Teaching*, 42(2), 38-43.

Kloser, M. J., Brownell, S. E., Chiariello, N. R., & Fukami, T. (2011). Integrating teaching and research in undergraduate biology laboratory education. *PLoS Biology*, 9(11)

Mattern, K. D., & Patterson, B. F. (2011). *Validity of the SAT for predicting third-year grades: 2006 SAT validity sample*. (No. 2011-3). New York: The College Board.

Mattern, K. D., Shaw, E. J., & Kobrin, J. L. (2010). *A case for not going SAT-optional: Students with discrepant SAT and HSGPA performance* (AERA 2010 presentation). Denver, CO.

National Research Council (U.S.). (2000). *Inquiry and the national science education standards*. Washington, DC: National Academies Press.

National Research Council (U.S.). (2003). *BIO2010: Transforming undergraduate education for future research biologists*. Washington, DC: The National Academies Press.

National Research Council (U.S.). (2009). *A new biology for the 21st century*. Washington, DC: The National Academies Press.

National Research Council (U.S.). (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

Noble, J., & Sawyer, R. (2002). *Predicting different levels of academic success in college using high school GPA and ACT composite score*. No. 2002-4. Iowa City, IO: ACT.

Patterson, B. F., & Mattern, K. D. (2013). *Validity of the SAT for predicting first-year grades: 2011 SAT validity sample*. (No. 2013-3). New York: The College Board.

Radunzel, J., & Noble, J. (2013). *Differential effects on student demographic groups of using ACT college readiness assessment composite score, ACT benchmarks, and high school grade point average for predicting long-term college success through degree completion*. (No. 2013-5). Iowa City, IA: ACCT.

Robinson, M., & Monks, J. (2005). Making SAT scores optional in selective college admissions: A case study. *Economics of Education Review*, 24(4), 393-405. doi:10.1016/j.econedurev.2004.06.006

Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics*, 121(1–2), 297-317. doi:http://iiiprxy.library.miami.edu:4357/10.1016/j.jeconom.2003.10.003

Sawyer, R. (2010). *Usefulness of high school average and ACT scores in making college admission decisions*. (No. 2010-2). Iowa City, IA: ACT.

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T., & Lee, Y. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching*, 44(10), 1436. doi:10.1002/tea.20212

Shaw, E. J., Kobrin, J. L., Patterson, B. F., & Mattern, K. D. (2012). *The validity of the SAT for predicting cumulative grade point average by college major*. No. 2012-6. New York: The College Board.

Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1-21. doi:10.1214/09-STS313

Stumpf, H. & Stanley, J. C. (2002). Group data on high school grade point averages and scores on academic aptitude tests as predictors of institutional graduation rates. *Educational and Psychological Measurement*, 62(6), 1042-1052. doi:10.1177/0013164402238091

Szelényi, K., Denson, N., & Inkelas, K. K. (2013). Women in STEM majors and professional outcome expectations: The role of living-learning programs and other college environments. *Research in Higher Education*, 54(8), 851-873. doi:http://dx.doi.org/10.1007/s11162-013-9299-2

Timmerman, B. E., Strickland, D. C., & Carstensen, S. M. (2008). Curricular reform and inquiry teaching in biology: Where are our efforts most fruitfully invested? *Integrative and Comparative Biology*, 48(2), 226-240.

Tuckman, B. W. & Kennedy, G. J. (2011). Teaching learning strategies to increase success of first-term college students. *The Journal of Experimental Education*, 79(4), 478-504. doi:10.1080/00220973.2010.512318

Whalen, D. F. & Shelley, M. C. (2010). Academic success for STEM and non-STEM majors. *Journal of STEM Education: Innovations and Research*, 11(1/2), 45-60.

Xu, H. & Talanquer, V. (2013). Effect of the level of inquiry on student interactions in chemistry laboratories. *Journal of Chemical Education*, 90(1), 29-36. doi:10.1021/ed3002946

Corresponding Author:
Brandon S. Diamond
dr.bdiamond@gmail.com
David Posnack Jewish Day School
Paul and Maggie Fischer High School
5810 South Pine Island Road
Davie, FL 33328
754-800-4373