Chapter 2

SCIENTIFIC RESEARCH SELF-EFFICACY AMONG UNDERGRADUATES: CURRENT CONTEXTS AND APPROACHES FOR MEASUREMENT

Christopher T. Goode1, Shari L. Britner2, Phillip Gagné1, John L. Pecore3, Melissa K. Demetrikopoulos4, Brian A. Williams1, Laura L. Carruth1, Robert L. DeHaan5 and Kyle J. Frantz1

1Georgia State University, US
2Bradley University, US
3Temple University, US
4The Institute for Biomedical Philosophy, US
5Emory University, US

ABSTRACT

To help address current concerns about the strength and diversity of the United States scientific and technical workforce, expert panels and organizations recommend recruiting young people into active research environments. Here we present a brief review of the construct of self-efficacy and its related social cognitive career theory, as they pertain to research and careers in Science, Technology, Engineering, and Math (STEM). We focus on the use of measures of self-efficacy (SE) for tasks that predict academic preparation in STEM fields and intent to persist in STEM careers, particularly for undergraduate students who are members of demographic groups currently underrepresented in the sciences (e.g., racial and ethnic groups, individuals from disadvantaged socioeconomic or educational backgrounds, and those with disabilities). We present a sample study from our own work, wherein we compare two approaches to providing

* Contact Information: Christopher T. Goode, PO Box 5010, Atlanta, GA, 30302-5010, Email: goode@gsu.edu, Phone: 404-413-6273.
undergraduates with a summer research experience in neuroscience: a traditional mentored apprenticeship model (AM) and a collaborative learning model (CLM). We used measures of scientific research self-efficacy and leadership/teamwork self-efficacy, among other instruments, as tools to assess and compare the models. We further present results of a qualitative case study of four individuals in our program to help describe their research experience and identify potential sources of self-efficacy. Both the AM and CLM significantly increased scientific research and leadership/teamwork self-efficacy. The qualitative data helped identify specific program components, essentially mastery experiences in scientific research, that were sources of gains in self-efficacy. We relate our measures of self-efficacy to those of other relevant constructs, such as science anxiety, identity, and commitment to science, and discuss these results in the context of social cognitive career theory. Continued work applying self-efficacy and social cognitive career theory will help strengthen and diversify the scientific workforce.

INTRODUCTION

By many accounts, too few of the most talented individuals in the United States are attracted to studies and careers in science, technology, engineering and mathematics (STEM). As a nation, we therefore push to strengthen and diversify the STEM workforce, in part by recruiting young learners into active research environments. We focus here on the construct of self-efficacy, using it to help explore students' motivations for pursuing academic preparation, their experiences in pre-professional research environments, and their progress toward subsequent careers in STEM fields. First described by Bandura (1977) as one's confidence in one's ability to carry out a particular task or perform within a given domain of skills, this construct has immense predictive value with respect to success within a specific domain. Self-efficacy thus forms the basis of social cognitive career theory (SCCT; Lent, Brown, and Hackett, 1994) which explains relationships between self-efficacy for job-specific skills, and interest in, progress toward, and successful attainment of those jobs. Extensive evidence links self-efficacy, especially in science and mathematics skills, to interest, progress, and success in STEM careers.

In this context, our research team aims to join others in incorporating self-efficacy measures into the portfolio of instruments used to assess and predict those strategies that will improve the STEM academic and career pipeline. Many measurements of STEM-related self-efficacy have been developed, from general science and mathematics self-efficacy (Betz and Hackett, 1983; Lent, Brown, and Larkin, 1984), to self-efficacy within a specific course of study, e.g. anatomy and physiology (Witt-Rose, 2003). Here, we focus on measurement of self-efficacy in scientific research skills, or scientific research self-efficacy. Although we do not attempt an exhaustive review of STEM-related measurements of self-efficacy, we do aim to highlight some related work to date. Because self-efficacy is a reflective construct, it is measured by self-report, and it is one's estimate of one's abilities, not actual performance, that is at the heart of the construct. Thus, self-efficacy measurement lends itself quite well to paper-and-pencil or online surveys. The dynamic nature of self-efficacy lends itself well to within-subjects, repeated measures, experimental designs measuring changes in self-efficacy over time, and/or mixed designs testing for differences between populations in their self-efficacy changes over time. A multi-method approach integrating traditional quantitative self-efficacy measures with varied qualitative assessments is also recommended. In this chapter,
we review some methods for measurement of self-efficacy, comment on challenges inherent in these methods, and provide our example of measurement and analysis.

We close with a summary of our own sample study in which scientific research self-efficacy is probed among undergraduate students participating in a summer research program, known as Behavioral Research Advancements In Neuroscience (BRAIN). Finally, we integrate discussion of some of our research results with calls for future research to answer outstanding questions about scientific research self-efficacy, recruitment, and retention of bright individuals in STEM careers.

MOTIVATION AND CONTEXTS FOR INVESTIGATING SCIENTIFIC RESEARCH SELF-EFFICACY AMONG UNDERGRADUATES

In an era of globalization and rapid technological advance, effective education in science is of major concern worldwide. In the U.S., declining interests in many STEM disciplines (Teachers College, 2005) and low levels of basic reasoning skills among college students (Cracolice, Deming, and Ehlert, 2008) are particularly alarming. For example, approximately 50% of early college science majors defect to other areas of study (Seymour, 1995a, 1995b; Seymour and Hewitt, 1997), and ultimately only 6% of current U.S. undergraduate degrees are awarded in the natural sciences (NSF, 2010a). Understanding factors that influence decisions about fields of study and career paths has therefore become a national priority (NIH, 2011; NSF, 2010b, 2011; USDOE, 2011). Self-efficacy in scientific domains and mathematics is a robust predictor of interest, intent to persist in, and actual progress through STEM academic preparation and careers (Lent, Brown, and Larkin, 1986), and enhancing science self-efficacy should therefore be pursued as a mechanism to improve STEM education and expand the STEM workforce.

Pressure also exists at the national level to increase diversity among professionals in STEM careers (NIGMS, 2010; NIH, 2009; NSF, 2011b). The rates of defection from science majors mentioned above is even greater than 50% for women and students from racial and ethnic groups currently under-represented in the sciences (Seymour, 1995a). Despite a recent doubling of the number of doctorates earned in STEM compared with other fields, diversity remains low among STEM graduate students, with only 7% of degrees attained by African Americans, 6% by Hispanics, and less than 1% by Native Americans or Alaskan Natives, figures that are less than half the proportional representation of these groups in the U.S. population (NSF, 2011a). With regard to women, some STEM fields boast increasing involvement by female trainees over the last 50 years, such that more doctoral degrees in the life sciences are now earned by women than men, for example (NSF, 2011a). Yet career progress remains a significant concern for women, as well as racial and ethnic minorities. In neuroscience (our field of focus), the proportion of women in professional positions declines as ranks rise, from 37% of post-doctoral fellows, to 34% of assistant professors, and 31% of associate professors, down to 26% of full professors, and only 25% overall in STEM management jobs; although notably the percentage of full professors in neuroscience who are women has risen from 9% to its current 26% over the last 25 years (SFN, 2009; USDOC, 2011). The most striking gender discrepancies remain in computer science, engineering, and
math jobs (with 24, 27, and 14% women at assistant, associate, and full professor levels, respectively)(USDOC, 2011).

These data demonstrate that the current cadre of scientific investigators charged with providing new knowledge to improve human health is not representative of the general population (Olson and Fagen, 2007), and thus the research agenda may proceed at odds with the needs of many U.S. citizens. Diversity among science and math professionals is also a goal in part because it maintains our competitive edge in the global economy (CED, 2006; Wenzel, 2004) and enhances some group dynamics (Milem, 2003; Nemeth, 1985). For example, in the context of scientific research, a diversity of research backgrounds among members of a team improves the generation of alternative hypotheses or models to test in lab groups (Dunbar, 1995).

As a targeted approach to retention of talented individuals from diverse backgrounds in STEM fields, numerous panels have recommended getting young people involved in scientific research (CED, 2006; NRC, 1996, 2003; NSB, 2004). There is perhaps no better way than an authentic research experience to introduce students to the practical application of science-related knowledge and skill, as well as the thrill of scientific discovery. It is a literal example of the “inquiry-based” learning that has been called for repeatedly (Hofstein and Lunetta, 2004; NRC, 2003, 2005). This type of experience may be especially important for novice scientists (Barab and Hay, 2001; Brooks and Brooks, 1999; Hofstein and Lunetta, 2004; Lederman, Abd-El-Khalick, Bell, and Schwartz, 2002), and for recruitment of minorities and women into the sciences (Lopatto, 2004; NRC, 2003). At the K-12 level, these recommendations have prompted science education standards to emphasize inquiry, problem-solving, and applied approaches (NCES, 2011). This is also a focus for those in higher education, professional trainees, and even young professionals, as federally-funded grant programs encourage top performers to incorporate research into their career paths, with emphasis on trainees from under-represented groups including racial and ethnic minorities, individuals from disadvantaged socioeconomic or educational backgrounds, and those with disabilities (NIGMS, 2010; NIH, 2009; NIMHD, 2011; NSF, 2011b).

In practice, recruitment of young people into the research environment is often carried out through apprenticeship programs, with generally positive outcomes. Offered during the summer or the academic year, apprenticeships such as the NSF-sponsored “Research Experiences for Undergraduates” (REU) match trainees with a mentor for full- or part-time work, and pay wages or count for course credit. In a thorough meta-analysis and ethnography, Seymour, Hunter, Laursen, and Deantoni (2004) concluded that undergraduate research can maintain positive attitudes toward science, elevate scientific thinking, and increase research skills. Almost all participants describe some benefit from this type of program, many in personal-professional growth (e.g. confidence, independence, responsibility in the laboratory). Additional learning gains include broadened views on the nature of science, enhanced understanding of the research process and how to approach scientific problems, as well as readiness for more demanding research (Lopatto, 2007; Ryder, Leach, and Driver, 1999). Whether or not these short-term gains translate into persistence in research fields is harder to assess, but a few recent studies provide excellent data in the affirmative. A survey of over 3,000 undergraduate program alumni indicates that sustained undergraduate research increases intent to pursue a PhD by a factor of four, compared to no participation (Russell, Hancock, and McCullough 2007). Another large, longitudinal study confirms that undergraduates participating in research sustain their intent to persist in STEM careers,
least over three years (Schultz et al., 2011), whereas a matched cohort of students who did not participate in research exhibited declining intent to persist. These longitudinal data are quite compelling, yet many investigators are challenged to maintain contact with the inherently transient population of undergraduate program alumni. Thus, a more proximal measure, robustly predictive of longer-term outcomes, is desired in order to evaluate the effectiveness of undergraduate research in achieving goals of strengthening and diversifying the scientific workforce. As summarized below, self-efficacy is a good choice. In fact, Kardash (2000) reported increased efficacy with research skills among undergraduates engaged in a summer research experience. Coupled with the aforementioned gains in personal constructs broadly related to cognition, emotion, and skill, this study suggests that a research apprenticeship is likely to increase self-efficacy as Bandura initially conceptualized it (Bandura, 1997), and thus promote career choices and actions related to scientific research (Lent et al., 1984, 1986; Lent et al., 1994). In our sample study described later in the chapter, we measured participant self-efficacy for a set of general scientific research skills.

In summary, our motivation to integrate the self-efficacy construct into the field of undergraduate scientific research stems mainly from our overarching goals to recruit and retain a diverse group of the most talented trainees and young professionals in pathways toward research careers, and to address gaps in current knowledge about readily measured proximal gains that might reliably predict long-term career success. Our context for conducting this work is the undergraduate research experience, given the widely accepted recommendation to demonstrate the wonder of discovery through research.

**SCIENTIFIC RESEARCH SELF-EFFICACY AND ITS POTENTIAL TO PREDICT PERSEVERANCE IN PATHWAYS TOWARD RESEARCH CAREERS**

As defined throughout this volume, self-efficacy is an individual’s confidence that he or she can successfully perform a task or behavior (Bandura, 1977, 1986). In academics, self-efficacy is students’ judgments of their abilities to perform specific academic tasks (Lent et al., 1984). It predicts one’s choice of activities, environments, efforts, and persistence toward a goal, and it is task- or situation-specific (Bandura, 1986, 1997). An individual may have high self-efficacy for a certain set of tasks, and very low self-efficacy for unrelated tasks. Although some measures of a “generalized” self-efficacy have been developed (e.g., Sherer et al., 1982), most instruments probe efficacy for a specific set of skills, e.g. mathematics, science, or anatomy. The predictive nature of self-efficacy is also task-specific, such that science grade self-efficacy, but not self-regulation self-efficacy, predicts science achievement (grades) for middle and high schoolers (Britner, 2008; Britner and Pajares, 2001, 2006); academic self-efficacy predicts academic performance, not athletic performance, etc. Different measures of self-efficacy must be developed for different skill sets. Self-efficacy is also mutable, changing based on feedback from the four main sources of efficacy beliefs: personal accomplishments/mastery experience (e.g. engaging in and successfully completing a task), vicarious experiences (e.g. exposure to successful role models), social/verbal persuasion (e.g. feedback from a mentor), and physiological/affective states (e.g. anxiety
while attempting a task; Bandura, 1986, 1997). As skills are practiced and mastered, the skill set itself may expand, and self-efficacy may increase.

We are interested in applying measures of self-efficacy to gauge individual likelihood of pursuing and succeeding in STEM careers. We are particularly focused on measures of self-efficacy to predict future successful performance of research, as well as incorporation of research into an existing career profile. Several scales have been designed to specifically measure laboratory and scientific research skills self-efficacy (Britner, 2002; Chemers et al., 2010; Smist, 1993). The first scale was developed for middle-schoolers, and individual items measure self-efficacy for general laboratory tasks such as using equipment, recording results, and drawing conclusions. Smist (1993) created a comprehensive science self-efficacy questionnaire that has been used in high school and undergraduate students to probe self-efficacy for biology, chemistry, physics and general laboratory skills. Respondents are asked to rate their confidence level for performing five to eight behaviors in each domain, e.g. "doing physics lab experiments well," or "lighting a Bunsen burner." Intercorrelations showed strong divergent validity among the sub-scales (e.g. biology, chemistry), demonstrating that self-efficacy for a single course of study was relatively independent. Another scale was designed by Chemers et al. (2010) for undergraduates. Individual items probe self-efficacy for more cognitively advanced skills such as generating a research question, analyzing data, creating explanations for results, and using scientific literature to guide research. Additional items probe self-efficacy for skills that can later be assessed directly, e.g. reporting results in an oral presentation or written report.

Although specific tests to determine whether scientific research self-efficacy predicts success in research careers remain to be reported, a rich body of literature suggests that self-efficacy in related domains predicts future behaviors, including career-related choices and success (Bakken, Byars-Winston, and Wang, 2006; Bandura, 1977; Betz and Hackett, 2006; Chemers, Hu, and Garcia, 2001; Gloria and Robinson-Kurpius, 2001; Hackett and Betz, 1981; Pajares, 1993). Because self-efficacy is domain-specific, different measurements must be created for different sub-domains of STEM. Lent et al. (1984, 1986) measured self-efficacy related to both academic preparation and occupational requirements for scientific and technical careers. For example, the self-efficacy for Scientific and Technical Fields instrument (STF; Lent et al., 1984) probes self-efficacy with respect to the educational preparation and job duties for 15 STEM job titles. Individuals with higher self-efficacy scores persist longer in STEM majors and achieve higher grades than those with lower self-efficacy. The same measure of self-efficacy also predicts STEM grades, persistence in the STEM major, and the breadth of career options selected by study participants, even after controlling for previous achievement, math ability, and STEM career interest (Lent et al., 1986). This same scale also predicted interests in both science and engineering careers directly (Lent, Larkin, and Brown, 1989). Similarly, math self-efficacy predicts interest in science and math careers (Hackett, 1985). The Mathematics Self-Efficacy Scale (MSES; Betz and Hackett, 1983) is a 52-item scale that probes self-efficacy related to math tasks, math problems, and college courses. On the other hand, Hollinger (1983) measured science self-efficacy as a part of a 12-item scale, and it predicted both math and science career interests and intent to persist. Many of these scales have been validated for both men and women (reviewed by Betz and Hackett, 1986).

The powerful relationship between self-efficacy and career success gave rise to social cognitive career theory (SCCT; Lent et al., 1994). Briefly, the theory concerns three aspects
of career choice and preparation: development of interest in specific careers (or aspects of those careers), making concrete career decisions (e.g. committing to a degree program), and persistence toward or performance in the career itself. Self-efficacy, arising from the four sociocognitive sources described above (mastery experience, vicarious experiences, social persuasion, and physiological states), is presented as one of three primary mechanisms directing career interest, choices, and performance. Two other social cognitive factors are critical to the theory: outcome expectations, the expected results of a particular behavior or choice, and goals.

When introduced, social cognitive career theory was a revolutionary way to view factors affecting selection of an individual's career path, as it departed from the previous reliance on aptitude as the single predictor of career success. This is not to say self-efficacy is sufficient to predict optimum career choices for an individual; the overall model involves feedback between actual skill development, successful performance, and goal fulfillment. It follows then, that another deviation from prior theories is that the social cognitive career model is dynamic; perceived self-efficacy is reinforced (or not) by successes (or failures) in the application of skills. Moreover, self-efficacy works as a feed-forward system when confidence level slightly outpaces skill level for a particular domain. What Lent et al. (1994) called "modest over-confidence" promotes engagement in challenging tasks that, when mastered, boost both efficacy beliefs and performance. In keeping with the classic "zone of proximal development" (Vygotsky, 1980), confidence levels that are only slightly beyond one's actual ability constitute an ideal zone to drive improved performance.

Our motivation to explore self-efficacy is part of a larger goal of diversifying the scientific workforce, so we are also interested in the validity of self-efficacy and social cognitive career theory across populations. In fact, many self-efficacy measures have been validated in varied populations, and demonstrate that self-efficacy may vary as a function of demographics including race/ethnicity, gender, and age (Betz and Hackett, 2006; Britner and Pajares, 2001; Gloria and Robinson-Kurpius, 2001; Hackett and Betz, 1981). For example, Gwilliam and Betz (2001) found five self-efficacy scales, including the self-efficacy in Scientific and Technical Fields (STF) and the Mathematics Self-Efficacy Scale (MSES) to be reliable and valid measures across both male and female, and African American and European American students. Klassen (2004) reviewed a number of cross-cultural evaluations of self-efficacy for a broad range of domains, and reported a remarkable difference in self-efficacy between collectivist and individualist cultures, with the latter showing greater self-efficacy in most studies. Scott and Mallinckrodt (2005) showed that high school girls with high science self-efficacy were more likely to choose a major in science than those low in science self-efficacy. Mathematics self-efficacy predicts math careers, outcome expectations, and interests in sixth graders (Turner, Steward, and Lapan, 2004). Math/science self-efficacy predicts math/science interests and goals for Mexican American students (Navarro, Flores, and Worthington, 2007). The models tested in these and other studies also include other domains relevant to self-efficacy and career choice, such as scientific and ethnic identity, family context, parental encouragement, gender-typing, and STEM career-related values. As such, these studies emphasize that social cognitive career theory must be applied in context, i.e. each participant in a study testing the theory presents with a unique personal history, social environment, emotional status, and value system. The influence of these types of factors on self-efficacy and career choice may be especially useful in describing differences between population subgroups, such as those currently underrepresented in STEM
professions. For example, particular sociocognitive experiences may be more likely to contribute to the self-efficacy of women of color than European American women (reviewed by Byars and Hackett, 1998). Gender-typical roles tend to be more fluid in African American than European American families, so African American girls may have more opportunities to master skills outside their traditional gender roles. Thus they may increase self-efficacy more broadly. Latina college students are more likely to be first-generation college students than European American women, and thus may not have as many vicarious experience opportunities. Native American women generally are more likely to experience adverse physiological/affective states due to conflict between personal and cultural/family goals with respect to education and careers. In terms of other factors critical to social cognitive career theory, outcome expectations are particularly relevant for women who are members of groups under-represented in the sciences (reviewed by Byars and Hackett, 1998).

**APPROACHES TO MEASURING SELF-EFFICACY**

Despite the complexity of self-efficacy, approaches to measuring this key construct can be surprisingly straightforward (Bandura, 2006). Paper-and-pencil surveys initially probed two dimensions of self-efficacy: magnitude, measured as the proportion of tasks in the domain of interest that the subject rated as 'achievable' on a binary scale (yes or no, on 10-15 listed tasks); and strength, measured as the depth of the subject’s confidence in his or her binary response for each task queried (i.e. confidence in one’s ability to perform a set of tasks; Bandura, 1977). In many cases, electronic surveys have now replaced the paper-and-pencil format, and many studies of self-efficacy focus on strength rather than magnitude of the construct (e.g. Lee and Bobko, 1994). Given that Bandura conceived of self-efficacy as a determinant of behavioral change, self-efficacy was originally measured at “critical junctions” in the process of learning, e.g. before treatment, after treatment, and again after a related test of performance or ability (e.g. Bandura, Adams, and Beyer, 1977). Thus, composite scores for magnitude and strength of self-efficacy are typically calculated, compared over time, and correlated with external ratings of performance on a series of related tasks. In terms of statistics, analysis therefore requires within-subjects, repeated measures, statistical designs. Further, early studies established the predictive nature of self-efficacy on performance, and thus self-efficacy became a measure of the effectiveness of a treatment to change behavior.

With regard to the specificity of self-efficacy, even early studies noted that self-efficacy for particular items predicted performance on matching tasks better than a composite score predicted overall performance, and others confirmed this finding (Pajares, 1997; Bandura et al., 1977). In the context of academics, composite scores may in fact have limited value in predicting particular outcomes (Pajares and Miller, 1995). This notion points out the struggle in design of self-efficacy instruments to balance specificity with relevance, i.e. researchers could create an instrument so specific to tasks at hand that the instrument would really bear no relevance to the real world, any other population, setting, or time, other than the one at hand (Lent and Hackett, 1987).

Only occasionally in the history of self-efficacy theory have researchers set out to manipulate self-efficacy directly, and measure it at a hypothesized critical junction in
behavioral change. One example is a study in which faculty members provided either positive or negative feedback to college students (thereby providing different types of ‘social persuasion’), measured task-specific self-efficacy before and after the feedback, and aligned the self-efficacy with subsequent approaches to problem-solving on a novel task (Bouffard-Bouchard, 1990). The sample study we outline below provides an example in which two different research experiences were compared for their ability to influence self-efficacy over a relatively short period of 10 weeks.

Additional instruments can be used to investigate the theorized sources of self-efficacy (e.g. Britner, 2008; Britner and Pajares, 2006). To this point we have been discussing quantitative approaches to probing and analyzing self-efficacy, which provide only a narrow profile of study participants. As described next, qualitative research methods can be coupled with these quantitative approaches in a mixed methods design to help address these concerns. The need for qualitative research lies in basic observations that the world is subjective, multidimensional, and holistic; and that people actively interpret the world around them (Merriam, 1988). Therefore, personal experiences narrated by study participants themselves, or documented by trained observers, are needed to provide richness and depth of description and explanation related to personal constructs such as self-efficacy (Zeldin, Britner, and Pajares, 2008). Qualitative case studies are especially appropriate when data collection and analysis are guided by hypothesis and theory, in order to describe the context in which participants may be changing, as well as their own interpretation about changes they experience. In the specific case of self-efficacy, qualitative research can help to explore an individual’s perceived sources of self-efficacy. For example, recent qualitative approaches have validated the influence of self-efficacy on STEM academic and career success, and showed that men and women report different primary sources of self-efficacy. Interviews revealed mastery experiences as the primary source of self-efficacy for men who entered science careers, but social persuasion and vicarious experiences as the primary sources of self-efficacy for women in science (Zeldin and Pajares, 2000; Zeldin et al., 2008). Multi-methods (quantitative and qualitative) studies of self-efficacy are perhaps best at providing a full profile of a population and its individual members, as quantitative measures can track change in the population, qualitative measures can describe the mechanisms of change for individuals, and each can inform the other regarding measure validation and refinement.

**Sociocognitive Constructs Likely to Interact with Self-Efficacy**

Although measuring self-efficacy for scientific research is perhaps the most salient to STEM careers, additional constructs are also likely to be influential. Career tasks, especially research endeavors in STEM fields, are necessarily group efforts. Despite popular stereotypes, there are no “solitary scientists.” Thus, abilities to lead others and/or to work effectively on a team are directly relevant to STEM careers. Chemers, Zurbriggen, Syed, Goza, and Bearman (2011) recently refined a scale to probe leadership and teamwork self-efficacy among undergraduate and graduate students, whereby respondents rate their agreement with statements such as, “I know how to cooperate effectively as a member of a team.” Just as math and science self-efficacy predict later choices and performance in STEM,
so does leadership and teamwork self-efficacy predict later performance in domains where teamwork is necessary (Chemers, Watson, and May, 2000; Watson, Chemers, and Preiser, 2001). Indeed, research experiences are associated with improvements in leadership and teamwork self-efficacy, which in turn is related to intent to persist in STEM careers (for graduate students and post-doctoral fellows; Chemers et al., 2011).

In addition to scientific research skills and leadership/teamwork self-efficacy, science anxiety, and science identity contribute to predictions of STEM career interests, choices, and persistence, particularly for women and under-represented minorities. Decades of measuring math anxiety reveal robust gender differences, with women scoring higher than men (see meta-analysis by Hembree, 1990). Several researchers have investigated a related construct, science anxiety, attributed to Mallow (1994), and found gender differences similar to those found with math anxiety (Britner, 2008; Britner and Pajares, 2006; Mallow, 1994). Because physiological and affective states constitute a primary source of self-efficacy, anxiety about science should be inversely correlated with science-oriented self-efficacy, and may thus help to predict STEM career progress.

Science can also be thought of as a "community of practice" (Lave and Wenger, 1991), a label that helps to drive the development of mechanisms for enculturation of novices into the discipline. This community of practice can provide experiences that constitute both vicarious and social persuasion sources of self-efficacy. The development of a sense of identity as a scientist is thus a central part of progress in STEM fields, especially when multiple social identities or group alignments are perceived as conflicting, such as identity as a scientist and identity as a woman (Settles, 2004) or as an African American (Brickhouse and Potter, 2001). Moreover, identity as a scientist predicts concrete steps toward STEM careers such as participating in research and applying to graduate schools, as well as persistence through the STEM career path from undergraduate to graduate status (Estrada, Woodcock, Hernandez, and Schultz, 2011).

Data from more qualitative work support the importance of identity as a scientist, particularly for women of color. Carlone and Johnson (2007) identified three types of scientific identity among women of color who were successful in science careers, including ‘research,’ ‘altruistic,’ and ‘disrupted’ scientists. ‘Research’ scientists saw the importance of science for the sake of science, ‘altruistic’ scientists viewed science as a mechanism for improving the human condition, and ‘disrupted’ scientists felt that at some point they were not recognized as a scientist and therefore their enculturation into the scientific community was disrupted.

A thoroughly validated and reliable instrument for measuring scientific identity remains to be established, but some researchers have made progress toward developing such a measure. The Science Identity Scale (Chemers et al., 2010) is a five-question survey whereby respondents rate their level of agreement with statements such as “I have come to think of myself as a scientist,” and “I feel like I belong in the field of science.” The internal validity of this scale is high (Estrada et al., 2011), and it has good face validity.

A final construct under current consideration is “intent to persist” in a given academic or career path. In science, one might refer to this intent as “commitment to science.” The construct relates to future intentions, plans, and goals, and thus can be measured at any point in a behavioral pathway of interest. Measures of intent to persist may predict concrete choices toward STEM careers, and later success directly, and/or may mediate the relationship between domain-specific self-efficacy and STEM career choices and progress. Although
often measured as responses to only one or two items within an instrument probing academics, social life, and/or career thoughts (e.g. Barker, McDowell, and Kalahar, 2009; Concannon and Barrow, 2010), intent to persist has been measured with a more thorough 12-item scale that includes sections on short-term career-related plans (e.g. "Next semester I intend to continue taking courses relevant to...") long-term career-related plans (e.g. "I intend to find a job as an..."), and degree attainment (e.g. "I would like to pursue a PhD in...") (Yonka Toker, unpublished measurements). The latter format emphasizes the idea that intent to persist may vary with time into the future and with the types of next-steps queried. As with the approaches to measuring self-efficacy discussed above, intent to persist appears to be considered somewhat 'static', and studies often seek to identify factors that determine the level of these intentions (Barker et al., 2009; Estrada et al., 2011; Mau, 2003). This work not surprisingly suggests that individual, departmental, institutional, and other environmental factors predict the intent to persist. The use of quasi-experimental design in which study participants engage in a program (or program components) designed to elevate their intent to persist appears rare, but would help to confirm which specific factors determine persistence, for which subgroups of participants.

For our sample study summarized next, we adapted measures of all of these constructs: scientific research self-efficacy, leadership/teamwork self-efficacy, science and neuroscience anxiety, and commitment to science. We implemented them in surveys at the beginning, middle, and end of a 10-week summer program for undergraduates conducting basic or applied research in the field of neuroscience. Our results suggest that scientific research self-efficacy and leadership/teamwork self-efficacy increase significantly, whereas anxiety declines and commitment to science does not change over this relatively short period. Individual interviews lend insight regarding the nature of the students’ experiences.

OUR SAMPLE STUDY

Introduction

To help recruit and retain talented individuals in pathways toward STEM careers, we designed a 10-week summer research program for undergraduate students. Known as "Behavioral Research Advancements in Neuroscience" (BRAIN), our program was designed to provide research experience in the field of behavioral neuroscience. Data from a pilot program were reported previously (Frantz, DeHaan, Demetrikopoulos, and Carruth, 2006). More recently we used a mixed-methods quantitative and qualitative approach to assess multiple program outcomes. Here, we present some results of both types of assessment to illustrate how we apply social cognitive career theory in the context of a summer undergraduate research experience. Specifically, we highlight some results from our quantitative analysis of responses on pre-, mid-, and post-program surveys, which probed scientific research self-efficacy, leadership/teamwork self-efficacy, science and neuroscience anxiety, science identity, and commitment to science. We also integrate those with a qualitative case study of four selected participants, whose responses in two semi-structured interviews focused on mastery of skills required for research as well as integration of research
with future career goals. We close the chapter with a discussion of our results, and general comments on science self-efficacy research.

Background

A specific challenge to effective undergraduate research experience may be recruitment of scientists willing to mentor novice undergraduates in summer programs. Thus, in our study we questioned whether undergraduate students could benefit from a collaborative, inquiry-based, student-driven research opportunity just as much, if not more, than those in a one-on-one research apprenticeship.

We used self-efficacy and several other measures to compare outcomes from two different program models: 1) a traditional Apprenticeship Model (AM) in which individual students join established research labs in basic science departments to work under individual faculty mentors; or 2) a distinctive Collaborative Learning Model (CLM) in which students work together in a dedicated teaching laboratory, in small student-driven research teams under the guidance of faculty, a post-doctoral fellow, and graduate student mentors to design and conduct original experiments, according to a defined but flexible curriculum.

Although research experience typically is not initiated until late in undergraduate careers, early participation in research may attract and retain undergraduates in science before they defect to other majors. Therefore, we biased admittance into our program to favor freshman and sophomore undergraduates, under-represented minority students, and women. Further recognizing that student retention and enculturation into the profession may be enhanced by effective mentors and successful, identifiable role models (Milem, 2003; NRC, 2003; Vygotsky, 1980; Wenzel, 2004), we recruited a diverse group of faculty members and advanced trainees (e.g. post-docs, graduate students) as research mentors and instructors for both program models.

Research Questions

1. Does a 10-week undergraduate summer research experience positively affect scientific research self-efficacy, leadership/teamwork self-efficacy, science and neuroscience anxiety, and commitment to science?
2. Do program outcomes vary between a traditional Apprenticeship Model (AM) and a Collaborative Learning Model (CLM)?
3. Do program outcomes vary with gender and ethnicity?
4. How do several selected participants describe their experiences in the program, especially with regard to scientific research self-efficacy and potential sources of gains or declines in this measure?

Participants

In spring 2009, 2010, and 2011, student participants were chosen in annual cohorts of approximately 40 from annual pools of approximately 200 applicants, recruited through local
advertisements, national and international conferences, websites for summer undergraduate research programs, and targeted electronic communication with colleagues nationwide. A committee of faculty members and senior trainees (post-doctoral fellows, senior graduate students) reviewed applications, first using a quantitative rubric, then a guided qualitative discussion of top-tier applicants. Review criteria included the following demographic characteristics: gender, ethnicity, academic year, home institution (in-state, out-of-state), research experience (yes, no), course preparation (biology, psychology, neuroscience, etc.), and grade point average (GPA). In order to favor admittance of applicants from populations currently underrepresented in STEM fields, and freshman and sophomore students with little research experience but adequate background knowledge from relevant course work, these characteristics were ranked on a weighted scale. For admissions purposes we also considered expressions of genuine interest in neuroscience research, academic degree aims, career plans, community involvement, leadership, writing skills, and letters of recommendation. Approximately 25% of original invitees declined for various reasons related to format assignment (see below), stipend, travel arrangements, family constraints, etc. Other applicants were invited from an extensive waiting list. To date, 97.5% of those who accepted the invitation completed the program that summer. Thus, 117 participants are considered in the present analysis, with 67% women, 56% academic underclassmen, 35% self-identified as Caucasian, 30% African-Descent/African-American, 16% Asian-Descent/Asian-American, 9% Hispanic/Latino/Latina, and over 10% selecting “Other” or “Do not wish to provide” categories for race.

Before invitation to the program, accepted students were assigned to AM or CLM models using stratified random assignment, such that AM and CLM cohorts were approximately matched on all the demographic characteristics listed above in the review considerations. Students were then invited into one of the two program models, and were not provided an opportunity to switch to the other. For the second and third cohorts, we provided the model assignments to students before they accepted our invitation to join the program in order to maintain participant satisfaction and maximize the likelihood that students could meet their own goals for that summer. For participants in the traditional AM, participant-mentor matches were facilitated in several ways. 1) Participants read summaries of research areas, and submitted five areas in rank order. 2) Program administrators suggested matches based on rankings and statements of interest from the program application. 3) Pre-program communication was encouraged to confirm or deny “fit.” Matching of mentors to students in the AM model was complete before participants arrived at the program site.

Local advertisements and targeted electronic communication were used to recruit interested mentors to provide research projects in their own neuroscience labs for participants in the AM. For the CLM, faculty, post-docs, and graduate students applied, interviewed, and were recruited to play a variety of instructional and mentoring roles in a single lab facility, including direct instruction, editing, coaching, and advising.

Procedure

Summer Research Program Structure

After a few days to move in and adjust to the local environment, the program began with one week of intensive classroom instruction in basic neuroscience, shared by all participants,
then nine weeks of neuroscience laboratory research in either the AM or the CLM group. The introductory classroom instruction addressed cellular and molecular neuroscience as well as behavioral neuroscience, using activities, lectures, and hands-on mini-experiments (approximately 9:00 a.m. – 5:00 p.m. daily over five or six days). During the subsequent nine weeks all participants were expected to work 35 hr/week in their laboratory settings. Participants also attended weekly 4 hr professional development workshops on topics including diversity in science career opportunities, graduate school preparation, stress management, science writing, how to develop effective poster presentations, and scientific ethics. The program culminated in the preparation of a written report (in the form of a journal article for the AM or a mini-research proposal for the CLM), as well as preparation of a research poster to be presented and judged at a closing research symposium. On successful completion of the program requirements, each participant received a stipend of $3-4,000 paid in increments.

Participants in the traditional AM (n=60 over three summers) joined in new or ongoing research projects in more than 30 different labs at five local research institutions. BRAIN program administrators exerted no influence over the nature of the research experience. Based on submission of weekly time sheets signed by mentors, participants fulfilled the expectation to conduct research activities for 35 hr/week, but daily schedules were designed individually by participants and mentors as they deemed fit for the diverse research paradigms, laboratories, and institutions that comprised the apprenticeship experiences.

Participants in the CLM (n=57 over three summers) all convened in a single dedicated laboratory (with neighboring seminar rooms and access to computers) to engage in various research techniques using an invertebrate animal model (red swamp crayfish; *Procambarus clarkii*). This species was chosen due to the extensive body of literature available on its cellular and molecular mechanisms of behavior, the relative simplicity of its nervous system, ease of care, and low level Institutional Animal Care and Use Committee oversight. A total of ten instructors were deployed over 9 weeks for the CLM (two faculty, three post-docs, three graduate students), with two or three present at any given time. They led demonstrations and experiments that required participants to use the following techniques: observation of animal behavior, anatomical dissection, histological staining, electrophysiological recording, RNA extraction from nervous tissue, quantitative PCR, and protein detection. During the first 4-5 weeks in the CLM, daily activities generally consisted of 1-2 hr introductions to new material (via lecture, demonstration, and discussion related to assigned readings), review of protocols, then initiation of experimentation in self-selected teams of two to three participants with assistance from instructors. Although all research teams used similar research techniques in a given week, their specific experimental questions were based on individual team interests. During the last three weeks in the CLM, each team designed and conducted its own pilot investigation on a unique topic chosen by team members. Usually, only one mentor was present during this period, but several instructors and mentors reviewed ideas, read proposed protocols, provided guidance, and assisted with data collection, during individual team meetings, consultations, and/or progress updates attended by all CLM participants. Weekly “journal clubs” facilitated comprehension of peer-reviewed articles on crayfish neurobehavioral research.
Program Assessment

All data collection was conducted with approval from the Georgia State University Institutional Review Board. Prior to arrival at the program location, participants were required to complete an on-line consent form followed by an electronic survey (pre-program survey). Again at program midpoint and finally after the closing research symposium, participants completed two more electronic surveys (mid- and post-program surveys, respectively). Surveys generally are valid and reliable measures of attitudes, behavior, and values of undergraduate students (Hinton, 1993). Six different inventories on the surveys are highlighted in this chapter. Survey completion was tracked by a unique numerical identifier and required for receipt of stipends. Personal identifiers were removed before data analysis to guarantee anonymity.

The first inventory probed participants’ scientific research self-efficacy, i.e. confidence in their own abilities to carry out research-related activities (adapted from Chemers et al., 2001). Because we were interested in pre-existing perceptions before the program and subsequent changes over the course of the program, we measured scientific research self-efficacy repeatedly: on the pre-, mid-, and post-program surveys. For our scientific research self-efficacy instrument, respondents were instructed to: “Indicate the extent to which you are confident you can successfully complete the following tasks,” followed by a list of ten tasks, such as “Use scientific language and terminology,” “Generate a research question to answer,” or “Relate results and explanations to the work of others.” A Likert scale of five possible responses ranged from “Not at all Confident” to “Absolutely Confident.”

A second inventory probed leadership and teamwork self-efficacy (adapted from Chemers et al., 2001). An introductory paragraph defined leadership as getting people to work together effectively, and defined teamwork as including “communication, collaboration, etc.” Respondents were asked to “Indicate the extent to which you disagree or agree with the following statements:” followed by a list of five items pertaining more to leadership, such as “I know a lot about what it takes to be a good leader,” intermixed with five items pertaining more to teamwork, such as “I know how to cooperate effectively as a member of a team.” A Likert scale of five possible responses ranged from “Strongly Disagree” through “Neutral” to “Strongly Agree”.

A third inventory addressed science identity (adapted from Chemers et al., 2001). Respondents were asked to rate their agreement with seven statements about scientific identity and belonging to the scientific community, such as “In general, being a scientist is an important part of my self-image,” or “I feel like I belong in the field of science.” A Likert scale of five possible responses ranged from “Strongly Disagree” through “Neutral” to “Strongly Agree”.

The next two inventories related to science and neuroscience anxiety, with nine items about science anxiety followed by eight more specifically about neuroscience (adapted from Britner, 2008). As above, participants were asked to rate their level of agreement with statements such as “It scares me to have to take science,” or “I feel a definite positive reaction to neuroscience.” A Likert scale of five possible responses ranged from “Strongly Disagree” through “Neutral” to “Strongly Agree”.

Commitment to science was investigated next, using a 7-item scale (adapted from Chemers et al., 2001). As above, participants were asked to rate their level of agreement with statements such as “I see the next steps in the field of science, and I intend to take them,” or
“I feel that I am on a definite career path in science.” A Likert scale of five possible responses ranged from “Strongly Disagree” through “Neutral” to “Strongly Agree”.

A final element of assessment included only on the post-program survey was intended as a formative assessment to guide future program structure, but it also provided a good measure of participants’ perception of the program’s value to their learning process. This is because we took a learning gains approach, modified from a Student Assessment of Learning Gains (SALG) instrument (Seymour, 1997; Seymour et al., 2004). According to SALG developers, students give clear indications regarding what they themselves gain from a learning environment, and the SALG is intended to assess the effects of specific teaching techniques or classroom activities on student learning. We probed this topic by asking participants to respond to this question: “How much did the following lectures, activities, and assignments help your learning?” A Likert scale of six possible responses included NA (for use if a participant was absent), then ranged from “No help” to “Very much help”.

To contribute to a rich understanding of undergraduate experiences throughout the summer program, each participant was also randomly assigned to participate in either an individual interview or a focus group discussion, each of which was conducted once 2-3 days before the start of the program, and again at the end of the program. Results from the focus group discussions are not described here. We used face-to-face, open-ended, semi-structured interviews in order to facilitate provision of personal narrative while ensuring consistency among interviews. Questions probed students’ ideas regarding the nature of scientific research, characteristics of the scientific community, characteristics of effective mentors, students’ beliefs about their ability to conduct scientific research, and their future goals and career trajectories. Interviews were conducted by one of two researchers, audio recorded, and transcribed verbatim.

**Data Analysis**

The data analysis highlighted here took two general forms: first, a quantitative summary of survey data; and second, a case study approach whereby four participants were selected to profile qualitatively in more depth. For the surveys, responses on individual items were assigned a numerical score, with 1 designated for “Not at all Confident,” “Strongly disagree,” or “No help,” depending on the inventory. When appropriate, items on the anxiety inventories were reverse-scored, so that greater anxiety or discomfort with science or neuroscience received a higher score. In order to confirm that each inventory demonstrated internal reliability and that individual items could be combined into a composite score, Cronbach’s alpha was calculated for each survey (raw reliability coefficients are reported). Scores were then summed to calculate a total for each inventory. In the event that participants were missing less than about 30% of individual items in a given inventory, then a hot-decking procedure was used to substitute values and insert an appropriate overall score for that inventory. In the case that a participant was missing more than about 30% of individual items, that participant’s score was dropped. Where a score was dropped from an individual time point, that participant’s data were not included in the repeated measures analysis of change over time, but available scores did remain in follow-up specific comparisons. Various repeated measures analyses of variance (ANOVA) were conducted separately for each inventory, with between-subjects factors such as gender, ethnicity (non-minority vs. under-represented minority), or program model (AM vs. CLM), and time as the repeated measure (pre-, mid-, post-program; Non-minority was defined as “Asian-Descent/Asian-American” or
"Caucasian", whereas under-represented minority was defined as "African-Descent/African American", "Hispanic/Latino/Latina", or "Native-American/Pacific-Islander"). In all cases, the Huynh-Feldt correction for possible violation of sphericity was at least 0.7 and univariate analyses are reported. SPSS (IBM, Armonk, NY) software was used for analyses.

For the interviews, four individual participants were chosen for in-depth analysis, based on their scientific research self-efficacy scores over the course of the program. We intended to analyze interviews from some participants whose self-efficacy rose during the program, some whose self-efficacy remained the same, and some whose self-efficacy fell. Very few students fell into the declining category, however, and all of these had been assigned to focus groups rather than individual interviews. Thus, we began with participants with below average pre-program self-efficacy, and identified those participants whose trajectory brought them near the mean at mid-program and above the mean post-program (rising self-efficacy). In the absence of viable candidates with declining self-efficacy scores, we identified instead those whose mid-program self-efficacy was relatively close to their pre- and post-program self-efficacy. In this way we selected four participants, two with rising self-efficacy and two with steady self-efficacy, with one of the latter in the mid-range of self-efficacy scores and the other very high (Figure 1).

Coding of transcript text for these individuals' interviews was done using NVivo (QSR, Cambridge, MA). A sample of the interviews was cross-coded to establish inter-rater reliability. Subsequent analysis occurred in two phases. To begin, we identified emerging first-level codes connected to students' descriptions of their scientific research self-efficacy, met periodically to discuss emerging findings, and created a set of codes that represented the first level codes. We then grouped first level codes into second level codes and identified emerging themes in the data, comparing ways in which students with different trajectories of self-efficacy discussed their confidence and their career plans.

![Figure 1. Scientific research self-efficacy as a function of time in the program. Four students' data are shown. Two students' research self-efficacy increased and two students' research self-efficacy remained relatively stable, with one remaining relatively high and the other at mid-range.](image_url)
Results

Case Numbers and Internal Reliability

In all reported analyses, there are missing cases. Some participants failed to respond to so many items on a particular instrument (>30%) that their case was dropped entirely from that measure. One participant withdrew consent to use data for reporting purposes.

To measure the internal reliability of our instruments probing scientific research self-efficacy, leadership/teamwork self-efficacy, science identity, science anxiety, neuroscience anxiety, and commitment to science, we computed coefficient alpha at pre-program, mid-program, and post-program, separately for each cohort in 2009, 2010, and 2011. Alpha levels ranged from .66 to .97, with the low reliability of .66 occurring on only one scale in one year (science identity in 2009) and far below the next lowest reliability score of .79.

Self-Efficacy

Mixed-design, 2x3 ANOVAs (α=.05) with Huynh-Feldt adjustments to degrees of freedom to correct for sphericity were used to test the effects of program model (AM vs. CLM) and time (pre-, mid-, or post-program) on both scientific research and leadership/teamwork self-efficacy. There was no significant effect of program model on either of the self-efficacy measures. There was a significant main effect of time for both measures of self-efficacy, F(1.81, 184.84)=124.92, p<.001; F(1.73, 176.91)=11.76, p<.001, with post-hoc testing revealing significant increases in scientific research self-efficacy from pre- (M=33.26, SD=7.13) to mid- (M=36.59, SD=5.52), t(105)=5.86, p<.001, to post-program (M=42.66, SD=5.78), t(107)=11.89, p<0.001, and in leadership/teamwork self-efficacy from pre- (M=42.51, SD=5.15) to mid-program (M=44.97, SD=4.66), t(105)=4.98, p<.001 (see Figure 2 and Figure 3).

Figure 2. Average scientific research self-efficacy by program model and time point. Error bars are one standard deviation. Scientific research self-efficacy increased significantly from pre- to mid- and from mid- to post-program.
Figure 3. Average leadership/teamwork self-efficacy by program model and time point. Error bars are one standard deviation. Leadership/teamwork self-efficacy increased significantly from pre- to mid-program and remained high from mid- to post-program.

Science Identity, Anxiety, and Commitment to Science

Again, 2x3, mixed-design ANOVAs (α=.05) with Huynh-Feldt degrees of freedom adjustments to correct for sphericity were used to test for the effects of program model (AM vs. CLM) and time (pre-, mid-, or post-program) on science identity, science anxiety, neuroscience anxiety, and commitment to science.

There were no significant main effects of program model on any of these variables. There was a significant effect of time on science identity, F(1.83, 186.53)=6.95, p<.01, with post-hoc testing revealing a significant increase from pre- (M=27.56; SD=4.94) to mid-program (M=28.35, SD=5.51), t(105)=3.05, p<.017, (see Figure 4).

Figure 4. Average science identity by program model and time point. Error bars indicate one standard deviation. Science identity increased significantly from mid-program to post-program.
Figure 5. Average science anxiety (left) and neuroscience anxiety (right) by program model and time point. Error bars indicate one standard deviation. Both science and neuroscience anxiety decreased significantly from mid- to post-program.

Figure 6. Average commitment to science by program model and time point. Error bars indicate one standard deviation. Commitment to science did not change significantly over the course of the program.

There were also significant effects of time on both science and neuroscience anxiety, F(1.92, 195.39)=5.07, p<.01; F(1.92, 209.11)=3.52, p<.05. Post-hoc testing revealed significant decreases in science and neuroscience anxiety from mid- (M=15.36, SD=5.57; M=13.52, SD=6.08) to post-program (M=14.02, SD=5.57; M=12.21, SD=5.57), t(107)=-3.18, p<.017; t(107)=-2.99, p<.017, (see Figure 5). On the other hand, commitment to science did not change over time in the program (see Figure 6).
Gender Differences in Scientific Research Self-Efficacy

In the absence of an effect of program model on scientific research self-efficacy, we collapsed data across models and analyzed for gender differences over time. A 2x3, mixed-design ANOVA was used to test for effects of gender and time point (pre-, mid-, post-program) on scientific research self-efficacy. Only the interaction was significant, F(1.85, 188.23)=3.21, p<.05. Post-hoc testing revealed that men reported a higher level of scientific research self-efficacy (M=35.57, SD=6.27) than women (M=32.13, SD=7.31) at the pre-program time point, t(83.82)=2.55, p<.05.

There were no significant differences in scientific research self-efficacy between men and women at either mid- or post-program. Women's scientific research self-efficacy increased significantly from pre- (M=32.01, SD=7.30) to mid- (M=35.99, SD=5.60), t(68)=5.54, p<.001, and mid- to post-program (M=42.61, SD=6.09), t(69)=11.84, p<.001; whereas men's increased significantly only from mid- (M=37.84, SD=5.21) to post-program (M=42.74, SD=5.21), t(37)=4.95, p<.01. Thus, men began the program reporting greater scientific research self-efficacy than women, but by mid-program the women had made up this initial difference; both groups' scientific research self-efficacy continued to increase through the end of the program.

Differences in Neuroscience Anxiety between Groups Under- and Well-Represented in STEM Fields

Again, data were collapsed across program models to test for differences across ethnic or racial groups over time. We used a 2X3 mixed-design ANOVA to test for effects of representation in the sciences (under-represented vs. well-represented) and time (pre-, mid-, or post-program).

Race was coded as a dummy variable to indicate whether individuals self-identified as belonging to a group that is demographically under-represented (including African-Descent/African-American, Hispanic/Latino/Latina) or well-represented (including Caucasian, Asian-Descent/Asian-American) in STEM fields. Individuals who did not identify a racial or ethnic category were not included in this analysis. Only the main effect of representation was significant, F(1, 95)=5.64, p<.05, with the represented group reporting lower neuroscience anxiety (M=12.07, SD=3.8) than the under-represented group (M=14.20, SD=5.5), regardless of time point.

Student Assessment of Their Own Learning Gains

We asked students to rate on a scale of one to five to what extent different program components helped their learning process. On average, program components were rated somewhere between "moderate help" and "much help" (M=3.43 SD=0.74). Of those components offered every year, three items were rated highest, with an average between "much help" and "very much help" (i.e. with means greater than 4): preparation of a research poster, presentation of that research poster, and preparation of a research report. Generally, participants valued the professional development workshops more than the orientation curriculum (M=3.80 SD=.75 vs. M=3.22 SD=.84), and 55 respondents in the CLM valued the program components more than 57 respondents in the AM (M=3.6 SD=.73 vs. M=3.28 SD=.73).
Qualitative Case Study Results

We selected four participants to include in a case study. They differed in the trajectory of their scientific research self-efficacy scores over the course of the program. As shown in Figure 1, the scientific research self-efficacy scores of two of these participants were low initially (compared to the group as a whole), but continued to increase through the midpoint to the end of the program. A third participant started the program with very high scientific research self-efficacy, which stayed high through the midpoint to the end, and a fourth participant began the program with a medium to low score which remained fairly steady through the midpoint to the end of the program. We call them Lisa, Helen, Sarah, and Jennifer, respectively. All were women, and all but Helen were in the AM group.

The interview data revealed some patterns not evident in our quantitative results. First, although these students experienced one of three very different trajectories with respect to scientific research self-efficacy, they all discussed perceived gains in self-efficacy and attributed them to their participation in research, and to overcoming challenges resulting from working independently or coping with unavailable mentors. Considering the qualitative data from these four students as a whole, it is clear that the students attributed changes in their own self-efficacy to mastery experiences, one of the four sources of self-efficacy. Even Jennifer and Sarah, whose self-efficacy did not change significantly during the program, identified mastery experiences in research design, specific neuroscience procedures, writing, leadership, and teamwork, as well as independent study. All four students also reported experiencing something like the process of enculturation into the scientific community. To bolster interpretation of the interview comments by each individual, we integrated results from our more quantitative measure of learning gains, as below. Lisa and Helen participated in the CLM and AM, respectively. They both began with moderately low scientific research self-efficacy scores and attributed them to a lack of research experience. Lisa and Helen both gained in scientific research self-efficacy during the program.

At the end of the program they both attributed their increased confidence to the greater understanding of the scientific process they gained through the program. Out of a highest possible average of 5, Lisa valued the program components (M=3.77), more than the average for her cohort (M=3.34), and Helen valued them even more than that (M=4.16).

Jennifer started with low-medium scientific research self-efficacy and remained stable throughout the program. Her assessment of the value of various program elements was also relatively low, averaging 2.70. Late in the program she admitted that she had previous research experience before beginning the program. Further review of her program application materials revealed her stated desire to pursue a medical degree for clinical practice rather a research degree, and this was confirmed by statements in her interview.

Sarah started with very high scientific research self-efficacy, and remained high throughout the program. As with Jennifer, she valued various program components less than her cohort, with her average SALG score of 3.03. She had graduated from college, was the oldest participant in her cohort, and had comparatively extensive research experience before the program began. She did mention gains in leadership, which indeed helped to validate her high leadership/teamwork self-efficacy scores at 49, 50, and 49 (out of 50) on the pre-, mid-, and post-program surveys, respectively.
Discussion of Results from the Sample Study

Our implementation of a scientific research self-efficacy instrument, alongside four other assessments of sociocognitive constructs, yielded valid and reliable data on student outcomes from an undergraduate summer research experience. The answer to our primary research question is generally affirmative: the research experience positively affected student outcomes, and resulted specifically in gains in both scientific research and leadership/teamwork self-efficacy, as well as declines in science and neuroscience anxiety. None of these measures varied by program format (AM vs. CLM), confirming our hypothesis that a collaborative learning experience produces outcomes as good as those of a traditional apprenticeship. Only scientific research self-efficacy varied by gender, and by mid-program women and men reported equal levels of confidence. Anxiety about neuroscience was higher among students from groups currently under-represented in science than in those from well-represented groups, however, a difference that was not resolved during the program. Students’ assessment of their own learning gains revealed they valued the process of creating and presenting scientific products the most. Finally, our qualitative data reveal that despite three distinct self-efficacy trajectories through our program, similar attributions of changes in self-efficacy to mastery experience can be made.

Both the AM and the CLM produced significant gains in scientific research and leadership/teamwork self-efficacy. We are continuing to track our students through their undergraduate programs, potential graduate programs, and beyond; we hope to confirm that these gains in domain-specific (Betz and Hackett, 1983; Hackett, 1985) and leadership/teamwork self-efficacy (Chemers et al., 2011) predict persistence and success in STEM careers. One of the ultimate goals in applying social cognitive career theory to STEM career development is to adopt a more proximate measure of desired outcomes to help evaluate programs such as ours. Although medical colleges maintain such a rich repository of data on progress through their degree programs that thorough evaluations of training programs can be made (e.g., Jaffe and Andriole, 2011), it is more difficult to maintain contact with program participants through other research-oriented STEM graduate degree programs, post-doctoral fellowships, and finally into careers themselves. Once the predictive validity of a given STEM domain-specific self-efficacy instrument is established, more efficient program evaluation can be conducted.

Interestingly, we found that women initially had significantly lower scientific research self-efficacy than men, but that they had caught up by the middle of the program. A follow-up study could be applied to determine whether these gains in self-efficacy, which we show here occur at different rates for undergraduate men and women, arise from different sources of self-efficacy, as has been demonstrated specifically for successful men and women in science (Zeldin et al., 2008). If we can identify different sources of self-efficacy in future years, it may also be the case that we find that the CLM provides more opportunity for vicarious experiences of mastery, to which successful women in science previously attributed their gains in self-efficacy more than did men. Our interview analysis of four women in the program provided no evidence of this difference in attribution, however, suggesting either that undergraduate women and successful professional women in science do not attribute self-efficacy gains to the same sources, or perhaps that our four cases are not representative of the larger population of women in science.
Although our initial analyses have not revealed any gender x program model interactions, nor any race x program model interactions, it may be the case that the CLM provides a more ideal fit for some groups of students, e.g. academically younger students with less research experience or lower scientific research self-efficacy. In a pilot program similar to the one just described, students had the opportunity to choose which program format they preferred; we found numerous cases in which those who preferred the CLM justified their choices by citing insecurity about their own research abilities and their desire to work on a team at the early stages of research experience (Frantz et al., 2006). Furthermore, women successful in science retrospectively cite more influence of social persuasion on their own science self-efficacy than men (Zeldin et al., 2008), suggesting that a collaborative experience could be more suitable for women. Finally, African American, Latina, Asian and Native American individuals are subject to unique cultural influences that affect sources of self-efficacy and thus persistence toward careers (Byars and Hackett, 1998). We must, therefore, perform more detailed analyses to determine whether under-represented minority students may benefit more from one or the other program format under certain circumstances. For example, a profile may emerge such that undergraduate, African American women with less previous research experience are most likely to benefit from the CLM. Similar detailed analyses may contribute to explanations for the lack of resolution of the higher anxiety levels among our study participants from underrepresented groups, compared to those in well-represented groups in the sciences. To further explore sources of self-efficacy in the current dataset, we look to our SALG, as well as our qualitative interview data.

Students' own ratings of what program components benefited them the most showed that research products, i.e. a research paper, a poster, and the presentation of the poster in a conference-like setting, were the most valued. These results align well with those of a recent, large-scale survey of undergraduates participating in summer undergraduate research experiences (Lopatto, 2007). The single program component that predicted a favorable overall evaluation of the experience was a final presentation. Furthermore, each of the products valued most highly by our program participants conforms to the best practices outlined in recent calls for inquiry-based writing as an essential component of science education in science lab courses (Moskovitz and Kellogg, 2011). Our research product program components are examples of "writing as professionalization," but in a context where students are able to attempt to relate their own results to the current literature, unlike in a traditional lab course setting. Application of social cognitive career theory here would suggest that the process of creating the poster and paper, and verbally explaining the poster in a conference-like setting constitute mastery experiences, which, if completed successfully, boost self-efficacy.

Although we did not inquire directly about sources of self-efficacy among program participants, the results of our four in-depth interviews all pointed to mastery experiences as critical to gains in research confidence. Of these four, the two who valued the program components the most (Lisa and Helen) were the ones who increased in scientific research self-efficacy. The student who entered the program with very high scientific research self-efficacy (Sarah) and maintained this level throughout the program, was already experienced in research, and may not have been challenged as much as the more novice students. Although all four students attributed any gains in confidence they experienced during the program to research participation, it is only the ones whose scientific research self-efficacy increased who valued highly the program components which constituted mastery experiences.
Our quantitative measures of science and neuroscience anxiety complement both the quantitative and qualitative self-efficacy results. Anxiety constitutes a physiological and affective state, and is one of the four sources of self-efficacy (Bandura, 1986, 1997). Our data show that although scientific research self-efficacy rose throughout the program, anxiety did not decline until our post-program measurement. Although scientific research self-efficacy at program entry was certainly not low (Figure 2), neither were students so comfortable that their science and neuroscience anxiety levels were at the minimum (9 and 8, respectively). Considering the qualitative attributions to mastery experiences, the increases in scientific research self-efficacy, and the value placed on scientific products that exemplify mastery of research skills, we interpret the whole profile as an indication that students were in the zone of “modest overconfidence” described by Lent et al. (1994), and were sufficiently challenged by the program tasks to gain some sense of mastery, thus boosting self-efficacy and lowering anxiety. It may be the case that the differences in the trajectories of scientific research self-efficacy in the qualitative case studies described here, and the variability in the quantitative measures of scientific research self-efficacy, can be explained in part by variability in the level of challenge for the tasks. Thus we detect no change in self-efficacy aligning with work either below the optimum (e.g. Sarah’s trajectory, Figure 1) or perhaps above it (e.g. Jennifer’s trajectory). Exploring this relationship further may allow us to optimize training programs for individual students, or for specific research settings.

The lack of differences in any measure between the AM and CLM groups as a whole is encouraging for the recruitment and retention of undergraduates in research career paths. Even at a research university, mentor supply is often the limiting factor in offering research experiences to undergraduates. Mentors for certain STEM disciplines may be even more scarce at Historically Black Colleges and Universities, Hispanic-Serving Institutions, and Tribal Colleges and Universities. The CLM may provide an alternative strategy to maximize scarce mentor resources in these environments. Future studies should include attempts to scale this strategy to these types of institutions.

Another priority is to perform a detailed analysis comparing the costs of each program. It would be ideal if the CLM could be scaled to institutions with fewer financial resources as well as fewer research faculty members. Our initial calculations, however, do not suggest that this is the case. Whereas most research mentors can be recruited to take on undergraduate trainees without financial compensation, we did not and would not expect to find individuals willing to teach in a CLM without pay. Moreover, purchase of research equipment, supplies, and animals (e.g. crayfish) must be carried out by a central program administration, and a teaching laboratory facility must be available nearly 24 hr per day, 7 days per week, if the CLM is to provide maximal research flexibility and authenticity to its participants.

Of final note, our participant population applied for the opportunity to conduct research, and thus ‘self-selected’ with a pre-existing interest in scientific research and motivation to spend a summer conducting research. Our conclusions regarding the effects of the program on their self-efficacy and future career intentions therefore relate to retention of talented individuals in science. STEM fields also need education research focused on recruitment of talented individuals into science. Elementary, middle, and high school students clearly are not too young to begin to develop academic and career interests, and thus should remain a focus for improvement of science education, perhaps in part through recruitment of very young learners into real research environments, to the extent appropriate with regard to safety and ethical conduct.
Our study overall provides an example of how to integrate scientific research self-efficacy and other sociocognitive constructs into education research in the context of undergraduate STEM experiences. Our study is challenged with several concerns typical of self-efficacy and undergraduate education research. Coupled with studies by other investigators asking similar questions, the study provides recommendations for how best to retain a diverse population of good students in pathways toward research in STEM fields.

**GENERAL CONCLUSIONS**

The aim of the present chapter is to help draw attention to the need for application of self-efficacy assessment and its associated social cognitive career theory, in the context of undergraduate scientific research. We have provided an example from our own research for implementation and analysis of scientific research self-efficacy and leadership/teamwork self-efficacy, both of which increased over the course of a 10-week summer research experience. We have also used self-efficacy and other constructs to compare outcomes from two different program formats, and provided data suggesting that a collaborative learning model is just as good as an apprenticeship model in this regard. These quantitative data were complemented by qualitative assessments in the form of individual interviews, which revealed some potential sources of the quantitatively measured self-efficacy, and thereby exemplified the benefits of integrating quantitative and qualitative assessment approaches. Although future work with our program participants (e.g. alumni surveys) and our existing data set will be required to confirm the predictive nature of scientific research self-efficacy for intent to persist in and actions taken to pursue research-related careers, numerous studies testing social cognitive career theory strongly suggest this will be the case. Thus, this chapter suggests that at least two different approaches to getting students involved in research produce short-term gains in self-efficacy that are likely predictive of long-term career progress in science. These data arrive at a critical time during which national and international councils recognize the need to improve science education in order to address concerns in globalization and technology.

**REFERENCES**


DOI: 10.1002/tea.20131


