Journal of Science Education and Technology, Vol. 11, No. 4, December 2002 (© 2002)

Improving Biology Performance with Workshop Groups

Wendi K. Born,[1] William Revelle,[1] and Lawrence H. Pinto[2,3]

This 2-year quasi-experiment evaluated the effect of peer-led workshop groups on performance of minority and majority undergraduate biology students. The workshop intervention used was modeled after a program pioneered by Treisman (1992). Majority volunteers randomly assigned to workshops ($n = 61$) performed significantly better than those assigned to the control group ($n = 60$, $p < 0.05$) without spending more time studying. Workshop minority students ($n = 25$) showed a pattern of increasing exam performance in comparison to historic control minority students ($n = 21$), who showed a decreasing pattern ($p < 0.05$). Volunteers ($n = 121$) initially reported that biology was more interesting and more important to their futures than to nonvolunteers' ($n = 435$, $p < 0.05$). Volunteers also reported higher levels of anxiety related to class performance ($p < 0.05$). The relationship of anxiety to performance was moderated by volunteer status. Performance of volunteers was negatively associated with self-reported anxiety ($r = -0.41$, $p < 0.01$). Performance of nonvolunteers was unrelated to self-reported anxiety ($r = -0.02$). Results suggest elevated anxiety related to class performance may increase willingness to participate in activities such as workshop interventions. In addition, students who volunteer for interventions such as workshops may be at increased risk of performance decrements associated with anxiety. Even so, workshop programs appear to be an effective way to promote excellence among both majority and minority students who volunteer to participate, despite the increased risk of underperformance associated with higher levels of anxiety.

KEY WORDS: group learning; biology performance; minority groups; academic achievement; stereotyped–attitudes.

INTRODUCTION

According to the National Science Board (NSB), the college years are a time of such profound attrition from the sciences that in 1987 they described this attrition as "a grave long-term threat to the Nation's scientific and technical capacity, its industrial and economic competitiveness, and the strength of its national defense" (National Science Board, Task Committee on Undergraduate Science and Engineering Education, 1987, p. 1). Further research identified the primary factors associated with students majoring in the sciences as (a) participating in the college-preparatory track in high school; (b) taking the most demanding math and science courses; (c) being European American or Asian American; (d) being male; (e) coming from a family with high socioeconomic status; and (f) having a scientist/engineer for a parent (U.S. Congress, Office of Technology Assessment, 1988). Although the number of science and engineering degrees increased until 1996, and has remained level since then (National Science Board, 2002), concerns about the number of scientists working in the United States remain. These concerns center on the increasingly rapid changes in the areas of science and technology and the influence of technology on the economic welfare of the United States as the world's leader in high-tech exports (Doyle, 2002).

The NSB originally suggested two fronts on which to attack the problem of attrition from the

[1]Department of Psychology, Northwestern University, Evanston, Illinois.
[2]Department of Neurobiology and Physiology, Northwestern University, 2153 North Campus Drive, Evanston, Illinois 60208.
[3]To whom correspondence should be addressed; e-mail: larry-pinto@northwestern.edu

sciences in college: improve the quality of education and attract a new pool of future scientists. The three major problems identified in undergraduate institutions were (a) uninspired laboratory instruction; (b) faculty with out-of-date disciplinary knowledge; and (c) unimaginative, out-of-date courses and curricula. The final recommendations of the report suggested that the National Science Foundation stimulate spending in the sciences and expand efforts that would lead to increased participation by female and minority students (National Science Board Task Committee on Undergraduate Science and Engineering Education, 1987, p. 4). This research was designed to address both of these issues.

Underrepresentation of African American and Hispanic American students at the highest levels of achievement remains acute. (Hereafter "minority" will refer to African American and Hispanic American students; "majority" will refer to Asian American and European American students.) Minorities make up a rapidly increasing proportion of the population, but are not making commensurate gains in the highest levels of achievement. The difference in the proportion of students of different ethnicities scoring above average on the Scholastic Aptitude Test (SAT), Quantitative section (SAT-Q) is marked. One third of all Asian American students and one fifth of all European American students score 600 or higher, while only 3.4% of African American students and 7.4% of Mexican American students score at this level or above (Miller, 1995). Unfortunately, even those minority students who do have above-average qualifications frequently underperform in relation to their European American peers. In one study of equally qualified engineering majors, European American students were 20–25% more likely to graduate with engineering degrees than were their minority peers (Miller, 1995). Minority students of both sexes continue to be drastically underrepresented at the doctoral level, especially in the fields of engineering and the natural sciences, earning 2.1% of the PhDs in engineering and 3.6% of the PhDs in the natural sciences (Hill, 1996). Should minority students enter the sciences in levels proportionate to their population in the United States, there would be greater overall gains in the number of U.S. scientists than if European American males returned to their 1980 level of participation in the sciences (Doyle, 2002).

Across ethnicities, women are more likely than men to enroll and graduate from high school, college, and master's programs, but less likely to enter natural science or engineering fields, and less likely to earn doctoral degrees, regardless of field of study (Hill, 1996; Vetter, 1990). There continue to be more men than women pursuing science at every level of the "pipeline" from middle school on. Unlike the comparison between minorities vs. nonminorities, however, there is evidence that women are as likely or *more* likely to have high achievement in the sciences when compared with men who have similar standardized test scores and attitudes toward science. Nevertheless, these women are still more likely than men to leave the science pipeline after a brief exposure, resulting in a loss of talented women from the sciences (Hanson, 1996).

Of the factors that have been associated with the pursuit of math and science majors, few are amenable to intervention at the undergraduate level. Improving college preparation *after* arrival would seem to be the most likely avenue for intervention. In the past, as many as 25% of science and engineering majors have come from outside of the traditional college-preparatory track (U.S. Congress, Office of Technology Assessment, 1988), suggesting that academic preparation during, rather than before the undergraduate years, is a viable way to increase pursuit of math and science careers.

For the last 30 years, many institutions have attempted to compensate for underpreparation at the college level by requiring remedial course work from students identified as likely to have trouble in certain areas. Often the students identified are minorities. Sadly, there is little evidence that remedial course work in math and science has been effective at increasing participation in the sciences, encouraging college persistence, or contributing to degree attainment. It could be argued that remedial math and science courses are even more uninspiring, more unimaginative, and more out-of-date than are regular college courses, and therefore less likely to result in increased academic interest in these areas for any student. Some, such as Bonsangue and Drew (1995), have proposed that "achievement among underrepresented minority students in mathematics, science, and engineering disciplines may be less associated with pre-college ability than with in-college academic experiences and expectations" (p. 32).

Steele and colleagues (Steele, 1992; Steele and Aronson, 1995) have proposed that the mere existence of well-known stereotypes undermines the performance of women and minorities at all levels of education by placing them in situations that are threatening to their sense of themselves as able human beings. Because widely held stereotypes identify women and minorities as less able in math and science, these individuals may constantly feel as if they were suspected of being less able. When minority students go to college, they are likely to be more of a minority of the population than ever before, and placement in a remedial program sends the implicit message that they need help because of their race. Negative stereotypes posit an additional burden for academically stigmatized groups with which other groups do not have to cope. Steele (1992) hypothesized that most minority students find themselves in a no-win situation where there are only negative associations to their academic efforts and that these students often choose to disinvest from areas in which they feel threatened, rather than risk proving their critics right. Steele suggested that remedial programs harm retention efforts by exaggerating prevalent negative stereotypes.

The *21st Century Program* (Steele, 1997) was an experimental intervention modeled after The Professional Development Program Mathematics Workshop (Treisman, 1992; reviewed later), but more extensive in scope. A preliminary report of outcome data from this ongoing study offered evidence that remedial programs did not improve performance of African American students (and may even have depressed performance). However, interventions designed to be more academically challenging than regular course work did improve performance (Steele, 1997). In the 21st Century Program, African American students in remedial programs performed worse than any other students, and their grades were not reliably related to their SAT scores. African American students who were not in any type of program (control) performed slightly better than students who received remedial attention. Importantly, grades for this group of students were positively correlated with SAT scores. African American students in both the control and remedial groups performed worse than a group of European American control students of equal ability.

In contrast, African American students in the experimental program showed a strong relationship of SAT scores to performance, and their grades were as high as the grades of European American students. At the highest levels of ability (defined by SAT scores), African American students in the experimental program far outperformed African American students in the control and remedial programs. This relationship did not hold for the bottom third of the distribution of SAT scores, where African American students in the experimental program performed similarly to those in the remedial and control groups. There was also a trend for European American students in the experimental program to perform better than European American control students, suggesting that the benefits of participation in this type of program are not limited to minority students.

Although the evidence that remedial programs hurt minority students is preliminary, there is no evidence that remedial programs have been successful in encouraging women and minorities to pursue even advanced course work in math and the sciences, to say nothing of majors or careers. This track record is unlikely to provide colleges with any successful way to recruit additional women and minorities into these programs of study. Fortunately, evidence that small group workshops have improved performance for participants in general, and for minority students in specific, is beginning to accrue.

Workshop Interventions

The use of workshop groups as a method of encouraging academic excellence grew out of an investigation of the study habits of different groups of calculus students. Treisman (1992) noticed that Asian American students frequently obtained high grades in freshman calculus, while African American students often obtained low grades. He and colleagues studied African American and Asian American students and concluded that the differences between thee groups were due to study habits. African American students most often spent the recommended 6–8 h a week studying for calculus. They generally worked alone and were academically isolated, frequently unaware of how their efforts or grades compared with those of other students. In contrast, Asian American students devoted about 14 h a week to calculus. Similar to the African American students, 8–10 of those hours were spent studying alone, but the additional 4–6 h were spent in group study. Asian American students were often part of a network of peers who got together in groups to share knowledge about calculus and other common interests, and to critique each other's work. These groups typically included someone who had already taken calculus who quizzed the other students, or helped them to see flaws in their thinking. Given this frame of reference, Asian American students were able to assess their progress and performance in the class and identify areas where improvement was needed. Treisman and associates created the workshop model with the immediate goal of encouraging success through academic community among calculus students, and the long-term goal of producing more, and more diverse, mathematicians (Treisman, 1992).

The Professional Development Program Mathematics Workshop was initiated in 1978 to improve calculus performance by eliminating academic isolation. The program was described to students as a noncredit honors program, which would involve four additional hours of group study per week on material more challenging than that assigned in the regular course. Students of all races were recruited, but on average, 80% of participants were African American or Hispanic American. Graduate students who served as a group facilitators monitored individual groups of 5–7 students. Facilitators were meant to serve as peer advisors by asking questions and encouraging productive efforts without doing the work for the students.

This program was remarkably successful in increasing the number of minority students who enrolled in and successfully completed a first semester calculus course designed for science and engineering majors (Fullilove and Treisman, 1990). During this period, only 3% of African American workshop participants received grades of D or F in calculus, compared with 40% of nonparticipants and 33% of a historical control group. What is more, African American students who participated in the workshop as freshman went on to earn degrees at Berkeley, "at rates comparable to those of (European American and Asian American) students" (p. 468). Participants were 25% more likely to persist in a major requiring mathematics and earn a degree than were African American students who did not participate in the workshop and a historical control group of African American students (Fullilove and Treisman, 1990). This pattern of success was evident at all levels of precollege preparation, including students who were classified as economically disadvantaged and those who were classified as "special" admissions because they did not meet the standard admission criteria.

Although a dozen colleges are reported to have instituted programs in calculus modeled after Treisman's workshop at University of California at Berkeley (Selvin, 1992), and more than a hundred colleges have conducted trial runs of similar programs (Bonsangue and Drew, 1995), there are few comprehensive outcome evaluations of these and similar programs. Those programs for which there are some data available include *MathExcel*, a calculus program at the University of Kentucky at Lexington (Freeman, 1995, 1997), the *Academic Excellence Workshop Program*, a calculus program at California State Polytechnic University of Pomona (Bonsangue and Drew, 1995), the *Emerging Scholars Program*, a calculus program at the University of Texas at Austin (McCaffrey and Myers, 1994), the *Medical Scholars Program*, a 1st-year medical program at the University of California San Francisco School of Medicine (Fullilove *et al.*, 1988), the *Workshop Chemistry Project*, a chemistry program at City College of New York (Gosser *et al.*, 1996), and *BioExcel*, a new workshop program in Biology at University of Kentucky (Cohen, 1997).

These interventions have reported similar effects, despite the different subject areas and populations on which they have focused. All have reported improvements in graded performance, and many have reported improvements in student retention in the sciences and subsequent research participation. Unfortunately, given the various methods of reporting data, it is not possible to report effect sizes for these programs (for a detailed review of these programs, see Born, 2000).

Workshop programs extending Treisman's work share common elements with cooperative learning groups. In cooperative learning groups, students work together with the goal of maximizing achievement for all members. Members are encouraged to help each other and do not compete with others in their own group. Although most of the research on cooperative learning has been done with children, there is some evidence that cooperative learning groups conducted with college students can yield effect sizes that range from 0.26 (Slavin, 1995) up to 0.6 (Johnson, *et al.*, 1991). Although these researchers disagree about the magnitude of the advantage attributable to cooperative learning, they provide extensive reviews of the vast literature on cooperative learning groups, and advocate cooperative learning as an educational tool (Johnson *et al.*, 1981; see also Slavin, 1983).

Motivational Theories of Underperformance

Underperformance occurs when performance falls significantly below what would be expected on the basis of ability. Many different researchers have investigated the problem of underperformance, and the result has been a seemingly confusing array of theories that highlight the importance of a number of cognitive, affective, motivational, situational, and social variables related to performance. These theorists have primarily tried to explain achievement behavior on an episodic level, explaining behavior during circumscribed achievement tasks as a function of specific beliefs, thoughts, or goals. Some of the most influential theories that have come out of this tradition are learned helplessness (Seligman, 1975), reactance theory (Wortman and Brehm, 1975), attributional style theory (Abramson *et al.*, 1978), and mastery vs. performance goal orientation (Dweck, 1986; Dweck and Leggett, 1988).

In contrast, classic motivational theories have historically related achievement behaviors to the dynamic interplay of two *separate* motivational systems, one focused on the pursuit of reward and the other on the avoidance of punishment (Atkinson and Birch, 1970; Lewin *et al.*, 1944; McClelland *et al.*, 1953), which were later applied specifically to achievement motivation (Atkinson and Birch, 1974, 1986). This dynamic model proposed that behavior in achievement contexts is the product of an individual's expectancy of success and the subjective value placed on success vs. failure. This model predicted that stable individual differences in motivation (i.e., pursue achievement, avoid failure) would interact with situational variables (i.e., task difficulty, consequences of success/failure) to result in temporary emotional states (i.e., anxiety, pride, shame) that would have predictable motivational effects on future achievement behavior. Other theorists have contributed to the tradition of dynamic thinking based on approach and avoidance motives by examining such things as the motivational consequences of achievement outcomes (Kuhl and Blankenship, 1979; Revelle and Michaels, 1976), future orientation in achievement motivation (Raynor, 1969, 1970), the relationship of approach and avoidance motivation to the adoption of proximal achievement goals (Elliot and Church, 1997), test anxiety (Elliot and McGregor, 1999), and the relationship of motivational differences to academic self-regulation and achievement (Pintrich and de Groot, 1990).

When an individual seeks achievement despite fear of failure, the predictable result is a state of anxiety (Atkinson and Birch, 1986; Gray, 1987). Even individuals who have high motivation to avoid failure often choose to participate in achievement tasks, either because they also have a high need for achievement, or because there is an external incentive that is valuable to them. Because anxiety is unpleasant, the experience of anxiety can affect the valuation of and participation in future opportunities for achievement (Raynor, 1970).

The relationship of anxiety and achievement has long been a focus of research (Covington and Omelich, 1991; Mandler and Sarason, 1952; Sarason, 1961; Wine, 1971). People experiencing anxiety during achievement tasks/tests have been found to devote more attention to rumination and worry about evaluation, while the attention of those who are not anxious remains focused on the task (Kuhl and Blankenship, 1979). When a task requires a great deal of cognitive resources, the diversion of attention from the task to the self will result in performance decrements. People with high trait anxiety show increased attention to threat-relevant cues, whereas people who are generally nonanxious show decreased attention to threat cues (Mathews, 2002). A strong tendency (trait) to avoid failure will result in the experience of frequent anxiety states (Watson and Clark, 1984), which will further bias an individual toward cues for failure. On complex tasks, anticipation of failure, subsequent anxiety, and resulting decrements in performance can become a vicious, self-fulfilling cycle.

Often failures can have far-reaching consequences for an undergraduate working to complete a particular major or degree. In competitive gateway courses such as biology, students are striving for grades that will increase their chances of participating in some future opportunity (research, medical school), often of even greater difficulty. This increasingly difficult contingent path serves to increase fear of failure (and therefore anxiety) associated with each new task (Raynor, 1969). Paradoxically, avoidance of anxiety often becomes an end in itself, leading to task avoidance, delayed task initiation, and reduced task persistence. Eventually both anxiety and failure may be avoided by quitting.

Stereotype Threat

The theory of stereotype threat (Steele, 1997; Steele and Aronson, 1995) can be rephrased in the language of approach-avoidance motivation with little difficulty. Because of the existence of prevalent stereotypes, the possible consequences of a task should be different for students to whom a relevant negative stereotype could be applied. The very nature of a negative stereotype may increase the subjective cost of failure, which would reflect on everyone in the individual's group, without a commensurate increase in the value of success, which would reflect only on the individual. The high cost of failure sets up a situation where the motive to avoid achievement failure is likely to be strong, regardless of individual disposition. On very difficult tasks it will be impossible to avoid failure at the item level, and it is likely that the individual will recognize that his or her current performance includes failures. This realization should increase anxiety and divert cognitive resources from the task to the self, which will result in overall performance decrements (Humphreys and Revelle, 1984; Wine, 1971). The more the individual values and identifies with the domain, the more sensitive he or she will be to both reward and punishment in the domain. Prior to frustration and failure, this value for the domain will ensure approach motivation and achievement behavior. With repeated frustration and failure, however, continued participation in the activity will become very costly because the expectancy for competency will decrease, creating the lose–lose situation described by Steele. Over time, repeated negative experiences will result in decreased expectation of success and increased expectation of failure. Inhibition of achievement activities and devaluing of achievement are ways to avoid the full negative effects of failure. Over time, stereotype-relevant failures (and successes) become less personally important (less punishing, less rewarding). When this happens, individuals will be less likely to seek experiences in the relevant area, since there is little left to motivate performance except the specific consequences of the task and the ever-present threat of confirming the stereotype. As long as the negative stereotype is prevalent, the risk of confirming it will always increase the negative consequences of participation for members of the stereotyped group.

This study conceptualized achievement in biology from a dynamic perspective, but was not designed to be a test of dynamic models of achievement. Workshop groups were expected to improve biology performance by imparting experiences and skills that should increase the subjective expectancy of success, reduce the fear of doing poorly, and thereby allow students to maintain their value for biology and remain engaged and enrolled. This was expected to have especially beneficial effects for minority students who may have had additional pressure to avoid failure at all costs.

METHOD

Participants

Participants were 585 students enrolled in Biology 210, Quarter 1, in 1997 or 1998, who took at least one exam. Students' ethnicities were reported as either African American ($n = 16$), Asian American ($n = 260$), European American ($n = 292$), Hispanic American ($n = 14$), Native American ($n = 2$), or unknown ($n = 2$). Volunteers were 121 European American and Asian American students (majority students) who asked to participate in the Biology Honors Workshop program, enrolled in biology, and took at least one exam. These students were randomly designated as either workshop participants ($n = 61$) or motivational control participants ($n = 60$) prior to the 1st week of class. Nonvolunteers were 435 majority students who did not ask to participate in workshops, enrolled in biology, and took at least one exam.

Selection

All African American and Hispanic American students (minority students) enrolled in the class were recruited to participate in workshops and 25 did so (4 declined). Minority students enrolled in biology during the prior year (1996; $n = 21$) were used as an historic control group for minority students who participated in workshops.

Selection of the 121 majority participants was a more complex process. In the 1st year (1997), majority students eligible to enroll in biology ($n = 615$) were contacted by mail and asked if they would be interested in participating in the workshop program if invited to do so. A median split of combined SAT-Verbal (SAT-V) and SAT-Quantitative (SAT-Q) scores, or if none, American College Test—English (ACT-E) and American College Test—Math (ACT-M) scores, was performed for the 72 majority students who responded positively. Students were grouped by race, sex, and high vs. low test scores, stratified within groups by GPA, and divided into blocks of 3. Within each block, one of the three students was randomly selected to participate in the workshop program; the remaining two were designated as control participants. Originally, 25 students were selected to participate in the workshop and 47 were left in the control group. Of the 25 majority students invited to participate, 5 did not enroll. These students were replaced by returning to the original block from which each declining student had been selected, and randomly selecting one of the remaining students to participate. All of these replacements participated in the workshop. Of the control participants, 12 did not enroll, resulting in a total of 25 majority workshop and 30 majority control participants in 1997.

During the 2nd year (1998), the method of soliciting and selecting majority students was simplified. Students eligible to enroll in biology

($n = 326$) were contacted via e-mail. The e-mail briefly described the Biology Honors Workshop program that had begun the year before and requested that students reply if they were interested in making the weekly commitment necessary to participate. The 67 students who responded positively and preenrolled in biology were selected. These students were grouped according to race and sex and randomly assigned to either the workshop ($n = 35$) or control group ($n = 32$). In the end, 2 control participants were not enrolled following the drop-add period, and one student attended the introductory workshop meeting without invitation and was allowed to participate, resulting in 36 workshop and 30 control participants.

All 25 workshop minority participants completed Quarter 1. Of the 61 workshop participants, 2 dropped after the first exam, and 59 completed Quarter 1. Of the 60 control participants, 6 dropped after the first exam, and 54 completed Quarter 1.

Additional volunteers were added at the beginnings of Quarters 2 and 3 to fill vacancies, but no additions were made to the control group. At the beginning of Quarter 2, 48 students from Quarter 1 continued, and 23 new students were invited to participate. At the beginning of Quarter 3, 41 students from Quarter 2 continued (including 34 from Quarter 1), and 12 new students were invited to participate, including one newly enrolled African American student (male).

Assignment to Workshop Groups

Participants were assigned to workshop groups on the basis of the need for groups to be representative of the larger group makeup and scheduling constraints. Quarter 1 groups initially had 4–7 members which included both male and female students, both European and Asian American students, and at least one minority student. During Quarters 2 and 3 there were more groups than there were minority workshop participants, so it was not possible to have a minority student in each group.

Facilitators

The faculty member directing the workshops was responsible for selecting all 15 facilitators. Facilitators were junior and senior undergraduate science majors. In 1998, eight of nine facilitators had been workshop participants the prior year. Of facilitators, eight were European American (five females, three males), four were Asian American (two females, two males), two were African American (one female, one male), and one was Hispanic American (male).

Facilitators in 1997 completed 3 days of facilitator training at The University of Texas in Austin (UT) approximately 1 month before the intervention began. In 1998, two facilitators from 1997 returned as "senior facilitators." Senior facilitators helped with facilitator training and provided quality control for the 1998 workshop groups by monitoring each group two times during Quarter 1 and evaluating group functioning. In 1998, all nine facilitators and both senior facilitators completed 2 days of facilitator training at Northwestern University approximately 2 weeks before the workshops began. The training was based on the training experience the prior year, but tailored to fit the biology curriculum and the specifics of the Northwestern workshop program.

Facilitators met once a week for 1 h with the course instructor to go over the problem sheets for the coming week and to discuss concerns. Worksheets were developed independent of course exams, and facilitators were unaware of specific exam content. Problems were designed to be conceptually related to material covered in the biology class, but more challenging than homework problems. Problem sets used in the workshops were electronically posted each week so that they were available to all members of the biology class.

Group Activities

Groups met once a week for 2 h to work on problem sets. Group members were told that working together as a mutually supportive team had been shown to result in higher grades and better understanding of complex material. Members were encouraged to think of mistakes and initial failures as necessary steps to improve problem-solving skills. The group was characterized as a place for members to share problem-solving strategies and corrective feedback that would enable everyone to learn new ways to think and solve problems. Stereotypes were not directly addressed within the groups.

The role of the facilitator was to encourage student members to spend most of their time working on the problems in groups, and to ensure that group interactions resulted in movement toward problem solutions, but not to provide answers for students.

Table I. Measures of Ability for Nonvolunteer, Control, Workshop Majority, and Workshop Minority Students

	Nonvolunteers (Majority)			Volunteers: Control (majority)			Volunteers: Workshop majority			Volunteers: Workshop minority		
Ability	*M*	SD	*n*	*M*	SD	*n*	*M*	SD	*n*	*M*	SD	*n*
GPA[a]	3.40	0.37	401	3.34	0.38	57	3.38	0.41	60	3.07	0.46	25
ACT[b]	30.50	2.31	229	29.97	2.37	37	30.12	2.23	42	28.27	3.47	15
SAT[c]	1401	100	378	1362	117	54	1364	129	50	1250	120	21

[a]College Grade Point Average immediately prior to enrollment in biology; no differences between groups.
[b]American College Test Composite scores; no differences between groups.
[c]Combined Verbal and Quantitative Scholastic Aptitude Test scores; nonvolunteers differ from volunteers (control + workshop majority); $p < 0.01$.

Facilitators reinforced the idea that the groups were based on teamwork and that each member had a valuable contribution to make. Facilitators also recorded the attendance and participation of each group member.

Procedure

Students who volunteered to participate in the Biology Honors Workshop were randomly assigned to either the workshop or control condition as detailed previously. Workshop students attended a mandatory meeting at the beginning of the quarter. This meeting was used to introduce students to facilitators and other program personnel, introduce the idea of cooperative work groups, explain the role of the facilitators, and respond to any last minute needs that might require adjustment of the groups. Food and beverages were served and students were encouraged to ask questions and socialize. Workshop groups began meeting the 2nd week of classes and continued to meet weekly for 2 h through the end of the quarter (eight meetings).

In 1998, the entire biology class ($n = 303$), including workshop and control students, was surveyed via e-mail twice during the quarter. The first survey was sent during Week 5, between the first and second exams, and the final survey was sent after the second exam, and before the final exam. The surveys assessed hours of study per week, expected grade, feelings about biology as a subject, and the estimated importance of biology to the student's future. In addition, the first survey assessed anxiety related to class performance (as well as motivation to master the material, and motivation to do well in relation to peers, but these will not be discussed here), and sensitivity to prevalent academic stereotypes based on race and sex.

Measures

Premeasures

Enrollment status, sex, race/ethnicity, SAT scores (Verbal and Quantitative scores combined), ACT scores (Composite score), and GPA immediately prior to enrollment in biology (on 4.0 scale) were obtained from the university registrar. Table I lists SAT scores, ACT scores, and GPA for each group of students.

Performance Outcome Measures

Class performance was defined as the total number of points earned in each quarter, including three exam scores and a lab score, each worth 25% of a student's grade (obtained directly from the instructor). Exam performance was defined as the number of points earned on each exam. The examinations consisted of approximately equal parts of three types of questions (a) short-answer questions in which students were asked to write a brief explanation of why a certain result was obtained in a key experiment; (b) multiple-choice, matching, or fill-in-the-blank questions; and (c) problem-solving questions in which students were asked to interpret a certain experimental result or design an experiment. Retention in biology was operationally defined as the number of exams taken during the year. Only those students who took the first exam in Quarter 1 were included in analyses of retention.

Survey Measures

Students were surveyed twice via e-mail (during Weeks 5 and 10) of each quarter regarding a variety of attitudes related to biology. In Quarter 1, 201 replies

(66%) were received for Survey 1 and 163 replies for Survey 2 (54%).

On Survey 1, anxiety related to class performance was assessed with two items ("I worry about the possibility of getting a bad grade in this class," and "I often think to myself, what if I do poorly in this class"). Students were asked to report their experience of being stereotyped according to race and gender ("How often do others judge your academic ability based on prevalent racial stereotypes?" and "How often do others judge your academic ability based on prevalent gender stereotypes?"). Participants reported their level of agreement with these statements on an anchored 6-point scale.

The two-item scale of anxiety related to class performance had an alpha reliability of 0.84 ($n = 183$) and test-retest reliability of 0.63 ($n = 94$) over an interval of 16 weeks. Notably, a five-item anxiety scale, which was composed of the two anxiety related to class performance items, plus two items assessing worry, and one item assessing test anxiety, was not more reliable than the two-item avoidance of failure scale. Although the two scales were virtually interchangeable, further analyses used the anxiety related to class performance scale rather than the larger anxiety scale for theoretical reasons. The anxiety related to class performance scale specifically assessed avoidance of failure relative to biology class, which was the type of motivation the workshops might be expected to address. There is no theoretical reason to believe that participation in a biology workshop group would affect anxiety in general.

Other questions administered on both Surveys 1 and 2 asked students to "rate your current interest in the subject of biology," "rate the importance of biology to your future plans," and "rate how much you like the subject-matter in biology this quarter" on an anchored 6-point scale. Students were also asked to report the number of hours they spent studying for biology each week.

During Quarters 2 and 3, follow-up surveys included an item that asked students to indicate whether they would participate in summer research and to identify where they would be working. Students who answered both parts of the question were counted as participating in summer research.

RESULTS

Majority volunteers and nonvolunteers were compared to isolate preexisting differences in ability and to look for correlates of volunteerism. Workshop and control groups were then contrasted to test for motivational or performance effects of workshop participation. Workshop minority and historic control minority students were contrasted to isolate any preexisting differences on measures of ability. Comparing biology performance of historic control minority students and workshop minority students tested the immediate effect of workshop participation on minority performance in biology. Because there were no survey data available for the historic control group, workshop minority students were compared with workshop majority students to test for motivational differences. Male and female students were contrasted to test for differences in biology performance as a function of sex. The self-report of being judged academically on the basis of prevalent gender or racial stereotypes was examined both in relation to biology performance and as a function of sex and ethnicity. Retention and summer research in biology, both considered to be longer-term measures of engagement and performance in biology, were examined for relationship to workshop participation. Finally, class performance was examined as a function of group and anxiety related to class performance. Except where otherwise specified, variables that have been standardized were standardized on the basis of the mean and standard deviation of the nonvolunteers. Therefore, all standard scores are expressed relative to the majority students who were in neither the workshop nor control group. An alpha level of 0.05 was used for all statistical tests.

The Effect of Volunteerism for Majority Students

To isolate possible ability and motivational correlates of volunteerism, the academic credentials and initial survey responses of the volunteers were compared to those of the nonvolunteers.

Ability Measures

Although there were no differences found for GPA or ACT scores, volunteers had lower SAT scores ($M = 1363$) than did nonvolunteers ($M = 1401$; $F(1, 480) = 11.48$, $p < .01$, Table I). The difference was significant for both SAT-V and SAT-Q scores.

Motivation Measures

In contrast to the general similarity of groups on measures of ability, volunteers were quite different

Table II. Survey Measures of Motivation for Nonvolunteer, Control, Workshop Majority, and Workshop Minority Students

				Volunteers								
	Nonvolunteers (Majority)			Control (majority)			Workshop majority			Workshop minority		
Ability	*M*	SD	*n*	*M*	SD	*n*	*M*	SD	*n*	*M*	SD	*n*
Survey 1												
Anxiety[a]	0.00	1.00	131	0.46	0.93	17	0.27	0.96	35	0.60	0.97	14
Interest[a]	0.00	1.00	135	0.33	0.79	16	0.53	1.10	35	0.28	0.85	14
Importance[b]	0.00	1.00	133	0.13	0.57	16	0.40	0.58	36	0.42	0.53	14
Liking	0.00	1.00	133	−0.24	1.07	16	0.00	1.15	36	−0.17	1.23	14
Study hours	0.00	1.00	133	0.00	1.37	17	−0.13	0.76	36	0.57	1.67	14
Survey 2												
Interest	−0.32	1.20	106	−0.87	1.37	14	−0.04	1.07	30	−0.32	1.11	13
Importance	−0.05	0.98	106	−0.01	1.24	14	−0.03	1.17	30	0.00	1.01	13
Liking[c]	−0.50	0.97	106	−0.75	1.05	14	−0.18	0.82	30	−0.72	1.21	13
Study hours[d]	−0.72	0.85	106	0.39	2.66	14	−0.29	1.03	30	−0.26	0.85	13

[a] At Survey 1 nonvolunteers differ from volunteers (control + workshop majority); $p < 0.01$.
[b] At Survey 1 nonvolunteers differ from volunteers (control + workshop majority); $p < 0.05$.
[c] From Survey 1 to Survey 2 nonvolunteers show a steeper decline than volunteers (control + workshop majority); $p < 0.05$.
[d] From Survey 1 to Survey 2 nonvolunteers show a steeper decline than volunteers (control + workshop majority); $p < 0.01$.

from nonvolunteers on measures of motivation (assessed at Survey 1). Measures for nonvolunteers at Survey 1 were used to standardize variables at both Surveys 1 and 2. Therefore, all measures have been expressed relative to the nonvolunteer sample at Survey 1 (all means for nonvolunteers at Survey 1 = 0.00; see Table II).

On average, volunteers ($M = 0.33$) reported greater anxiety related to class performance than did nonvolunteers ($F(1, 181) = 4.29$, $p < 0.05$). Although volunteers and nonvolunteers reported similar liking of and hours of study for biology, volunteers reported biology to be more interesting and more important to their futures than did nonvolunteers. On the first survey, volunteers' reported interest in biology ($M = 0.46$) was higher than that reported by nonvolunteers ($F(1, 184) = 9.16$, $p < 0.01$). Volunteers also reported biology to be more important to their futures ($M = 0.32$) than did nonvolunteers ($F(1, 183) = 4.13$, $p < 0.05$), but this effect was somewhat smaller. Reported liking of biology and hours of study for biology did not differ as a function of volunteer status.

Each of the four self-report measures of engagement in biology was assessed again at the end of the quarter (Week 10). Univariate Repeated Measures Analysis was performed on each pair of variables to examine within-subjects change over the quarter and by group. Only students who completed both surveys were included in these analyses. Despite initial differences, volunteers and nonvolunteers reported similar interest in and importance of biology at Survey 2. Reported interest in and importance of biology dropped over the course of the quarter regardless of volunteer status. Interest in biology dropped a half of a standard deviation from Survey 1 ($M = 0.20$) to Survey 2 ($M = -0.31$; $F(1, 131) = 27.13$, $p < 0.001$). Importance of biology dropped a quarter of a standard deviation from Survey 1 ($M = 0.19$) to Survey 2 ($M = -0.05$; $F(1, 131) = 6.82$, $p < 0.01$). Interestingly, there was an interaction of group and change in liking of biology, such that nonvolunteers showed a much steeper drop over the course of the quarter than did volunteers ($F(1, 131) = 5.48$, $p < 0.05$). This same pattern occurred with hours of study. Nonvolunteers reported a steep drop in hours of study from Survey 1 to Survey 2, but volunteers reported similar hours of study ($F(1, 132) = 10.14$, $p < 0.01$). However, study hours were likely related to workshop participation, as detailed later.

Performance Measures

Consistent with expectations, volunteers and nonvolunteers earned an equal number of points in Quarter 1 (as well as in following quarters). Multivariate analysis of variance was used to examine performance across exams. Volunteers showed a pattern of increasing performance across exams, resulting in an interaction of volunteer status and exam performance over time ($F(2, 1056) = 5.60$, $p < 0.01$). Volunteers ($M = -0.27$) performed significantly worse than nonvolunteers on Exam 1 ($M = 0.00$; $F(1, 542) = 6.03$, $p < 0.05$), but were indistinguishable from nonvolunteers on Exams 2 and 3.

The Effect of Workshop Participation for Majority Students

As a result of workshop participation, workshop majority students were expected to report lower anxiety related to class performance than were control students. Consequently, majority volunteers assigned to participate in the workshop were expected to earn more points in biology than were those assigned to the control group.

Ability Measures

As a test of the assignment to conditions, the academic credentials of the workshop and control students who participated in the biology course were compared. As expected, no differences were found on measures of ability taken before the class. Workshop and control groups were compared on the basis of their academic credentials in subsequent quarters as well to test for the effect of attrition. In every case (GPA, ACT, SAT), there was no difference between workshop and control students (see Table I).

Motivation Measures

Because workshop participants were selected randomly from volunteers, workshop and control students were assumed to be equivalent on motivational measures prior to the class. Contrary to expectation, workshop and control students reported equal levels of anxiety related to class performance at Survey 1. Workshop and control students also could not be distinguished on the basis of reported interest in, importance of, liking of, or hours of study for biology at either Survey 1 or Survey 2. (Although reported study hours did not differ reliably from Survey 1 to Survey 2, ($F(1, 39) = 1.75$, $p = .19$), the direction of movement was different for the two groups. Workshop students reported reduced hours of study from Survey 1 to Survey 2 ($M = -0.21$), while control students reported increased hours of study ($M = 0.52$). Previous analyses indicated that there was a drop in study hours from Survey 1 to Survey 2 for nonvolunteer students ($M = -0.72$), making the increase for control students even more unusual.) Measures of motivation are listed in Table II.

Performance Measures

Despite equivalent ability and motivational goals, workshop students ($M = 0.17$) earned more total points than did control students ($M = -0.31$; $F(1, 111) = 6.25$, $p < 0.01$). Multivariate analysis of variance was used to examine performance across exams. Interestingly, both groups showed a pattern of increasing performance from Exam 1 to Exam 3. Quarter 1 performance as a function of group (controlling for GPA) is reported in Fig. 1. Workshop students continued to demonstrate an advantage over control students during follow-up at Quarters 2 ($F(1, 117) = 7.84$, $p < 0.01$) and 3 ($F(1, 93) = 21.04$, $p < 0.001$). Performance across all three quarters as a function of group (controlling for GPA) is reported in Fig. 2.

The Effect of Workshop Participation for Minority Students

Minority students who participated in the workshops in 1997 and 1998 were expected to earn more points in biology than minority students in the 1996 historical control group did. Workshop minority students were expected to perform as well as workshop majority students with similar academic qualifications. Workshop minority students were not expected to differ from workshop majority students on the basis of motivational measures.

Ability Measures

Contrasts of initial ability measures revealed that workshop minority students had higher GPAs ($M = 3.06$) than did historic control minority students ($M = 2.75$; $F(1, 43) = 7.09$, $p < 0.01$), but no differences were found for ACT or SAT scores.

Performance Measures

Multivariate analysis of variance was used to examine performance across exams. All minority performance analyses controlled for GPA. As in prior analyses, the points earned have been standardized relative to same-year majority students (*nonvolunteer* majority students for 1997 and 1998).

Although there was no difference in total points earned during the quarter by workshop minority and historic control minority students, there was an interaction of workshop status and exam performance over time. Workshop minority and historic control minority students performed similarly on Exams 1 and 2, but workshop minority students showed an increase

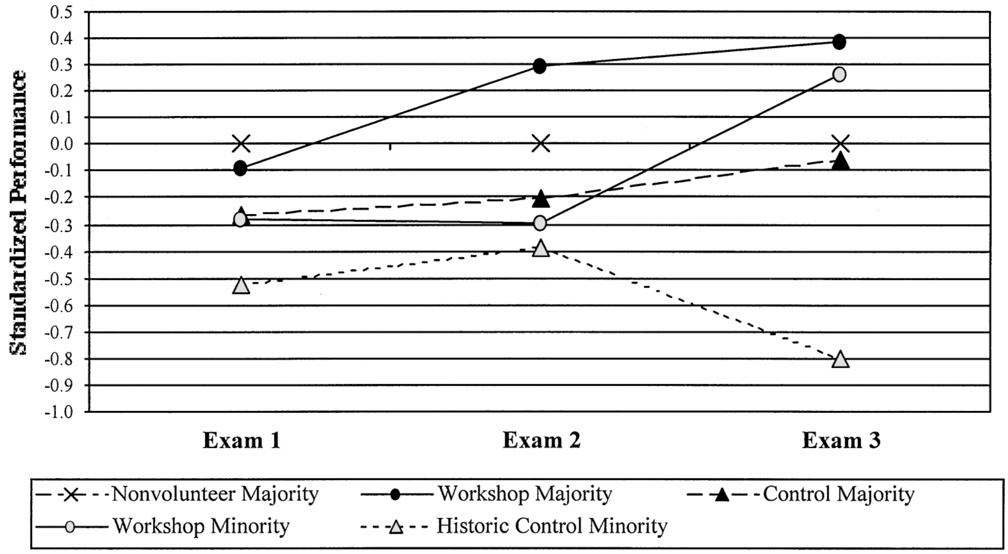

Fig. 1. Standardized Quarter 1 exam performance as a function of group, controlling for prior cumulative grade point average.

in performance on Exam 3, while historic control minority students showed a decrease in performance on Exam 3 ($F(2, 84) = 4.55$, $p < .05$). This interaction of group and exam can be seen in Fig. 1. The failure to detect overall differences between workshop minority students and historic control minority students on the basis of points earned is not very telling. The mean performance of workshop minority students consistently fell between that of control and workshop majority students. Controlling for GPA, workshop minority students could not be distinguished from either historic control minority students or workshop majority students with regard to overall performance at follow-up (see Fig. 2).

Although reported in a different metric, these results are similar to those reported by Treisman (1992).

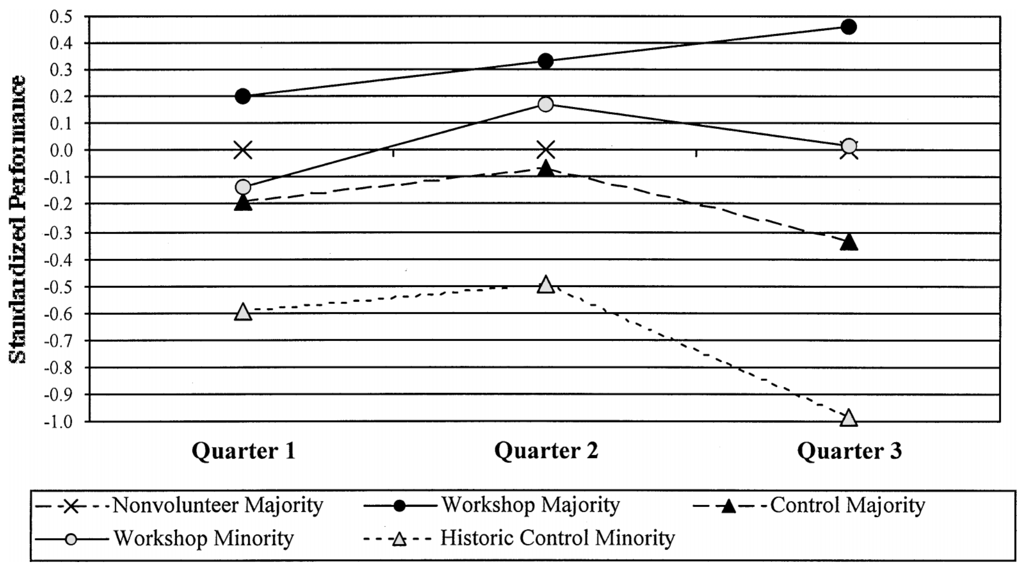

Fig. 2. Standardized performance in each quarter as a function of group, controlling for prior cumulative grade point average.

Table III. Report of Being Stereotyped by Gender and Race as a Function of Sex and Ethnicity

	Men			Women		
	M	SD	*n*	*M*	SD	*n*
All students						
Gender	0.10	0.99	71	0.55	1.14	107
Race	0.99	1.60	72	0.76	1.84	108
European American[a]						
Gender	0.00	1.00	35	0.29	0.88	63
Race	0.00	1.00	36	−0.24	1.03	63
Asian American						
Gender[b]	0.09	0.93	30	0.82	1.18	30
Race	2.00	1.55	30	2.36	1.74	31
African American						
Gender	0.85	0.67	2	1.32	2.14	6
Race	2.95	0.74	2	1.90	2.55	6
Hispanic American						
Gender	0.85	2.0	2	1.08	1.18	4
Race	1.90	0.74	2	2.69	0.53	4

[a]European American students reported being stereotyped by both gender and race less often than did Asian American, African American, and Hispanic American students; $p < 0.05$.
[b]Report of being stereotyped by gender differs as a function of sex; $p < 0.05$.

In the historic control group, 24% of all minority students who remained enrolled in Quarter 1 earned Ds or Fs. During this study, not a single workshop minority participant earned a D or an F or dropped the class. In contrast, of the four minority students who did not participate, two dropped the class after failing the first exam and one earned a D.

Motivation as a Function of Minority Status for Workshop Students

Workshop minority and workshop majority students were compared on the basis of the motivational variables with the expectation that they would be motivationally similar. At Survey 1, workshop minority students reported more hours of study ($M = 0.57$) than did workshop majority students ($M = -0.13$; $F(1, 47) = 4.08$, $p < 0.05$). Workshop minority and workshop majority students did not differ on self-report measures of anxiety related to class performance, interest in, importance of, or liking of biology at Survey 1. No differences at all could be detected between minority and majority workshop students at Survey 2.

The Experience of Being Stereotyped

The first survey asked students to rate how often others judged their academic ability on the basis of prevalent racial and gender stereotypes. Responses to the stereotyping questions were standardized on the basis of the responses of male, European American students and are depicted in Table III. Initial analyses were conducted using ethnicity and sex as factors.

As anticipated, Asian American ($M = 2.18$), Hispanic American ($M = 2.16$), and African American ($M = 2.43$) students all reported feeling more stereotyped by race than European Americans did ($M = -0.15$; $F(3, 166) = 41.05$, $p < 0.001$). Differences attributable to ethnicity accounted for 44% of the variance in self-report of feeling academically stereotyped because of race. Neither ethnicity nor self-report of being academically stereotyped according to prevalent racial stereotypes was related to performance when prior GPA was controlled.

Although female students reported being stereotyped by gender ($M = 0.55$) more often than male students did ($M = 0.10$; $F(1, 167) = 7.79$, $p < 0.01$), ethnicity was a better predictor than sex of feeling stereotyped by gender. African American ($M = 1.20$), Hispanic American ($M = 1.00$), and Asian American ($M = 0.46$) students of both sexes reported feeling more stereotyped by gender than did European American students ($M = 0.19$; $F(3, 167) = 3.61$, $p < 0.05$). Adding Ethnicity to the model allowed prediction of 10% of the variance in self- report of feeling stereotyped by gender. There was no interaction of sex and ethnicity.

Interestingly, Asian American students were the only students for whom feeling stereotyped by gender actually differed as a function of sex. Asian American women reported feeling more stereotyped by gender ($M = 0.82$) than did Asian American men ($M = 0.09$; $F(1, 58) = 7.80$, $p < 0.05$). Ratings of feeling stereotyped by gender did not differ as a function of sex for African American, Hispanic American, or European American students.

Because of the lack of a comparison group for minority students, female students were used to test whether or not performance was associated with the report of being stereotyped by gender, and if so, whether volunteer status or workshop participation affected the relationship. Contrary to prediction, biology performance was not directly related to self-report of being stereotyped by gender. However, female students' performance was effected by an interaction of volunteer status and report of being stereotyped by gender ($F(1, 89) = 8.99$, $p < 0.005$, $R^2 = 0.10$). Surprisingly, the report of being stereotyped by gender was reliably positively related to biology performance for volunteers ($r = .49$, $n = 31$, $p < 0.01$), but unrelated to performance for nonvolunteers ($r = -.15$, $n = 62$, $p = .25$).

Performance Differences as a Function of Sex and Group

Workshop participation was expected to ameliorate any male/female performance differences that occurred during the biology series in 1997 and 1998. In contrast to prior years, no sex differences occurred during this study (in any quarter). Sex did not interact with either volunteer status or workshop participation to predict performance in biology in any quarter.

Because the lack of sex differences in biology performance during this study was a surprising difference from the past year (1996), additional analyses were conducted to see if there were differences in initial ability measures for male and female students as a function of year, but none existed. The ability measures of students were strikingly similar across time, and failure to find sex differences during this study cannot be explained by a difference in level of ability of the students enrolled from year to year.

Retention Effects

Retention was defined as the number of exams taken over the course of the year. Three exams were given each quarter. Students who completed nine exams completed the entire biology sequence. Because prior cumulative GPA was related to retention ($r = .24$, $n = 517$, $p < 0.001$), all retention analyses used prior cumulative GPA as a control. Unexpectedly, volunteer status was related to retention. Volunteers took more exams ($M = 8.2$) than did nonvolunteers ($M = 7.7$; $F(1, 514) = 3.87$, $p < 0.05$). Contrary to prediction, however, workshop participation during the first quarter did not increase retention for workshop majority students, workshop female students, or workshop minority students over appropriate control students.

The proportion of students who successfully completed the biology sequence did not differ as a function of group status during Quarter 1, regardless of whether or not GPA was used as a control. Of the majority students, 75% of volunteers and 71% of nonvolunteers took all nine exams. Within the majority volunteers, 79% of workshop students and 72% of control students took all nine exams. Of the minority students, 64% of workshop minority students and 48% of historic control minority students took all nine exams. Of the four minority students who declined to participate in workshops, only one took all nine exams. Contrary to prediction, anxiety related to class performance was not reliably related to retention.

Participation in Summer Research

Contrary to prediction, workshop students ($M = 57\%$) were not more likely to participate in summer research in the natural sciences than were control students ($M = 38\%$; $F(1, 29) = 0.83$, $p = 0.37$) or nonvolunteers ($M = 40\%$; $F(1, 103) = 1.93$, $p = .17$).

Majority Student Performance as a Function of Anxiety and Group

Anxiety related to class performance was expected to be negatively associated with biology performance. No predictions were made about the relationship of anxiety, performance, and volunteer status. Anxiety related to class performance was expected to partially account for performance differences between majority workshop and control students.

Contrary to expectation, anxiety related to class performance was not related to performance for all majority students ($r = -.13$, $n = 158$). Further, the relationship of anxiety to performance was moderated by volunteer status, rather than workshop status.

For volunteers, anxiety related to class performance was highly negatively related to actual performance ($r = -.41$, $n = 50$, $p < 0.01$), whereas, for nonvolunteers, anxiety related to class performance showed no relationship to actual performance ($r = -0.02$, $n = 108$).

Among volunteers, anxiety related to class performance was equally negatively related to Quarter 1 performance, regardless of workshop participation and there was no interaction of anxiety with any other measure. Using a general linear model, the only significant predictors of Quarter 1 performance for students who volunteered were GPA, anxiety related to class performance, and workshop status ($R^2 = .44$).

DISCUSSION

This study was designed to test the effect of small group workshops on performance in biology and to investigate possible motivational explanations for those effects. The results offer conclusive evidence that workshops improved performance for majority participants and suggest that similar benefits were available for minority participants. These findings are consistent with the results of the other small group challenge interventions reviewed elsewhere (Born, 2000). Motivational results revealed significant differences between volunteers and nonvolunteers that were unanticipated. Because predicted motivational differences between workshop and nonworkshop students were not found, the mechanism for improved performance by workshop students remains unclear.

Performance Effects Attributable to Workshop Participation

Majority Students

Workshop majority students performed dramatically better than control students during Quarter 1, and the difference persisted during follow-up at Quarters 2 and 3. Because majority students were randomly assigned to groups and contrasts of initial measures of ability and motivation showed the groups to be equivalent, performance differences can be attributed to workshop participation.

The pattern of performance within Quarter 1 was similar for workshop and control students, yet workshop students performed at a higher level than did control students. Importantly, participating in the workshop appears to have improved performance without increasing the number of hours spent studying. If anything, the self-report data suggest that workshop students decreased their study hours near the end of Quarter 1 (as did nonvolunteers), while control students increased the hours they studied, although this tendency was not statistically significant. Report of study hours included hours spent in workshop groups, suggesting that the 2 h per week workshop students spent in group study allowed them to minimize study time outside workshops while still performing better than control students.

Minority Students

Workshop minority students demonstrated a distinct advantage over historic control minority students. Workshop minority participants showed a steep increase in performance by the end of Quarter 1, in contrast to the decrease in performance demonstrated by historic control minority students by the end of Quarter 1. According to predictions based on the relationship of GPA to the performance of majority students on the same exam, workshop minority students performed .64 standard deviations better than what would have been expected on Exam 3. Conversely, historic control minority students performed .24 standard deviations worse than what would have been expected on Exam 3. These positive effects of workshops on minority participants are consistent with earlier work by Treisman (1992) and Steele (1997).

Although workshop minority students reported more hours of study for biology at the beginning of Quarter 1 than did workshop majority students, they reported a decrease in hours of study from Survey 1 to Survey 2, when hours of study reported were virtually identical for all workshop students. At follow-up during Quarters 2 and 3, the performance of workshop minority students was indistinguishable from the performance of their majority counterparts.

During Quarter 1, both majority and minority workshop students gave their best performance on Exam 3. It appears that minority students first realized the benefits of workshop participation on Exam 3, while majority workshop students likely began to realize the benefits on Exam 2. This result suggests that, as with any new learning, the effect of workshop participation is likely to accrue over time. By Exam 3, students who attended all workshop sessions would have spent a total of only 16 h in the workshops. If it is the case that minority students are more likely to find group work a novel experience, then it may be necessary to increase time spent in workshops if benefits to minority participants are to be detected within a 10-week class.

Performance Effects Related to Motivation

Volunteerism

One of the most interesting results of this study was the finding that students who volunteered were motivationally different from students who did not. Volunteers initially reported that they found biology more interesting and more important to their future plans than did nonvolunteers, but they also reported higher levels of anxiety related to class performance. The anxiety related to class performance reported by volunteers was negatively related to performance, whereas anxiety related to class performance reported by nonvolunteers was not related to performance. This finding suggests an important self-selection effect. It may be that volunteers were aware of the detrimental effect of anxiety on their performance, and were motivated to seek ways to compensate for this problem. Indeed, participating in an honors workshop may have been only one of many ways volunteers acted on their concern regarding their performance in biology. Although nonvolunteers experienced a similar range of anxiety related to class performance, they likely had no reason to associate their anxiety with performance. It may also be the case that greater interest in biology and greater importance of biology were factors that encouraged volunteers to seek participation in the workshop program.

The finding that anxiety related to class performance was not detrimental to the class performance (nor was it reliably associated with GPA) of nonvolunteers contradicts prevalent assumptions about the general effect of anxiety on performance within the field of achievement motivation (Revelle, 1986; Sarason, 1961; Wine, 1971). It appears likely that there is a subgroup of people for whom anxiety and avoidance of failure may be particularly detrimental, so detrimental that the effect is detectable in the larger group, and generally assumed to influence everyone similarly. Volunteerism in this study may have been a behavioral measure of how subjectively important it was for students to do well in biology. Raynor (1970) found that decrements in performance were not a simple function of anxiety, but rather that anxiety interacted with importance of the domain for the subject; anxiety was particularly damaging only to students for whom the endeavor was highly important or personally relevant.

Volunteers also showed a pattern of increasing performance across the quarter, relative to nonvolunteers, although there were no differences in overall class performance in any quarter. Doing better later could have been an effect of the combination of high motivation (they kept trying), and a reduction in the effect of anxiety on performance over time. Because overall performance differences were not found between volunteers and nonvolunteers, either within a given quarter or across all three quarters, there is no evidence that nonvolunteers' slightly higher SAT scores played a role in performance of volunteers vs. nonvolunteers.

Workshop Participation

Workshop and control students were not distinguishable on most motivational measures. There was no evidence that workshop participation diminished anxiety related to class performance but measures were only taken once, and this design was not optimal for detecting change that may have occurred gradually. Future investigations should take multiple measures of both trait anxiety and specific anxiety related to class performance and compare the patterns of workshop and control students over time.

Stereotypes

Students' self-reports of how often they were academically stereotyped according to prevalent racial and gender stereotypes were used as an estimate of stereotype threat. Asian American, Hispanic American, and African American students all reported being stereotyped by race and sex more often than did European American students. Although Asian American students report being stereotyped by race, the prevalent stereotype associated with Asians' performance in the sciences is generally positive, rather than negative. The failure of this study to investigate the nature of the stereotype experienced makes it impossible to discuss this unanticipated finding with regard to Asian students. Because of the limited number of Hispanic and African American students in this study, no conclusions could be reached regarding the association between performance and the experience of being stereotyped according to prevalent, negative, racial stereotypes.

Although female students reported being stereotyped by gender more often than male students, this effect was largely due to ethnicity rather than sex. Asian, Hispanic, and African American women were more likely to report being stereotyped by gender than were European American women. Surprisingly, report of being stereotyped by gender was positively related to performance for female volunteers, but unrelated to performance for female nonvolunteers. Future investigation of the relationship between performance and the experience of being stereotyped should likely employ more and more detailed measures. Focus groups might be a good way to clarify some of the issues and experiences related to being stereotyped that Asian, Hispanic, and African American students feel are relevant to their class performance. This method might also shed light on the different experiences of European American female students and their female peers of other ethnicities.

Implications

To the extent possible, programs interested in promoting excellence in gateway courses should incorporate workshop programs similar to the one described here and make them available to students who are interested in making the commitment to participate. This research is consistent with the literature documenting the effectiveness of workshop programs in gateway courses. In this study, workshop participation improved the performance of all participants relative to appropriate controls, as well as to the larger class. The differences noted between volunteers and nonvolunteers also have important implications for education. Volunteers for programs like this one are likely to be vulnerable to the effects of anxiety and underperform in the absence of a challenge intervention, but able to achieve at levels higher than anticipated as a result of workshop participation. This result in particular may make the decision to have control groups in quasi-experimental studies like this one ethically problematic. In addition, it is likely that performance could be further improved by incorporating validated techniques for managing anxiety into the design of future workshop interventions. As an additional incentive, there is evidence that peer-led workshop programs actually save institutions money when compared to doing nothing (Bonsangue and Drew, 1995), and may save tremendously over remediation programs, which tend to be costly and of questionable benefit (Steele, 1997).

Grading is a potential problem for schools that choose to implement challenge interventions. When grading is done according to a strict curve, any increase in the performance of a subset of students will decrease the grades of other students. If workshops are attended by a large proportion of a class, and workshops increase the overall level of performance, but the curve does not change, then participants may find that higher achievement results in little or no change in their grades. Conversely, nonparticipants may earn lower grades for work that would have earned an A in years past. Workshops are typically not intended to increase the competitiveness and difficulty of a course for nonparticipants. Because grades are powerful ways to motivate (or discourage) and reward (or punish) academic habits, this issue may warrant serious contemplation. In competitive academic environments, the belief that workshop participants have an advantage could lead students to volunteer to participate even if they are not committed to cooperative, small group learning. A moderate proportion of uncommitted group members could potentially damage the effectiveness of the small group environment and compromise the entire program. To the extent possible, criterion grading, or an anchored grading curve would be optimal for use in courses with challenge interventions. In principle, if all students perform well, it should be possible for all of them to earn high grades.

Most workshop interventions are based on the assumption that effective study skills and problem-solving strategies will be disseminated among members of small work groups. Although this assumption was not directly tested in this research, it has intuitive appeal as an explanation for the differences observed between workshop and control students on an indirect measure of these skills: exam performance. Volunteer students entered biology with more than enough motivation, yet those assigned to the control group were unable to mobilize this motivation in their favor despite studying more hours. Workshop students, on the other hand, appear to have been able to focus and direct their motivation in a more productive manner, enabling them to maximize performance and minimize study hours.

It may be that effective study and problem-solving skills increased workshop students' perceptions of competence and subjective expectancies for success, and decreased the distracting effects of evaluation apprehension, allowing them to work more productively without becoming overwhelmed by fears of failure. Workshop students likely benefited from learning more effective learning strategies and detected a positive contingency between their efforts to do well and their improved performance as the quarter progressed. Conversely, control students may have been frustrated by the lack of contingency between effort and outcome. This vicious circle may have resulted in increased fear of failure, distracting worry about failure, less efficient study, and increased motivation to avoid failure. There is evidence that among intellectually capable students, high anxiety regarding exam performance is associated with lower self-efficacy and impaired performance (Benjamin *et al.*, 1987), and further evidence that students with higher self-efficacy use more self-regulation and adaptive problem-solving strategies during achievement tasks (Pintrich and de Groot, 1990) but no data on how these variables influence each other dynamically over time following differing achievement outcomes. Future investigations might benefit from the inclusion of periodic self-report ratings of perceived competence, perceived relationship of effort to outcome, and expectancy of success.

Investigation of the motivational mechanisms underlying the achievement gains associated with workshop participation was inconclusive. Future studies of motivational change might be more successful if measures of motivation were taken three or more times during the course of a class. Regardless of questions about why workshop groups have been effective, the results of this study are consistent with the growing body of work that suggests challenge interventions are highly beneficial for both majority and minority students who choose to participate in them.

ACKNOWLEDGMENTS

The authors would to thank Dr. Kenneth R. Bain for his expertise and support of this project.

REFERENCES

Abramson, L. Y., Seligman, M. E. P., and Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology* 87: 49–74.

Atkinson, J. W., and Birch, D. (1970). *The Dynamics of Action*, Wiley, New York.

Atkinson, J. W., and Birch, D. (1974). The dynamics of achievement-oriented activity. In Atkinson, J. W., and Raynor, J. O. (Eds.), *Motivation and Achievement*, Silver Springs, MD, V. H. Winston, pp. 271–236.

Atkinson, J. W., and Birch, D. (1986). Fundamentals of the dynamics of action. In Kuhl, J., and Atkinson, J. W. (Eds.), *Motivation, Thought, and Action*, Praeger, New York, pp. 16–48.

Benjamin, M., McKeachie, W. J., and Lin, Y. (1987). Two types of test-anxious students: Support for an information processing model. *Journal of Educational Psychology* 79: 131–136.

Bonsangue, M. V., and Drew, D. E. (1995). Increasing minority students' success in calculus. *New Directions for Teaching and Learning: Fostering Student Success in Quantitative Gateway Courses* 61: 23–33.

Born, W. K. (2000). *The Effect of Workshop Groups on Achievement Goals and Performance in Biology: An Outcome Evaluation*, Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

Cohen, W. (1997). *BioExcel at the University of Kentucky*. Unpublished Project Description.

Covington, M. V., and Omelich, C. L. (1991). Need achievement revisited: Verification of Atkinson's original 2×2 model. In Spielberger, C. D., Sarason, I. G., Kulcsar, Z., and VanHeck, G. L. (Eds.), *Stress and Emotion: Anxiety, Anger, and Curiosity*, Hemisphere, New York, Vol. 14, pp. 85–105.

Doyle, R. (2002). Filling the Pipeline: Are there enough Ph.D.s in science and engineering. *Scientific American* 286: 25.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist* 41: 1040–1048.

Dweck, C. S., and Leggett, E. L. (1988). A social–cognitive approach to motivation and personality. *Psychological Review* 95: 256–273.

Elliot, A. J., and Church, M. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology* 72: 218–232.

Elliot, A. J., and McGregor, H. A. (1999). Test anxiety and the hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology* 76: 628–644.

Freeman, M. (1995). *MathExcel: A Special Opportunity in Calculus*. Unpublished Report, Department of Mathematics, University of Kentucky.

Freeman, M. (1997). Collaborative math and science workshops in Kentucky and Appalachia. *Mathematics and Education Reform Forum* 9: 1, 5, 8.

Fullilove, M., Fullilove, R. E., Terris, M., and Lacayo, N. (1988). Is "Black achievement" an oxymoron? *Thought and Action* 4: 5–20.

Fullilove, R. E., and Treisman, P. U. (1990). Mathematics achievement among African American undergraduates at the University of California, Berkeley: An evaluation of the mathematics workshop program. *Journal of Negro Education* 59: 463–478.

Gosser, D., Roth, V., Gafney, L., Kampmeier, J., Strozak, V., Varma-Nelson, P., Radel, S., and Weiner, M. (1996). Workshop chemistry: Overcoming the barriers to student success. The *Chemical Educator* 1: 2–3.

Gray, J. A. (1987) *Psychology of Fear and Stress*, 2nd ed., Cambridge University Press, New York.

Hanson, S. L. (1996). *Lost Talent: Women in the Sciences*, Temple University Press, Philadelphia.

Hill, S. T. (1996). *Science and Engineering Degrees, by Race/Ethnicity of Recipients, 1987–94* (Detailed Statistical Tables, pp. 96–329), National Science Foundation, Washington, DC.

Humphreys, M. S., and Revelle, W. (1984). Personality, motivation, and performance: A theory of the relationship between individual differences and information processing. *Psychological Review* 91: 153–184.

Johnson, D. W., Johnson, R. T., and Smith, K. A. (1991). *Cooperative Learning: Increasing College Faculty Instructional Productivity*, School of Education and Human Development, George Washington University, Washington, DC.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D., and Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin* 89: 47–62.

Kuhl, J., and Blankenship, V. (1979). The dynamic theory of achievement motivation: From episodic to dynamic thinking. *Psychological Review* 86: 141–151.

Lewin, K., Dembo, T., Festinger, L., and Sears, P. S. (1944). Level of aspiration. In McHunt, J. (Ed.), *Personality and the Behavior Disorders*, Ronald Press, New York, Vol. 1, pp. 333–378.

Mandler, G., and Sarason, S. B. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology* 47: 166–173.

Mathews, A. M., and MacLeod, C. (2002). Induced processing biases have causal effects on anxiety. *Cognition and Emotion* 16: 331–354.

McCaffrey, J., and Myers, M. (1994). *The Emerging Scholars Program*, Unpublished program description, The University of Texas.

McClelland, D. C., Atkinson, J. W., Clark, R. A., and Lowell, E. L. (1953). *The Achievement Motive*, Appleton-Century-Crofts, New York.

Miller, L. S. (1995). *An American Imperative: Accelerating Minority Educational Advancement*, Yale University Press, New Haven.

National Science Board (2002). *Science and Engineering Indicators*, U.S. Government Printing Office, Washington, DC.

National Science Board Task Committee on Undergraduate Science and Engineering Education (1987). *Undergraduate Science, Mathematics and Engineering Education* (Source Materials NSB), U.S. Government Printing Office, Washington, DC, Vol. II, pp. 86–100.

Pintrich, P. R., and de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology* 82: 33–40.

Raynor, J. O. (1969). Future orientation and motivation of immediate activity: An elaboration of the theory of achievement motivation. *Psychological Review* 76: 606–610.

Raynor, J. O. (1970). Relationships between achievement related motives, future orientation, and academic performance. *Journal of Personality and Social Psychology* 15: 28–33.

Revelle, W. (1986). Motivation and efficiency of cognitive performance. In Brown D. R., and Veroff, J. (Eds.), *Frontiers of Motivational Psychology: Essays in Honor of John W. Atkinson*, Springer, Berlin, pp. 107–131.

Revelle, W., and Michaels, E. J. (1976). The theory of achievement motivation revisited: The implications of inertial tendencies. *Psychological Review* 83: 394–404.

Sarason, I. G. (1961). The effects of anxiety and threat on solution of a difficult task. *Journal of Abnormal and Social Psychology* 62: 165–168.

Seligman, M. E. P. (1975). *Helplessness: On Depression, Development, and Death*, Freeman, San Francisco.

Selvin, P. (1992). Math education: Multiplying the meager numbers. *Science* 258: 1200–1201.

Slavin, R. E. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin* 94: 429–445.

Slavin, R. E. (1995). *Cooperative Learning: Theory, Research, and Practice*, 2nd ed., Allyn & Bacon, Needham Heights, MA.

Steele, C. M. (1992). Race and the schooling of Black Americans. *The Atlantic Monthly* 269: 68–78.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist* 52: 613–629.

Steele, C. M., and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology* 69: 797–811.

Treisman, U. (1992). Studying students studying calculus: A look at the lives of minority mathematics students in college. *The College Mathematics Journal* 23: 362–372.

U.S. Congress, Office of Technology Assessment (1988). *Educating Scientists and Engineers: Grade School to Grad School* (OTA-SET-377), U.S. Government Printing Office, Washington, DC.

Vetter, B. M. (1990). *Who is in the Pipeline? Science, Math and Engineering Education* (Occasional Paper 90-2), Commission on Professionals in Science and Technology, Washington, DC.

Watson, D., and Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin* 96: 465–490.

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin* 76: 92–104.

Wortman, C. B., and Brehm, J. W. (1975). Responses to uncontrollable outcomes: An integration of reactance theory and the learned helplessness model. In Berkowitz, L. (Ed.), *Advances in Experimental Social Psychology*, Academic Press, New York, Vol. 8, pp. 277–336.