Running head: THINKING AND ACTING LIKE A SCIENTIST

Crashing the Gate: Identifying Alternative Measures of Student Learning in Introductory Science, Technology, Engineering, and Mathematics Courses

Kevin Eagan, Felisha Herrera, Jessica Sharkness, Sylvia Hurtado, and Mitchell Chang

University of California, Los Angeles

This study was made possible by the support of the National Institute of General Medical Sciences, NIH Grant Numbers 1 R01 GMO71968-01 and R01 GMO71968-05 as well as the National Science Foundation, NSF Grant Number 0757076. This independent research and the views expressed here do not indicate endorsement by the sponsors.

Introduction

Without significant improvement in scientific training in U.S. postsecondary institutions, America stands to lose its competitive edge in scientific achievement, innovation, and economic development. The National Academies' report, *Rising Above the Gathering Storm* (2007), revealed that only 15% of all undergraduates in the U.S. receive their degrees in natural science or engineering, compared with 67% in Singapore, 50% in China, 47% in France, and 38% in South Korea. To continue the United States' level of achievement and innovation in science and engineering, we must not only improve production of undergraduate science majors but also increase the number of diverse people in the scientific workforce. Indeed, increasing the diversity of talent in science, technology, engineering, and mathematics (STEM) represents a critical part of maintaining and increasing national prominence in science and engineering. As the Council of Graduate Schools (2007) wrote in a report on graduate education and American competitiveness, "it is imperative that the U.S. citizens from all population groups, including those who traditionally have not been highly represented, such as minorities and women, pursue STEM," for only in this way can the United States "maintain its competitive edge" (p. 15).

Central to improving the production of new scientists is the identification of successful students who express and maintain an interest in science. Recently, encouraging signs indicate that the U.S. has made progress in diversifying the STEM workforce and educational pipeline. For example, trend data show increases among entering freshmen in terms of interest in science and engineering majors as well as in the number of years of high school coursework completed in biology, math, chemistry and physics (Pryor, Hurtado, DeAngelo, Sharkness, Romero, Korn & Tran, 2008). Further, the National Science Board (2008) recently reported increases in the enrollment of African American, Latino, and Native American students in STEM undergraduate and graduate programs. However, the odds of remaining in science until degree completion are still currently very low: only 24% of underrepresented racial minority (URM) students and 40% of White students who begin college as science majors complete bachelor's degrees in science within six years of college entry (Center for Institutional Data Exchange and Analysis, 2000). Although many individual and institutional factors account for student attrition (Espinosa, 2009; Seymour & Hewitt, 1997), we wish to sharpen the focus on introductory coursework in science and mathematics. Students with aspirations for STEM bachelor's degrees encounter significant obstacles in the form of "gatekeeper" courses almost as soon as they begin their collegiate coursework. Success in introductory science and math coursework represents the necessary first step toward the completion of a bachelor's degree in STEM, as introductory courses provide foundational learning for all further coursework. Unfortunately, rather than providing tools for future study, many introductory courses tend to discourage many students from continuing in the sciences (Seymour & Hewitt, 1997).

Instructors of introductory courses often grade on a curve (allocating very few A grades) and narrowly assess student performance when assigning the grades. We question whether course grades actually capture the full set of scientific skills that students acquire throughout their introductory coursework. Indeed, instructors typically base grades in introductory science courses on students' ability to acquire and retain specific content knowledge rather than on their development of critical thinking skills, the latter of which are equally as necessary as the former for future science careers (Gainen, 1995).

Assessing students based primarily on mastery of content knowledge makes gatekeeper courses resemble sorting mechanisms that harvest student talent rather than develop it. Several scholars have concluded that prior academic achievement largely determines students' grades in introductory coursework (Gainen, 1995; Kamii, 1990; NCES, 2000; Payzant & Wolf, 1993; Waits & Demana, 1988); therefore, gatekeeper course grades may be more of a reflection of students' prior abilities than the acquisition of knowledge and skills during the course. Particularly in the case of underrepresented racial minority students (as well as women), prior achievement, motivation, and socioeconomic status represent key predictors of success in college science and engineering courses (NCES, 2000; Payzant & Wolf, 1993; Waits & Demana, 1988).

It is important to untangle prior preparation and background factors from performance assessment in courses and also to identify whether students in introductory courses develop the higher-order cognitive skills that they need future success in STEM-related careers. Drawing from research on science pedagogy and students' habits of mind for scientific work, we explore how student experiences in introductory courses affect academic achievement and the development of scientific skill sets. The purpose of this study is to assess the factors that predict the acquisition of the ability to think and act like a scientist as well as to assess the relationship between these scientific dispositions and academic achievement (grades) in introductory courses. Our goal is to provide research that will improve practice in introductory courses, lead to broader ways of identifying talent in science beyond academic achievement, affirm development of student skills in assessing performance, and ultimately increase the pool of students who will continue toward undergraduate degrees in STEM fields.

Literature Review

"Gatekeeper" Courses

Scholars and practitioners generally refer to introductory courses as "gatekeepers" because they represent the first course in a series of classes where knowledge becomes cumulative (Tobias, 1990). Success in introductory courses is theoretically a function of both scientific thinking dispositions and content knowledge (Conley, 2005; Hagedorn, Siadat, Fogel, Nora, & Pascarella, 1999). Unfortunately, in practice, college-level introductory science and mathematics courses tend to focus too much on the acquisition of content knowledge and too little on the development of meta-cognitive skills related to critical thinking and scientific literacy (Handelsman, Ebert-May, Belchner, Bruns, Chang, DeHaan, Gentile, Lauffer, Stewart, Tilghman, & Wood, 2004; Hurd, 1997; Williams, Papierno, Makel, & Ceci, 2004). As a result, introductory science courses tend to have relatively high failure and dropout rates (Seymour & Hewitt, 1997). Failure to succeed in gatekeeper courses can lead to difficulty in future courses and may prompt students to switch out of science majors (Labov, 2004; Seymour, 2001).

The high failure and attrition rates observed in gatekeeper courses have been attributed to many factors, including large class sizes, a lack of engaging pedagogy, and high competition among students (Handelsman et al., 2004; Seymour & Hewitt, 1997; Tobias, 1992). Because of the volume of students who enroll in introductory science and math courses, class sizes tend to be large, and instructors tend to rely on lecture to transmit course content. Although lecture may represent an efficient method for presenting domain-specific information to a large audience, it tends to encourage passive learning (Seymour & Hewitt, 1997).

In introductory science courses, lectures may fail to engage students in critical thought about the content being presented and may not provide the stimulation necessary to intellectually engage students in learning science and math (Gainen, 1995). Handelsman et al. (2004) note that "active participation in lectures…helps students develop the habits of mind that drive science. However, most introductory courses rely on 'transmission-of-information' lectures" that do not encourage participation (p. 521). Introductory courses that do not engage students can cause students to feel bored and disconnected from science; these feelings in turn may lead intended science majors, high achieving or otherwise, to switch out of the sciences (Seymour & Hewitt, 1997).

In addition to feelings of disengagement in introductory courses, students also often experience high levels of competition among their peers (Gainen, 1995; Mazur, 1992; Seymour & Hewitt, 1997; Tobias, 1992). Much of the competitive culture within introductory science and math courses originates from the grading practices of faculty teaching these courses, as they tend to grade students on a curve (Seymour & Hewitt, 1997). Grading on a curve discourages collaboration and cooperation among students because helping one's peers in class can disadvantage an individual student in terms of final (e.g., post-curve) test scores (Mazur, 1992). Indeed, grading on a curve engenders a "survival of the fittest" mentality among students in introductory courses (Epstein, 2006). This competitive mentality may be strongest among science students. For example, Vahala and Winston (1994) found that students taking laboratory science classes reported experiencing a more hostile classroom environment than did their peers in English courses.

Classroom Environments and Instructor Pedagogies

A competitive culture represents one of several climates that students may encounter in their collegiate courses. Classroom environments can have a significant effect—positive or negative—on student achievement (Fraser & Fisher, 1982; Prenzel, Kramer, & Dreschel, 2002; Seidel, 2006). Scholars consistently have demonstrated that competitive classroom climates have negative effects on students' learning and performance. For example, Walberg (1979) examined the effect of classroom climate on student achievement and retention and found that students who experienced more competitive classroom contexts tended to have higher rates of failure and lower levels of self-confidence than did their peers in more cooperative environments.

In a later study, Fraser and Fisher (1982) analyzed data collected from 1,083 science students across 116 classrooms and found positive correlations between students' adoption of scientific attitudes and their perception that the course encouraged participation and cooperation. In general, it seems that science students who perceive that courses have a sense of cohesion and goal orientation tend to have higher levels of academic achievement compared to their peers who had negative perceptions of their classroom context (Haertel, Walberg, & Haertel, 1981). These studies imply that if gatekeeper courses have competitive learning climates that reduce collaboration by forcing students to contend with one another for grades, most or all students may be negatively affected.

Because of the potential implications associated with how students perceive their learning environment, both in terms of engagement and the learning climate, many scholars have examined how pedagogical strategies can be adjusted to encourage student engagement and collaboration in class (Armstrong, Chang, & Brickman, 2007; Knight & Wood, 2005). For example, to examine how the extent of lecture and group work affected student learning, Knight and Wood (2005) conducted an experiment with two offerings of an upper-division biology course. In the first term, the course instructor used traditional lecture methods to teach the course. In the second term, the instructor employed not only lecture but also a number of cooperative and problem-solving activities. Knight and Wood (2005) compared students' scores on pre- and post-tests from the beginning and end of each term, respectively, and found that students who had exposure to the cooperative classroom environment experienced greater knowledge gains compared to their peers in the more traditional, lecture-heavy classroom environment. The authors repeated this experiment in a subsequent term and obtained similar results.

A similar study by Armstrong, Chang, and Brickman (2007) examined how students exposed to cooperative learning environments compared to their peers taking courses with a traditional lecture format. The experiment included two instructors, each of whom taught two sections of biology: one section utilized a traditional lecture format while the other section included more cooperative learning and group activities. Controlling for pre-test scores, students in the cooperative learning groups showed significantly more improvement in understanding than did their peers in the lecture condition. The cooperative group also had higher attendance rates and expressed more enjoyment and connection with the work.

A large number of studies over the last 25 years have concluded that incorporating small-group and collaborative learning activities in large STEM courses provides students with a number of benefits. Indeed, a meta-analysis by Springer, Stanne, and Donovan (1999) demonstrated that students who enrolled in STEM classes with a greater emphasis on small-group and collaborative learning had higher levels of academic achievement, higher rates of persistence in STEM courses, and more positive attitudes about academics in general than did their peers who had taken courses more dominated by lecture. The specific type of collaborative learning does not seem to matter; students have benefited from a wide range of interactive and active learning activities, including interactive-engagement strategies (Hake, 1998); student-faculty interactions (Astin, 1993; Smith, Sheppard, Johnson, & Johnson, 2005); course discussion (Allen & Tanner, 2002, 2005; Tanner & Allen, 2005), student feedback or clicker systems (Wood, 2004), and problem-based learning exercises (Shipman & Duch, 2001).

Pedagogical strategies that encourage engagement in discussion and group activities not only promote achievement and persistence but also provide students with opportunities to think critically about scientific concepts and their applications and thus promote higher-order thinking skills (Allen, Duch, & Groh, 1996; Freedman, 1994; Sagan, 1996). As Smith, Sheppard, Johnson, and Johnson (2005) describe, collaborative learning exercises offer an opportunity for the sharing of multiple perspectives on a problem, which allows students to evaluate several sources of evidence and rationales before deciding on a path toward problem resolution. Encouraging student participation in class discussions has been linked to higher-order critical thinking skills by many researchers (cf. Tsui, 2002), as has using students' ideas in class, providing opportunities for student interactions in class, and encouraging more classroom involvement in general (Smith, 1977, 1981; Terenzini, Theophilides, & Lorang, 1984). Scholars also have linked growth in critical thinking skills to other activities that encourage active and open-ended learning, including participation in research projects (Astin, 1993; Tsui, 1999), academic enrichment programs (Summers & Hrabowski, 2006), and taking essay exams (Astin, 1993).

Supportive Learning Environments and the Skills Needed for Scientific Success

As noted by Walberg (1979), Fraser and Fischer (1982), Vahala and Winston (1994), and others, students' perceptions of the classroom environment can have a direct effect on their academic achievement and cognitive development. Hostile or competitive environments tend to have negative effects on students' achievement whereas positive, supportive, and collaborative classroom contexts enhance student learning. Prenzel, Kramer, and Dreschel (2002) identified six necessary conditions for a supportive learning environment: relevance of content; quality of instruction; teacher's interest; social relatedness; support of competence; and support of autonomy. Environments that provide relevant content offer students the opportunity to see the real-world application of concepts and content knowledge. Quality instructors transmit information to students in clear, coherent ways that meet the learning needs of the particular audience in the classroom. Interested instructors demonstrate their commitment to student learning and an ethic of care in regard to students' academic problems. Classrooms that contain elements of social relatedness emphasize attributes of collegiality and cooperation. Instructors that offer students support of competence provide individualized, constructive feedback on coursework. Finally, learning environments containing support of autonomy offer students opportunities for exploration in problem solving, which allows for multiple approaches for reaching resolutions and conclusions.

Seidel (2006) writes that classrooms with these six elements engender greater self-motivation among students while allowing them to become more self-directed in their learning. Students who feel supported by their professors and have opportunities to engage in constructive ways with their peers increase the frequency with which they take risks in the classroom and offer alternative approaches to solving problems (Cobb, 1994; Driver, Asoko, Leach, Mortimer, & Scott, 1994). Additionally, participating in an environment that encourages the application of abstract scientific concepts to real-world experiments and problems provides students with the opportunity to understand the relevance and importance of the material presented in class. Exercises that allow for such application enable students to transition from abstract knowledge to more concrete understanding.

Specifically connected to the transition from the abstract to the concrete is students' ability to learn to think and act like scientists. The ability to think like a scientist involves asking questions, identifying problems, evaluating evidence to make appropriate judgments, and finding ways to make scientific findings relevant and accessible to society at large (Williams, Papierno, Makel, & Ceci, 2004). Koslowski (1996) describes two approaches related to how students learn to act and think like scientists: domain-specific and domain-general. Instructors who take a domain-specific approach to teaching science rely heavily on transmitting specific scientific concepts to students, having students memorize textbook definitions, and teaching algorithmic methods for problem solving. The domain-specific approach assumes that students will develop higher-order, critical thinking skills as they apply the scientific concepts and knowledge taught in the classroom (Barnett & Ceci, 2002).

In contrast to the domain-specific approach to teaching science, instructors who utilize domain-general pedagogical techniques tend to focus more on the development of reasoning and critical thinking skills that can be applied to a variety of disciplines (Koslowski, 1996; Resnick, 1987). The development of these higher-order thinking abilities provide students with the requisite skills to more accurately evaluate evidence and choose multiple paths in problem solving (Kuhn, Garcia-Mila, Zohar, & Anderson, 1995). Rather than merely memorizing a set of causal relationships, students who have developed appropriate levels of critical thinking skills can infer causal relationships from evidence presented, and this ability to reason and think critically about evidence represents a key difference between the domain-specific and domain-general approaches to teaching.

Additional influences on student success in introductory STEM courses

Course content, pedagogical practices, and learning environments all significantly affect students' achievement and development of scientific and general critical thinking skills in college classrooms. However, scholars have identified a number of experiences external to the classroom environment that also significantly contribute to growth in students' critical thinking abilities. For example, a substantial body of research has examined the benefits of research participation for undergraduate students, showing that engagement in research projects can help students tremendously in terms of their ability to develop scientific literacy and critical thinking skills because these opportunities provide students with hands-on learning experiences in which they must apply concepts learned in the classroom to real-world scenarios (Barlow & Villarejo, 2004; Foertsch, Alexander, & Penberthy, 1997; Lopatto, 2004; Sabitini, 1997). For example, Sabitini (1997) tracked the progress of several undergraduate students who conducted research with him, and found that these students reported substantial gains in their problem-solving abilities. The students also appreciated the opportunity to engage in a team setting.

Another activity that has been connected to growth in achievement and critical thinking skills is participation in peer tutoring (Bulte, Betts, Garner, & Durning, 2007; Glynn, MacFarlane, Kelly, Cantillon, & Murphy, 2006; Lockspeiser, O'Sullivan, Teherani, & Muller , 2008). Glynn et al. (2006) conducted a qualitative study in which fifth-year medical students served as peer tutors for groups of second-year medical students for a specific educational module related to the Early Patient Contact program. The second-year students reported that the peer tutors found ways to make the information contained in the module relevant to them, which encouraged a deeper understanding as well increased knowledge retention. Glynn et al. (2006) concluded that peer tutoring environments provide students with a safe learning environment, which encourages students to ask questions and discuss content more freely than they would in the context of a classroom. In a similar study, Lockspeiser et al. (2008) found that medical students who took advantage of peer-tutoring programs reported that they gained more support from their tutors than from the classroom context, and this helped to put them at ease about their ability to learn all of the information required of first- and second-year medical school students.

Overall, it seems that receiving tutoring or serving as a tutor can help students make learning connections they might otherwise not have made, especially if their only other sources of course content have been textbooks and lecture.

Conceptual Framework

Building on previous research this study takes a broad view of success that encompasses not only the narrow construct of students' course grade but also students' ability to think and act like scientists. We hypothesize that student success will be affected by a variety of factors that can broadly be broken down into five conceptual areas: students' experience of the learning environment in their introductory courses; course pedagogy; students' out-of-class experiences with research and tutoring; the amount of effort students expend on the course; and students' critical thinking dispositions. We propose that these factors will significantly affect student success after taking into account students' demographic characteristics, high school achievement, predispositions toward scientific thinking, and pre-college experiences.

We conceive of the student experience with the course learning environment as being multifaceted, encompassing how students feel about and interpret their instructors' teaching methods, coursework, and learning environment. We conceptualize course pedagogy as the kind of instructional methods that are used, the emphasis of the coursework, and how often feedback about course progress/performance is given. Student effort, background characteristics, and initial propensity toward thinking and acting like a scientist are included in the model because these factors have all been shown to affect student success in coursework; controlling for these variables allows for the examination of the unique impact of introductory courses and other related experiences on student success.

In terms of the outcomes in our model, we assess student "success" using three independent measures: thinking like a scientist, acting like a scientist, and students' final grade in the introductory course. Our intention was to separate students' thinking preferences and performance as a scientist from the evaluation students received in the course. We hypothesize that post-test scores of thinking and acting like scientists are significantly predictive of students' final course grade. We hypothesize the presence of these causal paths for two reasons. First, students reported their self-ratings on items related to thinking and acting like scientists prior to the posting of their course grade; thus, the temporal order in which we collected these measures dictated that students' end-of-term dispositions toward science may affect their final course grade. Second, we wanted to assess whether course grades were associated with the skills necessary to be a successful scientist or whether grades are independent of these skills.

Method

Data and Sample

During the spring of 2010, we surveyed students and faculty in introductory STEM courses at 15 institutions across the U.S. The sample of institutions was relatively diverse and included three historically Black colleges and universities (HBCUs), three Hispanic-serving institutions (HSIs), eight public institutions, one technical university, and two liberal arts institutions. Within each institution, we sampled between five and six introductory STEM courses. We defined introductory as the first course in a sequence of courses where knowledge is cumulative. In other words, success in any one of the courses in this study was necessary to move onto the next course in the sequence. Each institution provided at least one biology course and at least one chemistry course. We also had a mix of introductory calculus, statistics, engineering, and computer science courses across all of our participating institutions; however, the distribution of the types of courses represented varied across each institution.

At the beginning of the academic term, students in these courses completed the web-based 2010 STEM Student Pre-Questionnaire, which collected information on students' self-rated academic and science abilities, the frequency with which they articulate and apply science concepts, and demographic information. Many of the items on the pre-survey were based on Conley's (2005) work in identifying the skills and dispositions students should have for success in introductory science courses. Participation in the study was voluntary, and students who completed the survey received a $10 gift card. During the last two weeks of the academic term, students completed the web-based 2010 STEM Student Post-Questionnaire, which re-asked many of the same questions from the first survey while also including a number of items related to students' experiences in their introductory courses. Students who completed the post-survey received a $10 gift card. Faculty who taught these courses completed an online instructor survey at the end of the course, which included items related to the pedagogical techniques faculty used in the course, their perceptions of student learning, and their priorities for undergraduate education. Finally, we collected course grade information from registrar's offices at each of our institutions. In all, we had 3,205 students across 88 classrooms in 15 institutions respond to both student surveys, which translated into a 42.1% response rate.

Measures

This study focuses primarily on three outcome variables, measured at or near the end of the academic term. Two outcomes are latent constructs that represent the frequency with which students reported acting and thinking like scientists, and these same two constructs were measured as pre-tests in the first survey in order to control for prior abilities and experiences.

The third outcome measure is a variable representing students' final grade in their introductory course, which we collected from registrar's offices at each institution. The two latent constructs, frequency of thinking like a scientist and acting like a scientist, were composed of a set of indicator variables identified through factor analysis. A confirmatory factor analytic model was run in EQS 6.1 to confirm that the relationships between the indicator variables and the constructs held up for both the pre-test and the post-test. Table 1 presents the indicator variables explained by the latent constructs, their factor loadings, and the fit indices for the measurement model.

Analyses

Through structural equation modeling (SEM), we analyzed relationships among exogenous and endogenous variables in an effort to simultaneously estimate the relationships among sets of variables and confirm latent constructs (Bentler, 2006; Bentler & Wu, 2002). Parameter estimates are generated by analyses of estimated covariance matrices. SEM accounts for measurement error and provides overall goodness of fit indices to determine the adequacy of the model, both of which represent advantages over traditional path analysis (Laird, Engberg, & Hurtado, 2005). To assess model fit, we relied on three fit indices: non-normed fit index (NNFI), comparative fit index (CFI), and root mean square error of approximation (RMSEA). NNFI and CFI values above 0.90 indicate adequate model fit, while RMSEA scores below 0.06 indicate an appropriate level of fit (Raykov, Tomer, & Nesselroade, 1991).

Our analytic approach began with a confirmatory factor analysis that tested the adequacy of our measurement model. As mentioned above, the measurement model included the observed indicator variables and their associated latent constructs for both the pre- and post-surveys. This measurement model confirmed the factor structure of the two pre-test factors and their associated post-tests. Next, we added to the measurement model all of the hypothesized predictors and paths to test the full structural model. LaGrange Multiplier tests provided guidance, in conjunction with prior literature and theory, about adding relational paths among variables in the model. If all paths from a variable were removed, we dropped the variable from the analysis. As the final model, presented in Figure 1, illustrates, we removed several paths from measures of course pedagogies to the three outcome variables. Similarly, we deleted several of the paths that connected measures of the course environment with the three outcome variables.

Limitations

A number of constraints with the methodology and the data may limit the generalizability of our findings and conclusions. First, we situated our study within 88 classrooms across 15 campuses of varying type, size, selectivity, and mission, and the composition of our sample may limit the generalizability of our findings to other types of institutions and classroom contexts. Second, these data have a nested design in which students are clustered within classrooms that are clustered within institutions. Because of sample size considerations within each classroom and within each institution, we could not disaggregate data by classrooms or by institutions. Finally, because this study examined students' experiences in a single introductory course, the short timeframe in which we administered the surveys may have affected the amount of change detected in students' frequency of thinking and acting like a scientist. This study chose to focus on student changes in these items over the span of a single academic term; had students been tracked over a longer period of time, we may have detected more substantial changes in these outcomes.

Results

Table 1 presents the results of the measurement model for the four latent variables in the study: pre- and post-tests of thinking like a scientist and acting like a scientist. The model statistics suggest that the model fits the data well. The Satorra-Bentler chi-square statistic is 300.69, and the fit indices are NNFI = 0.98, CFI = 0.98, and RMSEA = 0.03. Table 1 shows the factor loadings of observed variables on each of the four factors; all variables loaded highly on the relevant factor.

Table 2 shows the results of the final structural model, including unstandardized regression coefficients, standardized regression coefficients, standard errors, and significance levels for the model's direct effects. Figure 1 diagrams the causal paths in the final structural model, using solid lines to represent significant effects and dotted lines to represent non-significant paths. The Satorra-Bentler chi-square statistic for the final model was 2281.07 (df = 808, N = 3205, $p < 0.001$), and the fit indices were: NNFI = 0.93, CFI = 0.94, and RMSEA = 0.03. Although the chi-square statistic was significant, it is highly dependent on sample size and degrees of freedom and thus cannot be relied upon as an indicator of model fit (Bentler, 2006). The other indices suggest that the model appropriately represents the relationships among students' characteristics at the beginning of the course, their experiences and exposure to pedagogical strategies within the course, and the three primary outcomes.

Considering the total effects on the three outcome variables, the results in Table 2 indicate that composite SAT score was the most important predictor of students' end-of-course grade. Higher SAT scores corresponded to significantly higher grades in introductory STEM courses. Likewise, students who reported higher class rank also tended to have significantly higher grades than their peers who ranked lower among their graduating high school classes. These findings connect to prior research that demonstrates that students with higher SAT scores and greater levels of academic achievement in high school typically earn significantly higher grades early in college compared to their peers with lower test scores and high school grades (Astin, 1993).

Another positive predictor of end-of-course grade was students' indication that they had participated in a pre-college research program. By contrast, students who indicated aspirations for a research career tended to have significantly lower course grades compared to their peers who did not report such aspirations. This variable may have served as a proxy for students with medical school aspirations; in other words, students with medical school aspirations may not have been as likely to aspire to a research career.

Among the variables corresponding to students' experiences in the course, findings in Table 2 suggest that students who crammed for exams in the course and who felt bored in class more often tended to earn significantly lower grades in the course compared to their peers who crammed for exams less often and did not feel as bored. Students who crammed for exams may have not spent as much time in weeks prior to the exam internalizing and fully learning the material, and this finding affirms prior work that suggests student learning and performance can be hindered by trying to study and learn significant amounts of material at the last minute (Beck, Koons, & Milgram, 2000; Wesley, 1994). Additionally, students who felt bored in class may have disengaged with course material more often, and this disengagement contributed to their significantly lower grades in the course (Struthers, Perry, & Menec, 2000).

Two predictors linked to the pedagogical strategy utilized by faculty teaching introductory courses significantly predicted students' end-of-course grades. Taking a course where faculty members more frequently relied upon multiple-choice exam questions significantly predicted higher course grades. Although more frequent use of multiple choice exam questions by faculty may lead to higher grades, these questions may not reflect whether a student has actually learned the material, as students do not have to demonstrate critical thought to get the correct answer. Instead, memorization of the material, which may be forgotten later, or luck may actually be contributing to the higher grades. Additionally, enrolling in introductory STEM courses where the faculty member spent a greater proportion of class time allow students to work in groups corresponded to significantly higher end-of-course grades. Research consistently shows that students learn well from one another (CITE); thus, allowing students to engage with each other on course material provides them with more opportunities to critically reflect on and internalize course content.

In predicting students' end-of-course grade, we initially tested a hypothesized relationship between students' frequency of thinking and acting like a scientist and their final grade. Course grades were not significantly related to either of these constructs; however, the findings in Table 2 provide details on the characteristics and behaviors that contribute to students' increased frequency of thinking and acting like scientists. As expected, students' frequency of thinking like a scientist at the beginning of the academic term served as the strongest predictor of this construct at the end of the term. The second most important predictor was students' self-rated scientific ability, as students who felt more confident in their scientific ability tended to report that they more frequently thought like a scientist. Unlike the results for GPA, neither SAT scores or class rank had a relationship with students' frequency of thinking like a scientist. Participation in a pre-college research program, aspiring to a research career, and identifying as White, as compared to non-White, significantly predicted students thinking like a scientist significantly more often. Perhaps pre-college research careers primed students' interest in STEM, which contributed to a greater engagement with STEM in introductory STEM courses.

Likewise, identifying with STEM early, perhaps with an early aspiration for a research career, increases students’ engagement with science (Carlone & Johnson, 2007).

Among the course-related predictors, three variables had a significant relationship with students’ frequency of thinking like a scientist. Students who reported having joined or created a study group had significantly higher frequencies of thinking like scientists. Similarly, taking a course where faculty used a greater percentage of class time for group work also corresponded to higher frequencies of thinking like a scientist. Similar to the finding with grades, students who engage with other students on course material are able to better learn from their peers. The third course-related variable related to thinking like a scientist was the extent to which students reported feeling bored in class, which negatively predicted their frequency of thinking like a scientist. To the extent that feeling bored in class leads to greater disengagement (CITE), it is not surprising that students who felt bored more often tended to think about science less often.

Linked to thinking like a scientist was the frequency with which students reported acting like scientists. Similar to thinking like a scientist, the pre-test for acting like a scientist represented the most important predictor for this outcome at the end of the course. Self-rated scientific ability was the second most important predictor of acting like a scientist. Neither SAT scores or class rank significantly predicted students’ frequency of acting like a scientist by the end of their introductory STEM course. Students who aspired to a research career and who participated in a pre-college research program reported acting like a scientist significantly more often. Self-rated scientific ability, research career aspirations, pre-college research participation, and identifying as White significantly and indirectly, through the pre-test, predicted higher frequencies of acting like a scientist.

Direct effects on acting like a scientist included joining or creating a study group, taking an introductory course where the faculty member felt strongly about preparing undergraduate students for graduate education, and the frequency with which faculty used electronic quizzes for immediate feedback. As with thinking like a scientist, students who joined or created a study group reported thinking like a scientist significantly more often than their peers who did not participate in or create a study group. Additionally, students whose introductory course faculty felt more strongly about preparing undergraduate students for graduate education tended to report thinking like a scientist significantly more often. By contrast, having a faculty member who utilized electronic quizzes for immediate feedback more often tended to significantly decrease how often students reported acting like a scientist. Although these quizzes offered students immediate feedback about their understanding of the material, electronic quizzes may have reduced the time students had to devote to behaviors indicative of acting like a scientist, such as conducting an experiment or synthesizing several sources of information.

Discussion and Conclusion

One of the primary goals of higher education is to educate cadres of graduates whose talents can improve society and global economic competitiveness. The National Science Foundation's National Science Board notes that science and engineering serve as the primary drivers of both economic growth and national security, and that "excellence in discovery and innovation in science and engineering…derive from an ample and well-educated workforce" (National Science Board, 2003, p. 7). Improvement in science teaching and classroom practice is now more critical than ever to maintaining America's competitive edge. Promoting success in college coursework is necessary for creating a large and diverse group of students who have the potential to become the next generation of scientists.

Although students' grades in introductory courses may be useful for sorting students, they do not seem to be useful for capturing gains in dispositions for scientific work. Instead, students' course grades were related to cramming for exams, working in groups, and feeling bored in class. Further, as we expected, students' course grades were in large part predicted by high school preparation (class rank, test scores, and prior research participation), which means that "success" in introductory courses was more related to previous preparation than to science skills developed in these courses.

If grading practices serve as the primary sorting mechanisms that colleges employ in science, and if grades in large introductory are meted out on a curve with relatively few students earning high marks, then talented students who do not out-compete their peers may be weeded out of science majors very early in their college career. Students who do not earn top grades do not necessarily lack the skills needed to be a good scientist; they may simply lack the prior preparation or study skills needed to perform well in lecture-based classes that reward cramming for exams. To keep talented students in science majors, we need to broaden performance criteria and assessment techniques. Grades alone will not identify the nascent scientific talent that exists among college students.

Interestingly, students earned higher grades when they were involved in courses where group work was encouraged. This is congruent with the literature that shows the myriad benefits of group work in terms of student performance (Armstrong, Chang, & Brickman, 2007; Knight & Wood, 2005; Shipman & Duch, 2001; Springer, Stanne, & Donovan, 1999). Providing students with opportunities to share ideas and evaluate multiple sources of evidence reinforces content and concepts presented in class (Armstrong, Chang, and Brickman, 2007; Knight & Wood, 2005). Such reinforcement of learning may have longer-term benefits in terms of students' retention of domain-specific knowledge (Koslowski, 1996; Kuhn, Garcia-Mila, Zohar, & Anderson, 1995).

Implications

Many NSF-funded projects are specifically devoted to interventions that are designed to improve the teaching and learning of science, yet there remains incredible resistance to change. Our study suggests that science faculty must confront the questions of whether we can afford to cram content into students at the expense of the development of scientific skills and thinking, and whether we can continue to let grading practices reflect previous preparation rather than actual learning in the classroom. With increased interest in STEM among entering students (Higher Education Research Institute, 2010), the U.S. is at a critical crossroads in terms of its opportunity to improve the production of science degrees. In order to move forward most productively, faculty must reexamine current teaching and grading practices.

Faculty cannot move forward on this path alone; they will need help from researchers and practitioners, and, in particular, institutional researchers and evaluators. Currently, the vast majority of students are still in large lecture venues in introductory science and mathematics courses, and this is not likely to change. However, this does not mean that faculty cannot change the way that they teach students and develop scientific talent. Further research is needed to understand the impact of more varied and engaging pedagogies used by faculty in science.

If introductory science curriculum continues to emphasize only the transmission of content knowledge at the expense of more general higher-order thinking skills, we risk losing a significant number of future independent thinkers. Instead, if introductory courses can instill in students and reward them for mastery of critical thinking skills, we have the opportunity to develop young scientists equipped not only to master scientific concepts and knowledge but also to critique pre-existing knowledge. Investments made in these areas are necessary to open the valve in the pipeline that is preventing the movement of current students past introductory science coursework.

References

Allen, D. E., Duch, B. J., & Groh, S. E. (1996). The power of problem-based learning in teaching introductory science courses. In Wilkerson, L. and Gijselaers (Eds.), *New Directions for Teaching and Learning, 68.* San Francisco: Jossey-Bass.

Allen, D., and Tanner, K. (2002). Answers worth waiting for: One second is hardly enough. Cell Biology Education, *1*(1), 3–5.

Allen, D., & Tanner, K. (2005). Infusing active learning into the large-enrollment biology class: seven strategies, from the simple to complex. *Cell Biology Education, 4,* 262-268.

Armstrong, N., Chang, S., & Brickman, M. (2007). Cooperative learning in industrial-sized biology classes. *Life Sciences Education*, *6*(2), 163-171.

Astin, A. W. (1993). *What matters in college? Four critical years revisited*. San Francisco, CA: Jossey-Bass.

Barlow, A. E. L., & Villarejo, M. (2004). Making a difference for minorities: Evaluation of an educational enrichment program. *Journal of Research in Science Teaching, 41*(9), 861-881.

Barnett, S. M. & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612-637.

Beck, B.L., Koons, S.R., & Milgram, D.L. (2000). Correlates and consequences of behavioral procrastination: The effects of academic procrastination, self-consciousness, self-esteem, and self-handicapping. *Journal of Social Behavior and Personality*, *15*, 3-13.

Bentler, P. M. (2006). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software, Inc.

Bentler, P. M. & Wu, E. J. C. (2002). *EQS 6 for Windows user's guide.* Encino, CA: Multivariate Software, Inc.

Bulte, C., Betts, A., Garner, K., & Durning, S. (2007) Student teaching: Views of student near-peer teachers and learners. *Medical Teacher*, *29*(6), 583-590.

Center for Institutional Data Exchange and Analysis. (2000). *1999-2000 SMET retention report*. Norman, OK: University of Oklahoma.

Cobb, P. (1994). Where is the mind: Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher*, *23*(1), 13-19.

Committee on Science, Engineering, and Public Policy. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Retrieved February 25, 2008, from National Academies Press website, http://www.nap.edu/catalog/11463.html

Conley, D. T. (2005). *College knowledge: What it really takes for students to succeed and what we can do to get them ready.* San Francisco: Jossey-Bass.

Council of Graduate Schools. (2007). *Graduate education: The backbone of American competitiveness and innovation*. Washington, DC: Council of Graduate Schools.

Driver, R., Asoko, H., Leach, J., Scott, P., & Mortimer, E. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher*, *23*(7), 5-12.

Epstein, D. (2006). So that's why they're leaving. *Inside Higher Education* Retrieved February 2, 2007, from http://insidehighered.com/news/2006/07/26/scipipeline

Espinosa, L. (2009). *Pipelines and pathways: Women of color in STEM majors and the experiences that shape their persistence.* Unpublished doctoral dissertation.

Foertsch, J., Alexander, B. B., & Penberthy, D. (1997). *Summer research opportunity programs (SROPs) for minority undergraduates: A longitudinal study of program outcomes, 1986-1996*. Madison, WI: The Lead Center, University of Wisconsin-Madison.

Fraser, B. J. & Fisher, D. L. (1982). Predicting students' outcomes from their perceptions of the classroom psychosocial environment. *American Educational Research Journal, 19*(4), 498-518.

Freedman, S. W. (1994). *Exchanging writing, exchanging cultures: Lessons in school reform from the United States and Great Britain*. Cambridge, MA and Urbana, IL: Harvard University Press and National Council of Teachers of English.

Gainen, J. (1995). Barriers to success in quantitative gatekeeper courses. In J. Gainen & E. W. Willemsen, (Eds.), *Fostering student success in quantitative gateway courses* .New Directions for Teaching and Learning, 61. San Francisco: Jossey-Bass.

Giancarlo, C. A. & Facione, P. A. (2001). A look across four years at the disposition toward critical thinking among undergraduate students. *The Journal of General Education*, *50*(1), 29-55.

Glynn, L. G., MacFarlane, A., Kelly, M., Cantillon, P., & Murphy, A. W. (2006). Helping each other to learn: A process evaluation of peer-assisted learning. *BMC Medical Education*, *6*(18), 1-9.

Hagedorn, L. S., Siadat, M. V., Fogel, S. F., Nora, A., & Pascarella, E. T. (1999). Success in college mathematics: Comparisons between remedial and nonremedial first-year college students. *Research in Higher Education*, *40*(3), 261-284.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*, 64-74.

Haertel, G. D., Walberg, H. J. & Haertel, E. H. (1981). Socio-psychological environments and learning: A quantitative synthesis. *British Educational Research Journal*, *7,* 27-36.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., et al. (2004). Policy Forum: Scientific teaching. *Science*, *304*(5670), 521-522.

Hurd, P. D. (1998). Scientific literacy: New minds for a changing world. *Science Education*, *82*(3), 407-416.

Kamii, M. (1990). Opening the algebra gate: Removing obstacles to success in college preparatory mathematics courses. *Journal of Negro Education*, *59*(3), 392-406.

Knight, J. K. & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, *4*(4), 298-310.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning.* Cambridge, MA: MIT Press.

Kuhn, D., M. Garcia-Mila, A. Zohar, and C. Anderson. 1995. *Strategies of Knowledge Acquisition*. Chicago: Society for Research in Child Development.

Labov, J. B. (2004). From the National Academies: The challenges and opportunities for improving undergraduate science education through introductory courses. *Cell Biology Education*, *3*(4), 212-214.

Laird, T. F. N., Engberg, M. E., & Hurtado, S. (2005). Modeling accentuation effects: Enrolling in a diversity course and the importance of social action engagement. *The Journal of Higher Education*, *76*(4), 448-476.

Lockspeiser, T. M., O'Sullivan, P., Teherani, A., Muller J. (2008). Understanding the experience of being taught by peers: The value of social and cognitive congruence. *Advances in Health Sciences Education, 13*(3), 361-372.
Lopatto, D. (2004). Survey of undergraduate research experiences (SURE): First findings. *Cell Biology Education, 3*(4), 270-277.
Mazur, E. (1992). Qualitative vs. quantitative thinking: Are we teaching the right thing? *Optics and Phonics News, February,* 38.
National Center for Education Statistics. (2000). *Entry and persistence of women and minorities in college science and engineering, NCES 2000-601.* Washington, DC: U.S. Department of Education.
National Science Board. (2003). *The science and engineering workforce: Realizing America's potential, NSB 03-69.* Arlington, VA: National Science Foundation.
National Science Board. (2008). *Digest of key science and engineering indicators 2008, NSB 08-2.* Arlington, VA: National Science Foundation.
Payzant, T. W., and Wolf, D. P. (1993). Piloting pacesetter: Helping at-risk students meet high standards. *Educational Leadership* 50(5): 42-45.
Prenzel, M., Kramer, K., & Drechsel, B. (2002). Self-determined and interested learning in vocational education. In K. Beck (Ed.), *Teaching–learning processes in vocational education* (pp. 43–68). Frankfurt .: Peter Lang.
Pryor, J. H., Hurtado, S., DeAngelo, L., Sharkness, J., Romero, L. C., Korn, W. S., & Tran, S. (2008). *The American freshman: National norms fall 2008.* Los Angeles, CA: Higher Education Research Institute, UCLA.
Raykov, T., Tomer, A., & Nesselroade, J. R. (1991). Reporting structural equation modeling results in *Psychology and Aging*: Some proposed guidelines. *Psychology and Aging, 6*(4), 499-503.
Resnick, L. B. (1987). *Education and learning to think.* Washington, DC: National Academy Press.
Sabatini, D. A. (1997). Teaching and research synergism: The undergraduate research experience. *Journal of Professional Issues in Engineering Education and Practice, 123*(3), 98–102.
Sagan, C. (1996), *The Demon-haunted World: Science as a Candle in the Dark*. New York: Ballantine Books.
Seidel, T. (2006). The role of student characteristics in studying micro teaching-learning environments. *Learning Environments Research, 9*(3), 253-271.
Seymour, E. (2001). Tracking the processes of change in US undergraduate education in science, mathematics, engineering, and technology. *Science Education, 86*(1), 79-105.
Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Boulder, CO: Westview Press.
Shipman, H. L. & Duch, B. J. (2001). Problem-based learning in large and very large classes. In Duch, B. J., Groh, S. E., & Allen, D. E. (Eds.), *The power of problem-based learning: A practical 'how to' for teaching undergraduate courses in any discipline.* Sterling, VA: Stylus Publishing.
Smith, D. G. (1977). College classroom interactions and critical thinking. *Journal of Educational Psychology, 69*(2), 180-190.
Smith, D. G. (1981, March). *Instruction and outcomes in an undergraduate setting.* Paper

presented at the meeting of the American Educational Research Association, Los Angeles, CA.

Smith, K. A., Sheppard, S. D., Johnson, D. W., & Johnson, R. T. (2005). Pedagogies of engagement: Classroom-based practices. *Journal of Engineering Education,* 94(1), 87-102.

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, *69*(1), 21-51.

Struthers, C.W., Perry, R.P., & Menec, V.H. (2000). An examination of the relationship among academic stress, coping, motivation, and performance in college. *Research in Higher Education*, *41*(5), 581-592.

Summers, M. F., & Hrabowski Iii, F. A. (2006). Diversity enhanced: Preparing minority scientists and engineers. *Science*, *311*(5769), 1870-1871.

Tanner, K. & Allen, D. (2005). Approaches to biology teaching and learning: Understanding the wrong answers—teaching toward conceptual change. *Cell Biology Education, 4*(2), 112-117.

Terenzini, P. T., Theophilides, C., & Lorang, W. G. (1984). Influences on students' perceptions of their academic skills development during college. *The Journal of Higher Education, 55*(5), 621-636.

Tobias, S. (1990). Stemming the science shortfall at college. In S. Tobias (Ed.), *They're not dumb, they're different.* Tucson, AZ: Research Corporation.

Tobias, S. (1992). Science education reform: What's wrong with the process? In S. Tobias (Ed.), *Revitalizing undergraduate science: Why some things work and most don't* (pp. 11-22). Tucson, AZ: Research Corporation.

Tsui, L. (1999). Courses and instruction affecting critical thinking. *Research in Higher Education, 40*2), 185-200.

Tsui, L. (2002). Fostering critical thinking through effective pedagogy: Evidence from four institutional case studies. *The Journal of Higher Education, 73*(6), 740-763.

Vahala, M. E. & Winston, Jr., R. B. (1994). College classroom environments: Disciplinary and institutional-type differences and effects on academic achievement in introductory courses. *Innovative Higher Education, 19*(2), 99-122.

Waits, B. K., and Demana, F. (1988). Relationship between mathematics skills of entering students and their success in college. *School Counselor*35(4): 307-310.

Walberg, H. J. (1979). *Educational environments and effects: Evaluation, policy, and productivity*. Berkeley, CA: McCutchan.

Wesley, J. (1994). Effects of ability, high school achievement, and procrastinatory behavior on college performance. *Educational and Psychological Measurement*, *54*, 404-408.

Williams, W. M., Papierno, P. B., Makel, M. C., & Ceci, S. J. (2004). Thinking like a scientist about real-world problems: The Cornell Institute for Research on Children science education program. *Journal of Applied Developmental Psychology, 25*(1), 107-126.

Table 1

Factor Loadings for the latent constructs in the model

	Pre-Test	Post-Test
Thinking Like a Scientist		
Make connections between different areas of science and math	0.67	0.70
Make sense of scientific concepts	0.71	0.72
Identify what is known about a problem	0.63	0.63
Ask relevant questions	0.60	0.64
Draw a picture to represent a problem or concept	0.46	0.51
Make predictions based on existing knowledge	0.69	0.79
Come up with solutions to problems and explain them to others	0.67	0.72
Investigate alternative solutions to a problem	0.67	0.68
Translate scientific terminology into non-scientific language	0.57	0.62
Acting Like a Scientist		
Relate scientific concepts to real-world problems	0.71	0.75
Synthesize several sources of information	0.70	0.70
Conduct an experiment	0.54	0.54
Look up scientific research articles and resources	0.59	0.57
Memorize large quantities of information	0.41	0.44

Table 2

Parameter estimates for direct effects in the structural model

	b	B	S.E.	Sig.
Grade				
Class rank	0.20	0.17	0.02	***
Aspire for a research career	-0.08	-0.06	0.03	*
Participated in a pre-college research program	0.08	0.06	0.03	*
Composite SAT score	0.20	0.25	0.01	***
Crammed for exams	-0.08	-0.08	0.02	**
Felt bored in class	-0.15	-0.14	0.02	***
Frequency: Professor relied on multiple-choice exam questions	0.10	0.05	0.03	*
Percentage of class time professor allotted for group work	0.01	0.07	0.00	*
Thinking Like a Scientist				
Pre-test: Thinking like a scientist	0.60	0.61	0.03	***
Class rank	-0.02	-0.02	0.01	
Aspire to a research career	0.04	0.06	0.01	**
Participated in a pre-college research program	0.06	0.09	0.01	**
Self-rated scientific ability	0.21	0.27	0.01	***
Self-rated math ability	0.05	0.07	0.01	***
White (non-White is the reference group)	0.07	0.05	0.01	**
Joined or created a study group	0.05	0.13	0.01	***
Felt board in class	-0.02	-0.03	0.01	*
Goal: Prepare students for graduate education	0.05	0.03	0.03	
Percentage of class time professor allotted for group work	-0.01	0.03	0.01	*
Acting Like a Scientist				
Pre-test: Acting like a scientist	0.65	0.62	0.03	***
Class rank	-0.02	-0.04	0.01	
Aspire to a research career	0.05	0.08	0.01	**
Participated in a pre-college research program	0.07	0.10	0.01	**
Self-rated science ability	0.23	0.28	0.01	***
White (non-White is the reference group)	0.05	0.04	0.02	*
Joined or created a study group	0.06	0.13	0.01	**
Goal: Prepare students for graduate education	0.08	0.04	0.03	*
Frequency: Used electronic quizzes for immediate feedback	-0.03	-0.04	0.01	*

Figure 1
Final Structural Equation Model