Title
The STEM Lecture Hall: A Study of Effective Instructional Practices for Diverse Learners

Author
Reimer, Lynn Christine

Publication Date
2017

UNIVERSITY OF CALIFORNIA,
IRVINE

The STEM Lecture Hall:
A Study of Effective Instructional Practices for Diverse Learners

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Education

by

Lynn Christine Reimer

Dissertation Committee:
Professor Mark Warschauer, Chair
Distinguished Professor Jacquelynne S. Eccles
Assistant Professor Rachel B. Baker
Assistant Professor Di Xu

2017

DEDICATION

To the loving memory of James Reginald “Tony” Burke.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

ACKNOWLEDGMENTS

I have been blessed with wonderful support throughout this process. Mentors, colleagues, students, friends, and family all played integral roles. While it is not possible to thank everyone, there are several who deserve special mention.

My husband and best friend, Kevin, told me for years that I would obtain my Ph.D. He is my inspiration and strongest advocate. He stands at my side and together we face life's challenges and celebrate life's joys. My daughter, Naomi, taught me many things as a colleague and friend. My daughter, Danielle, was my prayer partner. My sister, Margaret, paved the way as first in our family to earn a Ph.D. She and my brother-in-law, Dino, walked me through the final few months, providing emotional support and perspective. My dear friend, Wanda, hiked with me for miles filled with laughter, tears, and deep conversation. My colleague and friend, Amanda, pushed me to be a better writer. And finally, my stepfather James Reginald Burke, who served our country at great personal cost, showed me what it means to be courageous and persevere!

I am thankful for my committee, providing abundant opportunities and support as I pursued this project with an ambitious timeline. Dr. Xu asked hard questions to ensure my methodology was sound. Dr. Baker gave of her time and experience, reminding me of the big picture. Dr. Eccles inspired me with her knowledge, wisdom, and scientific expertise. Dr. Warschauer, my chair, welcomed me into the program and provided me with countless opportunities. I appreciate each of you.

I am deeply grateful for the University of California, Irvine. Many departments and individuals across the institution gave time and resources to support this project. Lastly, I thank the National Science Foundation. I was fortunate to begin the research under grant #1256500, completing it as a graduate research fellow (#187698).

CURRICULUM VITAE

Lynn Christine Reimer

2017	Ph.D. in Education University of California, Irvine
2009	M.A. in Education Fresno Pacific University
2001	California Clear Credential California State University, Fresno
1991	B.A. in Biological Sciences; B.A. in German Language University of California, Davis

HONORS AND AWARDS

2014-2016	Graduate Research Fellow (#187698), National Science Foundation ($132,000)
2014	Graduate Student Promise Award, American Educational Research Association
1986-1991	Regents Scholar, University of California

PUBLICATIONS

Reimer, L., Nili, A., Nguyen, T., Domina, T., & Warschauer, M. (2016). Clickers in the wild: a campus-wide study of student response systems. In G. Weaver, W. Burgess, A. Childress, & L. Slakey (Eds.), *Transforming institutions: Undergraduate STEM education for the 21st century* (pp. 383-398). West Lafayette, IN: Purdue University Press.

Reimer, L., Schenke, K., Nguyen, T., O'Dowd, D., Domina, T., & Warschauer, M. (2016). Evaluating promising practices in STEM undergraduate lecture courses. *Russell Sage Foundation Journal for the Social Sciences: Higher Education Effectiveness, 2*, 212-231.

Reimer, K., & **Reimer, L.** (2016). STAMINA: Persistence and character in youth mentor partnerships. In G. Peterson, J. Van Slyke, M. Spezio, & K. Reimer (Eds.), *Habits in mind* (pp. 132-156). Boston, MA: Brill.

Reimer, L., & Reimer, K. (2015). Maturity is coherent: Structural and content-specific coherence in adolescent moral identity. *Journal of Positive Psychology, 10*, 543-552. DOI: 10.1080/17439760.2015.1015159

SELECTED PRESENTATIONS

Hanna, K., **Reimer, L.,** & Nili, A. (2017, April). *First-generation Latino Students' Success in Undergraduate STEM: The importance of instructor-student interactions.* Paper presented at the American Educational Research Association (AERA) annual meeting, Houston, TX.

Ho, P., Nili, A., **Reimer, L.** (2017, April). *Using Coh-Metrix to analyze writing cohesion in introductory courses with and without learning assistants.* Paper presented at the American Educational Research Association (AERA) annual meeting, Houston, TX.

Link, R., He, W., & **Reimer, L.** (2016). *Getting students back on track: Effects of flipping organic chemistry on student achievement study strategies and perceptions of instruction*. Paper presented at the Biennial Conference on Chemical Education, Greeley, CO.

Link, R., He, W., & **Reimer, L.** (2016). *Sapling learning online homework in organic chemistry lecture*. Paper presented at the Biennial Conference on Chemical Education, Greeley, CO.

Reimer, L., Nili, A., & Warschauer, M. (2016). *Project SPROUT: Documenting and investigating STEM instruction.* Paper presented at the National Science Foundation/American Association for the Advancement of Science Research and Practice Symposium: Envisioning the Future of Undergraduate STEM Education, Washington, DC.

Macias, M., Nili, A, **Reimer, L.**, & Warschauer, M. (2016). *Three unique approaches to introductory biology: A comparison of the attitudes and academic outcomes of underrepresented groups*. Paper presented at the American Educational Research Association (AERA) annual meeting, Washington, DC.

Rivas, M., **Reimer, L.**, Nili, A., & Warschauer, M. (2016) *Clicker usage in STEM classrooms at UCI: A two-year study.* Paper presented at the American Educational Research Association (AERA) annual meeting, Washington, DC.

Reimer, L., Nili, A., Nguyen, T., Schenke, K., Domina, T., & Warschauer, M. (2015) *Problem-solving pedagogies: Enhancing undergraduate STEM outcomes for underrepresented students.* Paper presented at the American Educational Research Association (AERA) annual meeting, Chicago, IL.

Reimer, L., Schenke, K., Nguyen, T., O'Dowd, D., Domina, T., & Warschauer, M. (2014) *Evaluating promising practices in STEM undergraduate lecture courses*. Paper presented at the Russell Sage Foundation Conference on Higher Education Effectiveness, New York, NY.

Reimer, L. (2014). *Clickers in the wild: A campus-wide study of student response systems*. Paper presented at Transforming Institutions: 21st century STEM Undergraduate Education Symposium, Indianapolis, IN.

Reimer, L., & Reimer, K.S. (2014). Maturity is coherent: Structural and content-specific coherence in adolescent moral identity. Paper presented at the American Educational Research Association (AERA) annual meeting, Philadelphia, PA.

UNIVERSITY TEACHING EXPERIENCE

2016- Director of Education Programs and Instructor for Project-Based Integrated Math/Science Certificate, University Extension, University of California, Merced

2016- Adjunct Instructor, Fresno Pacific University
EDUC 643 (Teaching Elementary Mathematics)

2015-2016 Instructor, Academic Research and Writing Workshop for First Year Education PhD Students, University of California, Irvine

2015 Teaching Assistant, EDUC 126 (Ethics in Education), University of California, Irvine

2014-2015 Lab Instructor, EDUC 288A (Educational, Social, and Behavioral Statistics) and EDUC 288B (Structural equation Modeling), University of California, Irvine

2013-2014 Educational Technology Coordinator, Azusa Pacific University

2012-2013 Adjunct Instructor, Fresno Pacific University

SERVICE

2016- University of California, Extension and Continuing Education Committee to Review Certificate Programs

2016- San Joaquin Valley University of California (SJVUC) Alumni
Fundraising and support for UC Merced Undergraduates

2015-2016 Lead Researcher, Learning Assistants Pilot, Office of the Vice Provost for Teaching and Learning, University of California, Irvine

2015-2016 STEM Lecture Forum, University of California System

2015-2016 Dean Search Committee, School of Education, University of California, Irvine

2014-2015 Higher Education Faculty Search Committee, School of Education, University of California, Irvine

2014-2016 First Generation Faculty Initiative, University of California, Irvine

2014-2016 Learning Spaces Advisory Committee, University of California, Irvine

2014-2016 Mentor, DECADE program, University of California, Irvine

UNDERGRADUATE HONORS THESES SUPERVISED

Gonzalez, S. (2016). *Academic and social impacts of a semester versus quarter system*. Unpublished honors thesis, University of California, Irvine.

Ho, P. (2015). *Student-centered pedagogy in Writing 39A classrooms for English as Second Language (ESL) learners*. Unpublished honors thesis, University of California, Irvine.

Macias, M. (2015). *Three unique approaches to introductory biology: A comparison of the attitudes and outcomes of underrepresented groups*. Unpublished honors thesis, University of California, Irvine.

Park, H. (2015). *Scaffolding and inductive method of instruction in undergraduate physics*. Unpublished honors thesis, University of California, Irvine.

Phan, J. (2015). *Examining STEM instructional methods and related effectiveness for disadvantaged Asian American groups.* Unpublished honors thesis, University of California, Irvine.

AFFILIATIONS AND MEMBERSHIPS

American Educational Research Association (AERA)

American Association for the Advancement of Science (AAAS)

ABSTRACT OF THE DISSERTATION

The STEM Lecture Hall:

A Two-year Study of Effective Instructional Practices for Diverse Learners

By

Lynn Christine Reimer

Doctor of Philosophy in Education

University of California, Irvine, 2017

Professor Mark Warschauer, Chair

First-generation, low-income, underrepresented minority (URM) and female undergraduates are matriculating into science, technology, engineering, and math (STEM) majors at unprecedented levels. However, a disproportionate number of these students end up graduating in non-STEM disciplines. Attrition rates have been observed to spike in conjunction with introductory STEM courses in chemistry, biology, and physics. These "gateway" courses tend to be housed in large, impersonal lecture halls. First-generation and URM students struggle in this environment, possibly because of instructors' reliance on lecture-based content delivery and rote memorization. Recent social psychological studies suggest the problem may be related to *cultural mismatch,* or misalignment between independent learning norms typical of American universities and interdependent learning expectancies for first-generation and URM students. Value-affirming and utility-value interventions yield impressive academic achievement gains for these students. These findings overlap with a second body of literature on culturally responsive instruction. Active gateway learning practices that emphasize interactive instruction, frequent assessment, and epistemological instruction can be successful because of their propensity to incorporate values affirming and utility-value techniques. The present study observed instruction for gateway STEM courses over a three-year period at the University of California, Irvine (N = 13,856 undergraduates in 168 courses). Exploratory polychoric factor analysis was used to identify latent variables for observational data on gateway STEM instructional practices. Variables were regressed on institutional student data. Practices implemented in large lecture halls fall into three general categories: *Faculty-Student Interaction, Epistemological Instruction,* and *Peer Interaction*. The present study found that *Faculty-Student Interaction* was negatively associated with student outcomes for female and first-generation students; and *Epistemological Instruction* was negatively associated with student outcomes for Hispanic students. More importantly, *Peer Interaction* was positively associated with student outcomes for female, first-generation, and Hispanic students. Study implications and limitations are discussed with reference to the research literature.

CHAPTER 1: INTRODUCTION AND BACKGROUND

Recent academic trajectories of American undergraduates in science, technology, engineering, and mathematics (STEM) majors indicate a troubling paradox. While more underrepresented minority (URM), first-generation, low-income, and female college students matriculate into STEM majors from high school, growth in degree completion is not keeping pace, resulting in continued inequalities (Baber, 2015; Bailey & Dynarski, 2011; Chen, 2013; Eagan, Hurtado, Figueroa, & Hughes, 2014; Tibbets, Harackiewicz, Priniski, & Canning, 2016). Of those intending to major in STEM, only forty percent obtain a degree within six years (Eagan et al., 2014). When this number is disaggregated by ethnicity, the disparity in degree completion becomes clearer. Of those who declare a major in STEM, fifty-two percent Asian, forty-three percent white, twenty-nine percent Hispanic, twenty-five percent Native American, and twenty-two percent Black students successfully graduate with a STEM degree. Figures are similar for first-generation and low-income students compared with continuing-generation students and students from middle and upper-income brackets (Chen, 2013). STEM degree completion rates for women reveal that Asian women are overrepresented relative to the general U.S. population, while Black, Native American, and Hispanic women are underrepresented (Ong, Wright, Espinosa, & Orfield, 2011). Gender, for certain students, is an additional disadvantage. Institutional responses emphasize recruitment and extra-curricular or co-curricular support (Baber, 2015; Bailey & Dynarsky, 2011; Tinto, 2012). When instructional practices have been examined, much of the research comes from evaluations of highly motivated and trained instructors in low-enrollment courses. Results are often linked to measures of student attitudes and motivations without including measures of persistence (Nielsen, 2011).

Statement of the Problem

URM attrition from undergraduate STEM programs may be associated with student experiences in "gateway" courses. Taken in the first year, these courses use lecture halls to deliver introductory content to large student populations. Gateway courses are a pivotal transition point from high school to university, perhaps even more so for first-generation, low-income, URM and female students. Females drop out of certain STEM majors (e.g., astronomy) at a higher rate than men; other STEM fields (e.g., biology) no longer show this disparity. Most public research universities rely heavily on the gateway STEM format. Many of these courses are premised on independent learning expectancies such as taking notes, completing solo homework assignments, and generating rote content on exams (Baillie & Fitzgerald, 2010). Instructional reform of gateway courses has the potential to reduce or eliminate achievement disparities experienced by underrepresented groups (Eddy & Hogan, 2014). As the classroom remains the nexus of institutional contact, identifying gateway instructional practices capable of nurturing persistence—taking additional STEM courses rather than transferring out of STEM—is crucial to improved matriculation into STEM careers (Engle & Tinto, 2008; Tinto, 2012).

Purpose of this Study

The present study considered gateway learning through: (a) broad description—this study included nearly all STEM gateway courses at a major public university over the course of three years, identifying instructional practices within the classroom context; (b) inclusion of persistence measures—this study included the likelihood of progressing to the next course, grade in that course and likelihood of staying in a STEM major; and (c) focus on URM students—this study included a sample of students to identify how instructional practices are related to student outcomes over two years, focusing primarily on URM, first-generation, low-income and female individuals.

Significance of the Study

The present study implemented a comprehensive evaluation of large-enrollment gateway STEM course instruction. Results may provide insight into the utility of different instructional practices in lower-division gateway STEM courses at a major university, paired with implications for student outcomes. The study offers guidelines for instructional reform in gateway STEM courses at large public research universities. Findings may stimulate critical pedagogical reflection for STEM instructors, facilitate collaborative discussion within STEM departments, and promote conversation pertinent to the institution's mission. Because of its scale, study findings may prove generalizable with relevance to gateway STEM instruction in similar American university contexts. The tools used for this study—observation protocols, surveys, and interview protocols—were selected to maximize potential for replication in other settings.

Research Questions

This study used systematic observations of instruction across a variety of STEM disciplines in lower division, large-enrollment lecture courses, linking these observations to student outcomes. The present study sought to answer the following questions:

1) What kinds of instructional practices are occurring in lower division STEM lecture courses?
2) Which practices tend to co-occur, and which practices are conceptually related?

3) How do STEM instructional practices relate to student outcomes and for which students?

To address the first two research questions, the present study considered three years of observational data describing instruction in lower division STEM lecture courses and identifying correlated practices. To address the third research question, the study linked observational with institutional data over two years. Analyses focused on students exposed to multiple instructional styles across different gateway STEM courses and documenting the extent to which within-student variation is associated with variations in course persistence and success. Two bodies of literature framed the study of instructional practices and student outcomes. The first literature corpus considered social-psychological explanations for URM difficulties in STEM, particularly through cultural mismatch. Social-psychological interventions were described in terms of student flourishing (values-affirmation) and curriculum content (utility-values) (Tibbets et al., 2016). In the second literature corpus, interventions associated with the social-psychological literature were linked to the STEM instructional literature through culturally responsive teaching and three variants of active gateway learning.

Limitations

While this study is potentially generalizable to other public research institutions, there are noteworthy limitations. First, observational data were focused on the extent to which instructors utilized strategies rather than how effectively instructors implemented those strategies. The distinction is significant. Instructors are typically afforded a great deal of professional autonomy at research intensive universities such as the project site (University of California, Irvine; UCI). Until recently, instructors received sparse pedagogical training. Gateway STEM instructors were given few cues regarding the effectiveness of instruction and fewer incentives to invest time and energy into their instruction. Brief, informal interviews with participating instructors suggested that many refrained from implementing promising practices highlighted by the present study, preferring instead to stick with "tried and true" instructional techniques. Meanwhile, other instructors struggled to implement highly touted best practices in an effective manner. Additional study is required to determine how to efficaciously promote and implement promising practices in gateway STEM courses.

Second, it must be noted that UCI is a selective institution. To access introductory chemistry, biology, and mathematics courses, UCI students scored above 600 on the mathematics portion of the SAT or completed a rigorous set of preparation courses. The institution's selectivity may have blunted the statistical significance of findings. If this is true, certain instructional practices may provide added gains among first-generation, low-income, URM and female students enrolled in less selective colleges and universities. Additional studies of gateway STEM instructional practices should include community colleges and unranked universities.

Study Background

The present study extended work by Reimer et al. (2016) in which forty gateway STEM courses were observed and analyzed to identify the relationship between instructional practices and student outcomes. The goals of that study were to: (a) develop a comprehensive matrix for measuring instructional practices in higher education, (b) establish baseline data, (c) obtain a synoptic view of current STEM instruction, (d) promote synergy across departments and schools to share evidence-based instructional research, and (e) obtain full-scale funding supporting evidence-based STEM education. The work was guided by Clotfelter, Ladd, and Vigdor (2007) and Xu, Hannaway, and Taylor (2011). Reimer et al. (2016) found that two operational categories of instructional methods, *assessment* and *interaction*, were associated with a positive long-term benefit for first-generation undergraduates without negative implications for the average student. The present study continued this work with funding provided by an NSF Graduate Research Fellowship. In 2015 institutional funds were leveraged, expanding observations to include a small number of non-STEM courses. For the sake of context, a brief review of the entire project precedes a detailed description of the present study.

Pilot. The project was initiated in fall of 2012. The research team secured human subjects approval and selected an observation protocol for STEM instruction called the *UTeach Observation Protocol* (UTOP; Walkington et al., 2012). The team chose nine gateway STEM courses and obtained instructor consent for observations during the spring term of 2013. The recruitment process, including instructor and student consent, is presented under *Sample and Recruitment*. The research team created a student pre/post survey, a teaching assistant post survey, and instructor interview protocol. Preliminary findings suggested the observation protocol did not adequately measure instruction in large-enrollment courses, was evaluative in nature (e.g., a top-down approach), and generated weak inter-rater reliability. Because the project aimed to document instructional practices using a bottom-up approach, the research team created an entirely new observation protocol (described in chapter 3) and re-coded the nine observations from video. Summer 2013 was used to refine surveys and interview protocol, affording opportunity for triangulation and validation of the newly developed observation protocol.

Year 1. During fall of 2013 and winter of 2014 the research team collected data (observations, surveys, and interviews). Instructor interviews provided continuous feedback for improvement along with validation of the observation protocol. Interviews offered qualitative insights for a mixed methods investigation of student response systems (i.e., clickers; Reimer, Nili, Nguyen, Domina, & Warschauer, 2016). Spring 2014 was spent on data analysis (see Reimer et al., 2016 for results of pilot and Year 1 data, referenced as the *initial study*). During this period, undergraduate research assistants (RAs) were heavily involved in data collection and analysis.

Year 2. Observations were conducted for academic year (AY) 2014-2015, albeit without surveys of students or TAs as these were becoming an obstacle to instructor recruitment. Instructors were not interviewed in AY 2014-2015 because of insufficient financial resources. During the summer of 2015, internal funding was used to hire a project manager. This individual improved project functionality in anticipation of a third year of data collection with an expanded team of undergraduate RAs.

Year 3. Course observations were conducted for AY 2015-2016. A total of 69 instructor interviews for year 2 participants were completed; qualitative analyses were conducted on Year 1 and 2 instructor interview data. With strong RA support, interviews were conducted with undergraduate STEM students enrolled in large-enrollment courses. Additionally, Year 3 included observations and student surveys from several non-STEM courses.

The present study considered instructional practices within the classroom context of STEM gateway courses. All three years of observational data were used to analyze instructional practices and create conceptual composites (Research Questions 1 and 2). Institutional and observational data from Years 2 and 3 were used to determine the relationship between the instructional practices and student outcomes, particularly for underrepresented groups in terms of social-psychological explanation and intervention, coupled with culturally responsive teaching and active gateway learning practices (Research Question 3; RQ 3). Years 2 and 3 offer six consecutive terms of data, unlike the analysis conducted on the Pilot and Year 1 data, which had limited data ($N = 40$ courses). Years 2 and 3 included additional data collection providing a more comprehensive and congruent data set aligned with study research questions ($N = 168$ courses).

CHAPTER 2: CONCEPTUAL FRAMEWORK

The present study described instructional practices in gateway STEM courses and examined how these practices were related to student outcomes. Certain gateway STEM instructional practices were anticipated to support academic success for first-generation, low-income, URM and female students. Two parallel literatures addressed the core problem. The first literature corpus came from social/educational psychological researchers seeking to (a) explain the problem, and (b) design psychologically appropriate interventions to resolve the problem. Cultural mismatch was a prominent feature in social-psychological explanations of why first-generation, low-income, URM and female students struggle to persist in undergraduate STEM programs. Social-psychological interventions emphasized the importance of student flourishing (values-affirmation) and curriculum content (utility-value; Tibbets et al., 2016). Social-psychological explanations and interventions were used to provide an empirical frame for discussion of gateway STEM instructional practices.

The second literature corpus was concerned with undergraduate STEM instructional practices. Aspects of social-psychological explanation and intervention were evident in this companion literature. Culturally responsive teaching was used to introduce active gateway learning, and is understood as interactive engagement, assessment and feedback, and epistemology and metacognition. These three categories were used in the initial study (Reimer et al., 2016) and are reflective of the literature on instructional practices in STEM undergraduate lecture courses (e.g., Nielsen, 2011). Extant gaps in the research literature became the basis for study research questions and hypotheses.

Social/Educational-Psychological Explanations

Learning is a cultural process that unfolds in an environment shaped by identities and attributes (Hand, Penuel, & Gutierrez, 2012; Nasir, Rosebery, Warren, & Lee, 2006). Several influential educators have suggested that certain groups of students are disadvantaged because of conflicts between implicit expectations for autonomy in American higher education and cultural identity (Inzlicht & Schmader, 2012; Steele, 2011; Tinto, 2012). This argument is built upon twin assumptions, including (a) presence of autonomy norms (independent learning expectancies) in American higher education and (b) influence of autonomy norms on URM and first-generation student identity and academic achievement. To evaluate the relevance of these assumptions, Stephens, Fryberg, Markus, Johnson, and Covarrubias (2012) proposed *cultural mismatch,* an adaptation of social identity threat whereby "student social class backgrounds [are understood to] shape the cultural models of self they bring with them to college settings" (p. 1180). Experimental methods were subsequently deployed to evaluate the construct across multiple university sites, including private and public schools (Stephens et al., 2012). Cultural mismatch attempted to explain the academic achievement gap in terms of institutional responsiveness to the cultural identities of its students, rather than as student deficits subject to remediation (Engle & Tinto, 2008; Soria & Stebelton, 2012; Tinto, 2012).

Cultural mismatch combines social-psychological literatures on identity threat and culturally embedded self-understanding. Social identity threat is associated with negative stereotyping (dismissive recognition) of the identified group. Students experiencing social identity threat do not perform as well as students who are comparatively unencumbered (Inzlicht & Schmader, 2011; Steele, 2010). According to the literature on culturally embedded self-understanding, learning motivation is informed by identity consolidation and maturation

(Markus, 2008). An implication is that students experiencing an undergraduate STEM course as self-relevant are more efficacious, engaging with content and performing better on tasks, assignments, and tests (Fulmer et al., 2010; Markus, 2008). It follows that student perception of learning as self-relevant is dependent upon the presence or absence of external recognition of the member group. In the event recognition occurs as negative stereotyping or positive recognition should be absent from the learning experience, URM and first-generation students are at risk for diminished academic success (Stephens et al., 2012).

The Stephens et al. (2012) study involved multiple investigations using different samples and methods. Study 1 addressed the hypothesized pervasiveness of autonomy norms in American higher education. Two samples were recruited. The first involved senior administrators from top national universities and colleges (U.S. News and World Report, 2010). Participating administrators ($N = 254$) completed a survey listing institutional expectations of students, with half the items representing expectations of autonomy and half representing interdependence. A matching task required administrators to complete sentences associated with independent or interdependent student norms. Results indicated dominance of independent university norms and expectations in top academic institutions $\chi^2(1, N = 254) = 47.6$, $p = .000$), with no observed variation across administrator characteristics such as gender or ethnicity, nor institutional type (i.e., public or private). A second sample recruited administrators from unranked universities and colleges with diverse student bodies (U.S. News and World Report, 2010). Participants ($N = 119$) completed the same survey administered to the first sample. Results again indicated independent university norms and expectancies ($\chi^2(1, N = 119) = 17.0$, $p = .000$), although lower tier administrators (69%) were less likely to identify expectations for student autonomy than top tier administrators (84%). Overall, American university and college administrators consistently upheld norms and expectations for student autonomy as opposed to interdependence (Stephens et al., 2012).

Studies 2-4 addressed questions of whether independent norms eventuate into cultural mismatch for first-generation URM undergraduates (Stephens et al., 2012). Study 2 considered independent or interdependent models of self that were operative for incoming students (N = 1,528) at a large, selective public university. Students were categorized as first-generation if neither parent had completed a baccalaureate degree (N = 245), with continuing-generation students identified as having one or more parents with a college degree (N = 1,179). Fifty-seven percent of the first-generation cohort identified as URM, including Hispanic (37%), African American (13%), and American Indian (7%) students. Both cohorts completed surveys tapping autonomy (e.g., thinking independently) or interdependence (e.g., helping family, contributing to community) as a core feature of self. Students were followed for two years. Academic success was assessed through cumulative student GPA after Year 1 and Year 2, respectively. Relative to continuing-generation students, first-generation students linked self with fewer independent norms ($\beta = .13$), $t(1311) = 4.6$, $p = .000$ and more interdependent norms ($\beta = -.17$), $t(1311) = -6.5$, $p = .000$. Regressions assessed whether independent or interdependent models of self were linked to academic success at the conclusion of Years 1 and 2. Consistent with cultural mismatch predictions, interdependence negatively predicted cumulative GPA at the end of Year 1 ($\beta = -.11$), $t(1298) = -3.9$, $p = .000$, with autonomy weakly predicting GPA gains ($\beta = .04$), $t(1298) = 1.8$, $p = .07$. This finding became more pronounced at the end of Year 2, with interdependence negatively predicting GPA ($\beta = -.07$), $t(1291) = -3.1$, $p = .002$, and autonomy positively predicting GPA ($\beta = .05$), $t(1291) = 2.5$, $p = .01$. The analysis did not parse URM students from the first-generation cohort. Results confirmed the need for follow-on studies to ascertain whether first-generation students with interdependent learning expectancies experienced an "unseen disadvantage" relative to continuing-generation students (Stephens et al., 2012).

Studies 3 and 4 extended this work by manipulating university orientation materials to test whether norms of autonomy or interdependence impacted academic performance and perceptions of task difficulty (Stephens et al., 2012). Entering undergraduates ($N = 147$) were recruited from a less selective public university and more selective private university. Nearly half of students ($N = 67$) identified as first-generation. Two welcome letters were developed as primes, ostensibly from the university president. The first letter emphasized autonomy norms in four themes: (a) exploring personal interests, (b) articulating individual opinions and ideas, (c) developing a personal intellectual trajectory, and (d) involvement in personal research. The second letter emphasized interdependence with four themes: (a) learning as part of a community, (b) forming relationships with faculty and other students, (c) working closely with others, and (d) involvement in collaborative research (Stephens et al., 2012). In Study 3, incoming students were randomly assigned a welcome letter. Immediately after reading the letter, students completed a verbal reasoning task. As a hedge against stereotype threat (dismissing recognition), students were told the task assessed learning styles. Data were analyzed using a 2 X 2 analysis of covariance (ANCOVA; first-generation status X welcome letter) controlling for ethnicity and high school GPA. Findings demonstrated that first-generation students performed worse than continuing-generation students following receipt of the independent welcome letter $F(1, 38) = 6.1, p = .02$. The disparity vanished following introduction of the interdependence welcome letter $F(1, 42) = 0.1, p = .71$. Within-group analysis revealed that first-generation students performed significantly better in the interdependent condition than the independent condition $F(1, 38) = 4.2$, $p = .049$, whereas continuing-generation students performed the same in both conditions (Stephens et al., 2012).

Study 4 explored whether the university welcome letter influenced how students construed task difficulty (Stephens et al., 2012). After reading the welcome letter, students were asked to recreate an image (e.g., cat, dog, horse) using tangram puzzle pieces. Students were invited to complete as many images as possible within a 10-minute period, followed by completion of 7-point Likert scales to rate task difficulty. A 2 X 2 X 2 ANCOVA (first-generation status X welcome letter X university type) was conducted, controlling for ethnicity and high school GPA. First-generation students perceived the task as significantly more difficult than continuing-generation students when primed with the independent welcome letter $F(1, 133) = 12.4$, $p = .001$. Within group analysis showed that first-generation students construed the task as easier in the interdependent condition $F(1, 61) = 15.4$, $p = .000$, whereas continuing-generation students perceived the task equivalently in both conditions $F(1, 74) = 1.6$, $p = .21$. Taken together, Studies 2-4 provided support for cultural mismatch between institutional norms of autonomy and interdependent norms typical of first-generation students. The authors noted that group disparities in academic success are responsive to changes in university norms (Stephens et al., 2012). Transition from norms of autonomy toward interdependence may yield significant gains in academic success. In the classroom, recognition of group differences along with respectful affirmation of interdependent motives may go a considerable way in effectively promoting academic achievement for first-generation and URM undergraduates.

Phillips, Stephens, and Townshend (in press) conducted a follow-up study comparing first and continuing-generation students at an elite private institution, investigating whether models of self, sense of fit, and college outcomes differed on social class. The study addressed the problem that while college access has increased for first-generation URM students, degree completion has not kept pace (Baber, 2015; Bailey & Dynarski, 2011; Chen, 2013; Eagan et al., 2014). The findings debunked assumptions that (negative) ramifications of social class and cultural differences diminish with time spent in the university environment (Arnett, 2000; Gay, 2010; Soria & Stebelton, 2012; Tinto, 2012). By graduation, first-generation students did not endorse independent motives (e.g., *I want to become a more independent thinker*), continuing to favor interdependence (e.g., *I want to give back to my community*; Phillips et al., in press). In addition, student experiences of institutional fit revealed a growing gap by graduation. Graduating, continuing-generation students had higher GPAs compared with freshman year GPA whereas first-generation student GPAs remained unchanged. Perhaps most significant, first-generation students reported lower status at graduation when compared with continuing-generation counterparts. Status was assessed by having participants use a vertical ladder image to rank themselves compared with: (a) seniors at their institution, (b) people in society, and (c) peers at home.

Cultural mismatch underscores the importance of recognition as a potential equalizer, capable of affirming individual and group identity. URM and first-generation undergraduates may arrive at university with interdependent expectancies for learning and development. These students come as persons embedded within pre-existing cultural communities with local funds of knowledge. Rather than assume students must conform to independent learning norms, it seems more equitable to design gateway STEM courses in ways that capitalize on independent and interdependent expectations. The work on cultural mismatch suggests a win-win outcome, whereby first-generation and URM students experience gains in academic success without cost to other groups. To be sure, this would require considerably more of gateway STEM instructors.

Courses would need redevelopment to celebrate group differences through skillful deployment of interdependent norms, furnishing opportunities for quality interpersonal interactions and recognition. Such a course "reboot" could elevate independent learning norms without dismissing, misrecognizing, or displacing cultural groups.

Social/Educational-Psychological Interventions

The concern regarding cultural mismatch has been addressed through social/educational-psychological interventions, as a means of improving STEM persistence among first-generation and URM students and thus frames the importance of what occurs within the lecture hall setting. In a review essay, Tibbets et al. (2016) identify two categories of relevant intervention. The first category is aimed at cultural identity and student well-being through interventions characterized by *values-affirmation*. Values-affirmation attempts to reduce student anxiety associated with misrecognition and ameliorate personal struggles to find belonging (Lewis & Sekaquaptewa, 2016; Spencer, Logel, & Davies, 2016). The second category involves interventions given to student engagement with academic content and curriculum. Such *utility-value* interventions are designed to improve student motivation through goal identification and achievement (Eccles, 2009; Harackiewicz & Hulleman, 2010; Hulleman & Harackiewicz, 2009). Each category scaffolds different psychological interventions capable of improving URM and first-generation student achievement and persistence (Tibbets et al., 2016).

Several studies were designed to test the efficacy of these intervention categories. The values-affirming category was anchored by a project involving a gateway undergraduate biology course, with intervention designed to address the social achievement gap for first-generation students (Harackiewicz et al., 2014). The sample included first-generation students ($N = 154$) derived from a cohort of $N = 798$. The intervention included two values-affirming writing exercises and two control writing exercises. Values-affirming content domains included art, family relationships, politics, personal independence, learning, athletics, co-curricular involvement, music, career interests, spirituality, and humor. The writing exercises were complemented by self-attribution scales for the same domains. Relative to continuing-generation students, the achievement gap for first-generation individuals was reduced by fifty percent. First-generation students receiving values-affirmation manifested a GPA improvement of a quarter letter grade. The values-affirmation intervention improved overall semester GPA for first-generation students. First-generation students in the intervention condition were twenty percent more likely to take the next course in the STEM program sequence (proxy for persistence) when compared with controls (Harackiewicz et al., 2014).

A follow-up study tracked the same biology students over a three-year period (Tibbets et al., 2016). Values-affirmation effects persisted for first-generation students beyond three years. Informed by cultural mismatch, the authors considered autonomy and interdependence norms through the values-affirmation intervention. First-generation students writing about interdependent norms earned significantly higher grades than controls in the academic term during which the intervention was administered. The improved grades persisted over subsequent terms. The self-referencing effect was replicated in a controlled environment, with first-generation students improving scores on a math test following an invitation to write about personal interdependence (Tibbets et al., 2016). It was not clear whether the effect was related to diminished performance anxiety, developmental factors related to identity maturation and differentiation, or growing expertise associated with independent learning expectations in STEM programs.

Utility-value interventions were premised upon *expectancy value* (Eccles & Wigfield, 2002). Four types of expectancy value are detailed, including attainment (self-importance of task), cost (personal consequence), intrinsic (task enjoyment), and utility (usefulness of task related to goals; Eccles, 2009). Utility-value interventions with first-generation undergraduates focused on (a) student expectations for success, and (b) extent to which academic tasks are valued (Canning & Harackiewicz, 2015; Harackiewicz & Hulleman, 2010). An experimental study of success expectations with a small undergraduate sample provided support for utility-value suppositions (Durik, Shechter, Noh, Rozek, & Harackiewicz, 2014). In the first study (N = 62), utility-value information embedded in a math task resulted in improved performance for students with elevated expectations of success. In the second study (N = 148), administration of an "expectancy boost" for success resulted in improved task engagement for students with relatively low opinions of their academic capabilities (Durik et al., 2014). In sum, success expectancies explain noteworthy variance regarding student engagement with academic content and curriculum.

A second utility-value study considered student valuations of academic tasks in a gateway STEM course (Harackiewicz, Canning, Tibbets, Priniski, & Hyde, 2016). Pertinent to the present study, the sample (N = 1,040) included a cohort of first-generation URM students (N = 67) with continuing-generation URM students (N = 126) as comparators. Experimenters asked students to write a brief (500 word) essay on the personal significance of course content. The assignment was graded as homework and replicated twice during the term. The utility-value intervention was effective in improved academic achievement for all participating students. First-generation URM students experienced the greatest intervention benefit; the achievement gap for first-generation URM students was reduced by sixty-one percent. The gap lowered from

0.84 grade points to 0.51 grade points (a third of a grade; Harackiewicz et al., 2016). In step with cultural mismatch, first-generation URM students fused utility-value with interdependence goals, linking course content with perceived obligations to communities of origin.

Harackiewicz et al. (2016) noted that values-affirmation and utility-value interventions are neatly aligned with a companion literature on active learning in STEM disciplines. The active learning literature was initially premised on a well-known concept from P-12 education. *Culturally responsive instruction* details a positive plan for classroom instruction with diverse learners (Gay, 2010; Martell, 2013). Culturally responsive educators prize "cultural knowledge, prior experiences, frames of reference, and performance styles of ethnically diverse students to make learning encounters more relevant and effective" (Gay, 2010, p. 31). A classic example is the work of Jaime Escalante, who famously integrated Hispanic values, work habits, learning styles, and experiences into his advanced placement calculus course at Garfield High School in Los Angeles (Escalante & Dirmann, 1990). Decades before social-psychological research on URM academic success, Escalante established care and student concern as instructional priorities, framing pathways to academic success in terms of *ganas*, or "desire" in Spanish. Academic *ganas* tapped the interdependent, collaborative ethos of Hispanic students, culminating in a collective effort to master the advanced placement calculus test. While this example is somewhat removed from gateway undergraduate STEM courses, several elements are applicable. Equitable, culturally responsive instruction promotes group recognition and interpersonal interaction through: (a) instructor authenticity, (b) instructor care for students, and (c) interdependent tasks, assignments, and learning assessments.

The culturally responsive value of instructor authenticity was evident in a qualitative study of Hispanic high school students (Irizarry, 2007). Face-to-face instructor authenticity was robustly linked to group recognition for Hispanics and African-Americans. Authentic instructor interest and concern regarding the cultural specifics of student experience proved contagious, with students reciprocating across multiple instructor-student and student-student interactions. Not surprisingly, culturally responsive instruction depended on instructor expressions of care and concern for student success. Along these lines, Grantham, Robinson, and Chapman (2015) analyzed *thank-a-teacher* submissions from undergraduates at a less selective, diverse public university. Findings supported the importance of care and concern communicated through quality interpersonal interactions. Appropriately caring instructors created a trusting environment where worldview reflection could safely and effectively unfold in the public domain. The Grantham et al. (2015) findings mirrored key points identified by the National Survey of Student Engagement (NSSE). In this regard, care and concern involved thoughtful instructor feedback delivered in a timely manner to students. Caring instructors had high expectations of students. Caring interactions between instructors and students were characterized by freedom to discuss personal issues and career plans. Not surprisingly, these are key learning expectancies associated with interdependent cultures (Sanchez & Gunawardena, 1998; Smith & Ayers, 2006).

Social-psychological interventions capable of reducing the achievement gap are designed as extra-curricular, co-curricular, or supplemental to classroom instruction. The gateway classroom remains a key focus for instructional practices reaching students regardless of background (Harackiewicz et al., 2014; Stephens et al., 2014). As many American college students live and work off-campus, institutional efforts to increase persistence should emphasize the classroom, "the one place on campus, perhaps the only place, where the great majority of students meet the faculty and one another and engage in formal learning activities" (Tinto, 2012;

p. 5). Factors inhibiting academic performance such as cultural mismatch may be reduced when the classroom context does a better job of alignment with student norms and provides opportunity to safely build conceptual knowledge and scientific skills (Hill, Corbett, & St. Rose, 2010; Steele, Spencer, & Aronson, 2002; Stephens et al., 2012). A 2012 report by the President's Council of Advisors on Science and Technology (PCAST) points to the importance of instructional reform:

> Traditional teaching methods have trained many STEM professionals, including most of the current STEM workforce. But a large and growing body of research indicates that STEM education can be substantially improved through a diversification of teaching methods. These data show that evidence-based teaching methods are more effective in reaching all students—especially the "underrepresented majority"—the women and members of minority groups who now constitute approximately 70% of college students. (p. i)

Instructional practices in gateway STEM courses may enhance persistence, particularly noteworthy as the greatest attrition from STEM majors occurs during the first year (PCAST, 2012). Because gateway STEM courses at American public universities continue to be taught face-to-face in lectures halls, research must address the question of how gateway courses can be improved and made culturally responsive, especially for the growing number of URM, first-generation, low-income, and female students (Eagan et al., 2014).

A recent survey found that a majority (54%) of undergraduate STEM instructors at public universities hold fast to a didactic, independent learning lecture format (Smith, Vinson, Smith, Lewin, & Stetzer, 2014). To change this, attention shifted to active learning (Nielsen, 2011; Owens et al., 2017). Traditional lecture may be augmented to include a variety of complementary instructional practices favorable for student success (Hora & Ferrare, 2014). Along these lines, Freeman et al. (2014) conducted a meta-analysis of 225 studies: 158 included course outcomes (exams and concept inventories) and 67 measured the odds of completing the course with a passing grade. Results suggested that active learning is associated with one half standard deviation (*SD*) increase in student performance. Students in traditional lecture courses were 1.5 times more likely to fail the course. Active learning was understood "[to] engage students in the process of learning through activities and/or discussion in class, as opposed to passively listening to an expert. It emphasizes higher-order thinking and often involves group work" (Freeman et al., 2014: 8413-8414). Results indicated further study inclusive of the relationship between instructional practices and groups of students, especially in large-enrollment gateway courses.

Description and evaluation of gateway instructional practices, particularly in courses with average enrollments of 350 or more, may identify practices capable of improving URM and first-generation persistence and remediating achievement gaps (Ifill et al., 2016; Malcolm & Feder, 2016; Mervis, 2010; NAE, 2005). Accordingly, Reimer et al. (2016) observed instruction in gateway STEM courses at a public research university to determine prevalence and type of instructional practices, assessing implications for student success. The sample included 4,801 undergraduates; fifty-eight percent identified as first-generation and twenty-six percent identified as URM. Half the courses included instructional practices identified as effective in the Freeman et al. (2014) meta-analysis. Using a student fixed effects model, results indicated that students earned higher grades in courses where instructors utilized a combination of practices that fell into three categories: interactive instruction, frequent assessment, and explicit epistemological instruction. Interactive instruction and frequent assessment were positively associated with persistence—namely, the probability of taking the next course in the series and subsequent course grade—for first-generation college students with no detrimental impact on the other students (Reimer et al., 2016). It may be expedient to consider studies that specifically address URM and first-generation students in undergraduate STEM courses in terms of *active gateway learning*.

Active Gateway Learning: Interactive Instruction

In the gateway STEM context, interactive instruction can include instructor-student exchanges premised on induction, informal peer discussion with clicker questions, structured group collaboration on problems, or extended learning with flipped techniques (Freeman et al., 2014; Reimer et al., 2016; Stage & Kinzie, 2009). Several studies found that high quality instructor-student and student-student interactions were significantly associated with URM and first-generation success (conceptualized as course completion and grade in course; Figlio, Rush, & Yin, 2013; Jaggars & Xu, 2016; Ke, Chavez, Causarano, & Causarano, 2011). Norms of interdependence associated with interactive engagement practices were potentially able to attenuate cultural mismatch with positive implications for student efficacy and motivation.

A large, mixed-method study considered student attitudes relevant to interactive engagement practices in gateway STEM courses (Gasiewski, Eagan, Garcia, Hurtado, & Chang, 2012). Fifteen universities were evaluated, including 73 gateway courses with an overall sample of 2,873 undergraduates. Students from a variety of schools were surveyed, including two Hispanic serving institutions and three historically black colleges and universities. Gateway interactive practices included group sessions, clickers, and web-based pedagogy with a proprietary STEM curriculum. Students reported higher academic engagement in courses that included interactive instructional practices and strategies. Students perceived gateway STEM courses to be less internally competitive when interactive engagement practices were present, reflecting introduction of interdependence norms associated with classroom activities (Gasiewski et al., 2012). Considering earlier findings of student discomfort in STEM courses, interactive engagement in gateway courses may promote interdependence supportive of enhanced student attitudes and academic outcomes (Brint, Cantwell, & Hanneman, 2008).

A recent study at a Hispanic serving university used flipped classroom modules in gateway chemistry to enhance interactive engagement (Eichler & Peeples, 2016). The undergraduate STEM population was diverse, with thirty-six percent of participating students identified as first-generation and URM. The flipped modules made class time available for interactive practices. The study compared an identical course taught concurrently by the same instructor. The treatment condition included use of interactive practices for twenty-eight percent of the class time with students watching lecture modules as homework. The control condition used a traditional lecture format with interactive practices employed for thirteen percent of class time. Findings indicated that students in the interactive course condition earned significantly higher final course grades than students in the control condition (Eichler & Peeples, 2016). Although the study did not evaluate persistence, the diverse student population and positive outcomes suggest that active gateway learning with interactive engagement hedges against cultural mismatch and attrition.

A companion study considered identical gateway chemistry course sections, including an interactive condition and a lecture-based control condition (Ryan & Reid, 2015). Students in the interactive condition watched video lectures at home, making class time available for active learning through collective activities and practices. Outcomes were considered in terms of factors associated with diminished course persistence: Ds, Fs, and withdrawals (DFW). Students were administered a chemistry readiness pretest and subsequently placed into low, medium, and high achieving groups for comparison between course conditions. Interactive engagement significantly improved exam scores for the low scoring group of undergraduates with weak chemistry skills (0.43; $p < .05$). Perhaps more compelling, the DFW rate for gateway chemistry was reduced by fifty-six percent in the interactive condition relative to the control condition with no differences for the medium and high achieving groups (Ryan & Reid, 2015). Courses emphasizing interactive engagement potentially mediate social class and achievement gaps, especially during the first year in STEM when attrition rates are highest. The present study links interactive practices with student outcomes to identify which kinds of practices are related to student success for first-generation, low-income, URM and female students.

Active Gateway Learning: Frequent Assessment

Inclusion of frequent, formative assessment in gateway STEM courses is made easier with the advent of student devices such as clickers. Conceptualized as a way to increase active learning, these devices are used to provide instructors with real-time data pertinent to learning, providing opportunity to modify content delivery and improve effectiveness (Han & Finkelstein, 2013). Formative assessment practices may communicate instructor learning expectancies and norms. Several studies have shown that frequent, low-stakes (i.e., formative) testing improves student persistence, efficacy, and self-regulated learning (Black, 2013; Carpenter, 2012; Clark, 2012; McDaniel, Anderson, Dervish, & Morrisette, 2007; Pastötter & Bäuml, 2014). Clickers offer a non-threatening way for students to be assessed frequently (Blasco-Arcas, Buil, Hernandez-Ortega, & Sese, 2012; Wolter et al., 2011). However, the use of clickers in formative assessment may be a cultural mismatch for those students with interdependent learning norms. One study considered student success and persistence when clickers were incorporated across forty-three gateway STEM courses, each with an average course enrollment over 300 (Reimer et al., 2016). More than half of students in the study were first-generation and more than twenty-five percent were URM. Results were mixed: while clickers were associated with higher course grades on average, they were not associated with improved STEM persistence (i.e., successful completion of the next course in the series), nor were clickers associated with improved academic success (i.e., improved GPA). Unexpectedly, clickers were negatively associated with *quality* of gateway peer activities such as peer interaction (group work) and problem solving (Reimer et al., 2016).

A key concern emerges from this discussion. It is possible that instructors may deploy clickers in ways that reinforce autonomy norms associated with traditional didactic lecture in gateway STEM education while believing they are reforming instruction for greater inclusivity. If this is true, URM and first-generation students may experience gateway STEM learning with clickers as cultural mismatch, increased anxiety associated with a lack of belonging. Reimer et al. (2016) found that clickers were significantly and negatively associated with academic success for first-generation Hispanics, which may be indicative of cultural mismatch. Future studies should consider instructor expectancies (e.g., norms of autonomy or interdependence) associated with gateway STEM clicker deployment, with attention given to directional flow of feedback. For URM and first-generation students transitioning from high school to university, anonymity associated with clickers may conflict with learning expectancies that favor interdependence. Clickers do not necessarily facilitate relationship with instructors and peers. Active gateway learning incorporating assessment and feedback media such as clickers should be used in culturally responsive ways that affirm and support personally and relationally interdependent learning expectancies associated with URM and first-generation undergraduates.

Smith et al. (2009) focused on quality of peer interaction in gateway courses using clickers at a large public research university with a diverse student body ($N = 350$). Students were asked to respond to a question autonomously, followed by invitation to discuss answers with peers and revote on the same question. Baseline data were obtained for students using independent pathways, eliminating interdependent exchange or discussion in the classroom. Results indicated that students with opportunity for discussion demonstrated improved conceptual understanding, even when others in the group didn't know the right answer. Interactive engagement practices based on norms of interdependence positively impacted efficacy and motivation (Smith et al., 2009). Regrettably, the study did not differentiate between first or continuing-generation students. Frequent, formative assessment promoting interdependence may improve STEM persistence. The present study aims to identify what practices that co-occur with the use of formative assessment (e.g., clickers) are related to positive student outcomes for first-generation, low-income, URM and female students.

Active Gateway Learning: Epistemological Instruction

A benefit of frequent assessment relates to enhanced student capacity to frame scientific investigation (epistemology) and reflect critically on individual reasoning (metacognition). Epistemological instruction promotes facility with critical analysis, helping the student separate fact from opinion (Goldman, 1986; Grant, 2008; Hahn & Oaksford, 2012). Best practices are indicated through instructor reference to content from previous courses and instructor commitment to help students reframe scientific preconceptions and assumptions on sound empirical principles. Successful gateway STEM instructors model problem-solving techniques during lecture, make internal reasoning processes explicit, and provide learners with on-the-fly analysis of concepts (Bassok & Novick, 2012). This might involve disclosure of intellectual history for concepts under consideration and personalized to the instructor's research program (DeLuca & Lari, 2013; Pace & Middendorf, 2004). Gateway instructor concern for epistemological instruction should reflect interdependent norms, expectancies, and instructional practices, particularly as most scientific research occurs collaboratively.

A rugged approach to effective application of epistemology and metacognition involves *problem-based learning* (PBL; Hung, Jonassen, & Liu 2008). In large STEM courses, PBL includes instructional practices such as "worked" examples involving instructor modeling of deductive reasoning and problem-solving techniques. This provides opportunity for students to practice with each other, requiring individuals to think and behave as scientists. Examples and student collaborative practices are used to reference content from other courses, make connections to real-world situations, and prime the class for success on exams. The efficacy of PBL is documented through smaller, upper division courses in STEM (Dochy, Segers, Van den Bossche, & Gijbels, 2003; Eagan et al., 2014). Academic outcomes associated with PBL include enhanced student motivation and course satisfaction, content engagement, test performance, content retention and recall, and mastery of conceptual reasoning and problem solving skills (Chaplin, 2009; Newman, 2005; Strobel & van Barneveld, 2009). Studies of PBL with gateway STEM courses are few. Therefore, less is known about how epistemology and metacognition embedded in gateway learning may impact URM and first-generation students.

One study considered PBL through weekly instructional practices in a gateway biology course at a large public university (Haak, HilleRisLambers, Pitre, & Freeman, 2011). The course was selected based on achievement gaps for economically disadvantaged students, most of whom were URM. The achievement gap for the course was considered highest for any course at that university. Instructors modeled epistemic and metacognitive processes through worked examples, real time data analysis, and activities evoking higher order cognitive expertise. Analyses were confined to six offerings of the same course taught by one instructor. Two course offerings were traditional lecture, two were moderately interactive with incorporation of clickers, and two were highly structured around PBL principles. Students were identified through an opportunity program at the university. The structured course configuration emphasizing PBL principles had a significant impact, reducing the achievement gap from 0.80 to 0.44 grade points, a forty-five percent drop when compared with moderately interactive and traditional didactic course conditions (Haak et al., 2011, p. 1215). The study suggests that involving epistemological and metacognitive practices using interdependent norms may contribute to the persistence of disadvantaged students in STEM.

The literature underscores the benefit of including evidenced-based practices in STEM gateway courses as a means for increasing student persistence and degree attainment. However, most studies were limited in one or more of the following ways: (a) study was limited to a single, small enrollment course—only a few studies target large-enrollment courses; (b) study considered multiple sections of one course—few studies included multiple courses; (c) study was confined to a single discipline—very few studies included multiple disciplines; (d) study considered student self-report rather than achievement data—rarely did a study include persistence data; and finally (e) study made no distinction between first-generation and underrepresented minority students. Dividing instructional practices into three broad categories—interactive engagement, assessment and feedback, and epistemology and metacognition—suggests potential benefit for URM and first-generation students in terms of active gateway learning. Because the classroom is a location guaranteed to reach all students, inclusion of instructional practices capable of enhancing the learning environment and fostering cultural match would presumably reduce the achievement gap (Tinto, 2012). It is not a single practice, but rather a combination of interactive and collaborative practices that maximally benefit students, particularly those who are first-generation and/or struggling to maintain high grades in their STEM courses. A suite of interdependent teaching strategies within a gateway STEM course may change the culture of the classroom along with improvements to academic achievement and persistence. Interactive, collaborative practices have the potential to reduce cultural mismatch and achievement gap within gateway STEM courses, with positive implications for persistence.

The present study extended the work of the initial study (Reimer et al., 2016) and addressed the following research questions:

1) What kinds of instructional practices are occurring in lower division STEM lecture courses?

2) Which practices tend to co-occur, and which practices are conceptually related?

3) How do the instructional practices relate to student outcomes and for which students?

For the first question, three years of observational data were used (*N* = 190 courses). Two additional years of observational data expands on the initial study (*N* = 40 courses) to create a more comprehensive picture of instructional practices in lower division STEM courses. These three years of data were also used to identify which practices co-occur, and which practices are conceptually related. Neither of these analyses were included in the initial study. For the third question, two years of observational data was linked with student data to validate and provide a more detailed description of the relationship between instructional practices and student outcomes compared with the initial study. With these questions in mind, I expected to validate findings that courses include a variety of instructional practices beyond traditional lecture, such as use of clickers and group work (RQ 1). Additionally, I anticipated that specific instructional practices would co-occur, such as explicit instructor reference to exam content and open discussion of common student mistakes in a manner supportive of student success (RQ 2). Lastly, I expected to find certain practices were associated with positive student outcomes, conferring benefits for URM, first-generation, low-income students, and female students (RQ 3).

CHAPTER 3: METHOD

The study was conducted at the University of California, Irvine (UCI), a public research university operating on the quarter system. The study was approved by the UCI Institutional Review Board under protocol 2012-9277 and 2015-2522. Additionally, we obtained a letter from the university FERPA analyst confirming procedures for storing and analyzing student data. Data collection began in spring of 2013. The project was completed in spring of 2016. Data were obtained from the schools of Biological Sciences, Engineering, Information and Computer Sciences, and Physical Sciences. Taken together, these schools enroll half the undergraduate population and engage seventy-five percent of UCI undergraduates in STEM disciplines. As of 2015, the six-year degree completion rates for those students entering the four STEM schools were as follows: fifty-seven percent for Biological Sciences, sixty-one percent for Engineering, seventy-two percent for Information and Computer Sciences, and fifty-one percent for Physical Sciences. More than eighty-eight percent of these students earn a baccalaureate degree within six years. Those who switch out of STEM transfer to other major disciplines and the remaining twelve percent do not matriculate. Non-STEM majors in the other schools at UCI have higher completion rates. For the sake of this study, majors classified as STEM followed the National Science Foundation's Classification of Instructional Programs (CIP). The student population at UCI has changed significantly in the past ten years. In 2006, twenty-nine percent of these students were first-generation, twenty-four percent were Pell grant eligible, and thirteen percent were URM. In 2015, fifty-five percent were first-generation and forty percent were Pell grant eligible (i.e., low-income), with thirty percent identifying as URM.

Sample

Courses. We identified gateway courses in the schools of Biological Sciences, Engineering, Information and Computer Sciences, and Physical Sciences that were (a) prerequisites for courses in one or more STEM majors, (b) offered in multiple sections over the year, (c) enrolled 200 or more students,[1] and (d) are categorized as lower division, typically taken in the first or second year of study. Introductory chemistry (Physical Sciences) was the most popular STEM gateway course, required for nearly all STEM majors. Table 1 presents a breakdown of courses by school observed over three years to address the first and second research questions. Table 2 presents a breakdown of courses from Years 2 and 3 relevant to the third research question (the initial study analyzed the pilot and Year 1 data).

We recruited instructors from gateway STEM courses in four schools. The overall participation rate was eighty-six percent for three years (see Table 3). Eleven gateway instructors declined to participate; seven initially participated and later declined; three declined and subsequently chose to participate.[2] We focused on courses above 200 enrolled students to prioritize instructional practices in large lecture halls. The 200 cutoff is an acknowledged threshold for gateway lecture courses in public research universities (Freeman et al., 2014). The three-year observational sample used for the study included 190 course sections. Analyses linking observational data to student outcomes were conducted with data from Year 2 and Year 3

[1] We observed a few courses under 200 by dean or departmental request, typically associated with new course design or implementation of a novel learning technology. These are not included in the study, with the exception of Engr7B, which had a capacity for 225, yet only 149 were enrolled.

[2] On occasion, an instructor did not respond, and therefore was not observed.

and included duplicate sections ($N = 168$ courses), given that the pilot and Year 1 data were analyzed in the initial study.

--INSERT TABLES 1-3 ABOUT HERE--

Instructors. Instructors were active participants in the study. Once courses were identified, instructors were contacted prior to the start of the term and invited to participate. A follow-up email was sent a week later. Associate deans and instructors assisted recruitment by contacting colleagues and encouraging participation. Due to a limited number of observers, we prioritized a new instructor or course over courses that were repeats (e.g., same course taught by the previously observed instructor). The study included instructors in various positions: graduate students, researchers, lecturers, lecturers with potential for security of employment (LPSOE), lecturers with security of employment (LSOE), assistant professors, associate professors, professors, and professor emeriti. Participation included two video-recorded observations of the entire class session, along with an optional interview after the quarter was completed. We did not observe in either the first or last week of the term, nor on the day of a midterm exam or guest speaker, to better capture a typical day of instruction. In a 10-week quarter, the first observation occurred during weeks 2-5 and the second during weeks 6-9, intentionally scheduling at least two weeks between observations for the same instructor and randomizing within these parameters the observation date and observers. As mentioned in the Introduction, we rarely observed a lab or discussion section, as these were beyond the scope and resources of the study. An overview of instructor characteristics is provided in Table 4.

--INSERT TABLE 4 ABOUT HERE--

Detailed study information was provided to each instructor at the beginning of the quarter. After final exams, we provided instructors with a copy of all completed observation protocols (the Simple Protocol for Observing Undergraduate Teaching; SPROUT, described in the Instruments section below). We sent instructors dichotomous SPROUT scores pertinent to their course, department, and other courses observed that term. During this process instructors were invited to provide feedback (Malcom & Feder, 2016; Reimer et al., 2016). Considerable variation was observed for gateway instructional practices across courses and between sections of the same course taught by different instructors. However, when an instructor(s) taught duplicate sections, analyses determined these were nearly identical.[3]

Students. Students were passive participants. While some students were recruited for interviews, the present study focused on Year 2 and Year 3 data from the Office of Institutional Research (OIR) to address the third research question. The student analysis sample was diverse. Fifty percent were first-generation, thirty-six percent were Pell grant eligible (i.e. low-income), twenty-seven percent were URM, and fifty-three percent were female[4]. Some students fell into more than one category: twenty-eight percent were first-generation and low-income, seventeen percent were first-generation and Hispanic, and twenty-eight percent were first-generation and female. Nearly all students were in the first or second year of their STEM degree program. Because many students were enrolled in more than one observed course, it was possible that a single student provided more than one observation. The analysis sample included 44,981 student-by-course observations involving 13,856 discrete students. Table 5 shows the

[3] We tested this assumption by observing duplicate course sections. These observations returned a high degree of consistency, with instructional practices correlating at the 0.95 level.

[4] In certain courses, females were in the minority (e.g., engineering and information and computer science courses).

percentage of students that took multiple courses. The average number of courses taken by a single student was four (*Mean* = 4.13; *SD* = 2.83). Student characteristics are presented in Table 6: (a) all students in observed courses and similar courses in which the instructor declined to participate (*N* = 19,134); (b) all students that entered as freshman (*N* = 16,679); (c) the analysis sample (*N* = 13,856); and (d) the student fixed effects sample (*N* = 11,291). The analysis sample includes only those students who entered UCI as a freshman[5] and were in one or more observed courses. Thirteen students were dropped from the analysis sample as their SAT scores were missing at random. The analysis sample does not differ significantly from the full sample of students. The analysis sample was built from institutional and observational data. Appendix B provides the flow chart of how I systematically built and cleaned the data set.

--INSERT TABLES 5 & 6 ABOUT HERE--

Procedure

Instruments. The pilot study used an observation protocol that was poorly suited to gateway STEM courses. Consequently, we developed the SPROUT (Simple Protocol for Observing Undergraduate Teaching) with four goals in mind: (a) capture instructor and student behaviors in gateway lecture courses, (b) provide opportunities for instructors to reflect on their own instruction, (c) stimulate departmental conversations related to evidence-based instruction, and (d) link instructional practices to student outcomes. Observation protocols influencing the SPROUT included: U-Teach Observation Protocol (UTOP; Walkington et al., 2012); the Reformed Teaching Observation Protocol (RTOP; Sawada et al., 2002); and the Teaching Dimensions Observation Protocol (TDOP; Hora & Ferrare, 2014). The UTOP helped create

[5] Transfer students (eight percent) were excluded from this analysis. We conducted separate analyses on transfer students, with attention to their unique situation.

student-centered variables such as instructor monitoring of student understanding. The RTOP focused on math and science teaching such that it contributed instructional practices specific to those fields. The TDOP provided categories of instructional practices centered on higher education in STEM. Both UTOP and RTOP were designed for K-12 classrooms, but provided guidance on capturing qualitative data; the TDOP, while designed for higher education, captures only quantitative data. By incorporating salient features from these protocols, the SPROUT aimed to provide quantitative (dichotomous and count variables) and qualitative (descriptive) data on gateway instructional practices (see Appendix C for a copy of the protocol).

The SPROUT provided a multifaceted measure of active gateway learning suitable for the present study. To triangulate data, the research team developed a protocol to interview instructors at the end of the course. Interviews were semi-structured and included questions aligned with data fields captured by the SPROUT. In addition to instructor interviews, student surveys and teaching assistant surveys were administered in Year 1 as cross-validation. Protocol revisions were made through Year 1 to improve the clarity of the instructions. A few instructors provided feedback during the revision process. While there was initial concern about undergraduate RAs conducting observations, Cohen's kappa (κ) coefficient of reliability was acceptable at 0.75, with an interrater agreement rate of eighty percent.

Observations. Course observations were conducted twice per quarter. In the event instructors co-taught a gateway course, we conducted one observation for each instructor to capture all teaching practices. The observation protocol included a video recording and observer field notes which were subsequently transferred to the SPROUT. Observers were trained prior to conducting observations. Certain items on the SPROUT (e.g., clicker questions) required confirmation to determine whether an assignment was graded or merely awarded participation points. Observers checked items against course syllabi and supporting materials provided on the course website to verify content related to grading policies.

Institutional data. Following final exams, student data were obtained from OIR including demographic data, previous academic data (e.g., high school GPA, scores on STEM AP exams, SAT Math and Verbal scores), and current academic data (e.g., course enrollments and grades—for observed courses and courses taken in subsequent terms). These data allowed us to track student progress toward STEM degrees.

Plan of Analyses

In the initial study, Reimer et al. (2016) created three measures using instructional variables from the SPROUT and course syllabi to analyze instructional practices with student outcomes from pilot and Year 1 data. These were composites capturing the degree to which instructors engaged in three categories of instructional practice during lectures, discussions and other co-curricular instruction directly linked to the course. Unlike the initial study, the current study focuses exclusively on what occurs during the large-enrollment lecture, "the one place on campus, perhaps the only place, where the great majority of students meet the faculty and one another and engage in formal learning activities" (Tinto, 2012; p. 5). The current study extended the initial study by addressing the following limitations: (a) collection of two new years of data providing year-over-year student and observational data; (b) addition of descriptive data pertinent to instructors and courses; (c) exploratory factor analysis of SPROUT variables made possible with a larger data set; (d) regression models linking instructional practices with student outcomes accounting for co-occurring strategies; and (e) evaluation of differential effects for underrepresented groups.

To address the research questions, I tabulated descriptive statistics for courses and instructors for all three years of data (RQs 1 and 2) and for Years 2 and 3 (RQ 3). Descriptive statistics were used to identify which gateway instructional practices were commonly implemented by department. This was followed with latent variable analysis to identify which instructional practices co-occurred and how co-occurring practices were conceptually related. Given the binary nature of the SPROUT variables, polychoric correlation and exploratory factor analysis (EFA) were indicated to identify latent composites of instructional variables (Flora & Curran, 2014; Gorsuch, 1970; Hair, Black, Babin, & Anderson, 2009). This process called for principal axis factoring (PAF) with promax rotation, a preferred (oblique) option for EFA using the following equation:

$$F_{OLS} = \frac{1}{2} tr[(S - \Sigma)^2] = \sum_{i} \sum_{j} + (s_{ij} - \sigma_{ij})^2$$

where s_{ij} and σ_{ij} are elements of observed sample correlation matrix and implied correlation matrix, respectively (de Winter & Doudou, 2012). PAF makes no assumption regarding the type of error and minimizes the unweighted sum of ordinary least squares of the residual matrix. Estimates of the loadings are obtained by minimizing the discrepancy between the observed sample correlation matrix S and the model-implied (filled) correlation matrix Σ (de Winter & Doudou, 2012). This provided the best fitting model of co-occurring practices.

Following identification of co-occurring instructional practices and conceptual relationships, the final analytic step called for regression to consider association between instructional practices and measures of student success: (a) final grade in observed course on a four-point scale, (b) odds of taking a follow-on course, (c) grade in the follow-on course, and (d) odds of staying in their STEM major. Because some students take courses at other institutions, especially during the summer, any course in the same discipline at a higher level was coded as a follow-on course, with priority given to the next course in the series (e.g., Chem1B after Chem1A). I linked observational with student-level data to provide a non-experimental, population-based evaluation of the relationship between instructional practices and student outcomes. I took advantage of variation across sections of the same course and across different courses to measure the relationship between active gateway learning practices and student outcomes within a given student.

The plan of analyses took advantage of the fact that many students were enrolled in multiple courses under observation (average of 4.13 courses; *SD* = 2.83). Repeated observations made it possible to estimate the extent to which exposure to promising instructional practices influenced academic achievement in that course and subsequent course, net of observed and unobserved student characteristics: "under a strict exogeneity assumption on the explanatory variables, the fixed effects estimator is unbiased," (Wooldridge, 2013, p. 482)[6]. For the initial study, only half of students comprised the student fixed effects sample. However, with two years of student institutional data linked to observational data for six consecutive quarters (two academic years), the student fixed effects sample included eighty-two percent of students with nearly identical demographics (see Table 6). The models took the following general form:

$$Y_{ij} = \beta_0 + \beta_1 Instruction_i + \beta_2 Covariates + Course\beta_{3j} + \sum \beta_4 Student_i + \varepsilon$$

with Y_{ij} the outcome of interest. Four outcomes were indicated: (a) final grade in observed course on a four-point scale, (b) odds of taking a follow-on course, (c) grade in the follow-on course, and (d) odds of staying in their STEM major. *Instruction* was operationalized as a

[6] See Clotfelter et al., 2007 and Xu et al., 2011 for analyses using a similar design in public high school settings.

composite, guided by EFA, in some models and represented a vector of individual instructional practices in other models. *Course* represented a vector of course fixed effects controlling for aspects of content and instruction that did not vary across sections of the same course[7]. *Student* in this model was a vector of student fixed effects, controlling for student characteristics across courses including race, gender, economic and academic background, as well as invariant student characteristics such as interest and motivation.

To determine outcomes for underrepresented groups, a series of interaction terms were required (instruction * first-generation, instruction * low-income, instruction * ethnicity, instruction * female, and instruction * cumulative GPA at time of the observed course). These interactions were used to estimate the extent to which association between instruction and student outcomes was different for certain groups such as first-generation and/or low-income, compared with continuing-generation peers. Models using the full analysis sample clustered standard errors at the course section level. Models using the student fixed effects sample included course fixed effects with robust standard errors to account for any heteroscedasticity or within student correlation. All variables were standardized for ease of comparison.

[7] A vector of teacher (instructor) fixed effects was originally included in the model to control for instructor characteristics including race, gender, and position at the university, as well as invariant characteristics such as interest and motivation for teaching. However, is was omitted due to collinearity with the course fixed effects.

CHAPTER 4: RESULTS

Results are organized according to the plan of analyses from Chapter 3. Descriptive statistics are presented for the purpose of identifying which instructional practices were most common (RQ 1). Next, co-occurring instructional practices are considered with latent variable analysis using exploratory factor analysis (EFA; RQ 2). Finally, results from this process are used to guide the development of regression models examining the relationship between instructional practices and student achievement (RQ 3).

Descriptive Statistics

Using three years of observational data, I ran descriptive statistics to identify instructional practices in lower division STEM lecture courses from the schools of Biological Sciences, Engineering, Information and Computer Sciences and Physical Sciences (RQ 1). Table 7 provides an overview of instructional practices by department for three years of observational data, subsequently used for the latent variable analysis as described below and Table 8 provides the same information for Years 2 and 3 used in the regression analyses and with overall trends described below. The tables are organized by instructional composites. Engineering and Information and Computer Sciences (ICS) courses were combined into one department, as many of the courses are cross-listed. Math and statistics were combined for two reasons: (a) because they share similar content, and (b) because most students in the study were only required to take one of each.

--INSERT TABLES 7 & 8 ABOUT HERE—

Faculty-student interactions. The first composite included nine items. Instructors asking display questions (e.g., fact-based or yes/no; 65% - 79%) and checking if students had questions (59% - 79%) were found most often across all departments. Instructor open-ended questions were observed less often (34% - 13%), most common in biology. An instructor pausing and checking for questions was associated with a 0.52 standard deviation increase in students asking a conceptual question (49% - 74%). Similarly, when the instructor paused and checked for questions, this practice was associated with a 0.44 standard deviation increase in back-and-forth interaction between the instructor and a student. This back-and-forth element is one measure used in the literature to categorize a class as interactive (e.g., Freeman et al., 2014; Smith et al., 2014; Walkington et al., 2012). This practice occurred in nearly half the course across all departments, with the exception of general chemistry: biology (54%), general chemistry (28%), organic chemistry (44%), engineering/ICS (40%), math/statistics (47%) and physics (56%). Factor analysis confirmed strong correlations between an instructor asking if there were questions and students responding with questions.

Differences were noted in terms of how instructors engaged the whole class with a question, referred to as *formative assessment*. Student response systems such as iClickers were used more often in biology (74%), math/statistics (53%) and physics (52%). By way of comparison, instructors in general chemistry, organic chemistry, and engineering/ICS were less likely to use these systems, instead asking questions for which they required the whole class to respond in a different manner (e.g., raising their hand; 28%, 31%, and 29% respectively).

Epistemological instruction. Returning to the frequency of gateway instructional practices by department, the second composite, *Epistemological Instruction*, consisted of six items, most of which occur in over half of the courses observed. Instructors solving problems occurred in most courses, but most often in math/statistics (100%), general chemistry (96%), organic chemistry (94%), and physics (91%), compared with engineering/ICS (76%) and biology

(69%). Instructors used a variety of methods to solve problems in real-time, such as pen and paper projected with a document camera, a stylus on computer screen, or markers on the dry erase boards. Rarely did an instructor model problem solving by working through pre-made slides. Factor analysis confirmed that handwritten visuals and problem solving co-occurred, along with instructor notation of prior course content. An instructor including real-world examples occurred in over half of the courses: biology (80%); general chemistry (83%); organic chemistry (59%), engineering and ICS (55%), math and statistics (49%), and physics (72%).

Peer interaction. This composite consisted of three items. Group work was observed most often in biology (51%) and physics (26%) and rarely observed in chemistry (8%) or engineering/ICS (5%). Desk work did not occur often, except in general chemistry (19%). Peer discussion was frequently encouraged in the context of clicker questions.

Latent Variable Analysis

To identify co-occurring instructional practices, Years 1-3 of observational data from SPROUT were combined to create a polychoric correlation matrix for dichotomous data (Flora & Curran, 2004; Stata FAQ, 2016). Certain variables were returned missing on the matrix due to collinearity and/or lack of variation. I used the polychoric correlation matrix to run exploratory factor analysis (EFA) using principal axis factoring (PAF) with promax rotation (Gorsuch, 1970), which allowed me to identify co-occurring (e.g., latent composite) instructional practices in the SPROUT dataset. The initial extraction considered the variance explained by each factor. In situations lacking *a priori* assumptions regarding the optimal number of factors or percentage variance to merit inclusion, inferential decisions can be made by examining changes in variance between factors (Hair et al., 2009). A large change in explained variance between factors suggests a cut point. The initial EFA yielded four factors explaining seventy-seven percent of the total variance of the correlation matrix. However, weak to moderate change in variance was noted between Factors 3 and 4, with values as follows: Factor 1 = 26%, Factor 2 = 20%, Factor 3 = 18%, and Factor 4 = 15%. The change in explained variance between Factors 3 and 4 suggested that Factor 4 should be discarded. Item loadings were screened for the factors, eliminating items scoring less than .45 or items that loaded across more than one factor (Gorsuch, 1970; Hair et al., 2009). Factor 4 retained only two items after the screening procedure, which also supported discarding it. The remaining three factor solution explained sixty-four percent of the total variance.

Latent factors. Five items were associated with the first factor. Factor items were characterized by engagement between faculty and students ($\alpha = .58$), but not *between* students. Factor 1 identified which of these instructional practices co-occurred most often. For example, an instructor might check if there are questions (item 1) and students respond by asking questions about content, course, or lecture hall environment (items 2, 4, 5). Item 3 identifies moments when a question resulted in a conversation with a single student. Because factor 1 did not include any interaction between peers, I chose to divide *Interactive Instruction* into two separate composites, *Faculty-Student Interaction* and *Peer Interaction*, to identify any implications for independent versus interdependent learning in relationship with student outcomes. All composites are described following the latent variable analysis.

Three items were affiliated with the second factor ($\alpha = .49$). Instructors solving problems in front of the class (item 1), using handwritten visuals (item 2) and mentioning prior course content (item 3). The second factor informed the composite *Epistemological Instruction* described below. Three items were associated with the third factor ($\alpha = .38$), but were not conceptually related. Premade visuals (item 1) often co-occurred with the use of a pointer (item

2) and students completing desk work (item 3). Alpha coefficients of reliability were marginal for Factor 2 and inadmissible for Factor 3. Table 9 details items and factor loadings for the three-factor solution. Of note, while *assessment* was used as a composite in the initial study, nothing similar resulted from the factor analysis. This is most likely the result of differences between the two studies: the initial study including co-curricular practices from coded syllabi, whereas this study focused exclusively on what occurred in the lecture hall.[8] While certain practices co-occurred, they were not always conceptually related. Table 10 lists the factors by student demographics for Years 2 and 3 to identify how observable student characteristics were correlated with each factor. In summary, instructors used a variety of practices associated with active gateway learning, most were instructor-driven (e.g., an instructor asked questions or mentioned exam content), and some were student-driven (e.g., peer discussion). Co-occurrence of the practices (e.g., an instructor frequently used handwritten visuals when solving problems in front of the class) and EFA provided guideposts for development of instructional composites.

--INSERT TABLE 10 ABOUT HERE--

Instructional composites. I identified two distinct forms of interaction and thus created separate composites: *Faculty-Student Interaction* (nine variables; α = .70) and *Peer Interaction* (three variables; α = .59). Not all nine variables for *Faculty-Student Interaction* co-occurred but these were conceptually similar, capturing the extent to which faculty and students interacted with one another. The composite *Peer Interaction* captured the extent to which students interacted with one another. Research has shown that first-generation, low-income, URM and

---

[8] The initial study had three composites: Interaction included four SPROUT items and one from coded syllabi; epistemology only included SPROUT items; assessment included four SPROUT items and four coded syllabi items. *Assessment* from the initial study was dropped. As the goal was to document instructional practices, observations were not conducted during midterm or final exam days and quizzes tended to be given during discussion sections.

female students may hesitate to interact with an instructor in a gateway lecture hall, but benefit when afforded opportunities to interact with one another (Lewin et al., 2016; Smith et al., 2009; Stephens et al., 2012). The composite *Epistemological Instruction* included six variables that were conceptually similar to factor 2 ($\alpha = .50$). These practices captured the extent to which faculty modeled scientific thinking and reasoning, practices believed to promote deduction, problem solving, and causality necessary for STEM careers (Dunbar & Klahr, 2009). The composites were used in the regression analyses (RQ 3). Variables classified as *Instructional Tools* are included in Table 8 for descriptive purposes (RQ 1). Items comprising Instructional Tools were included when regression analyses were run on individual items (RQ 3). The three composites were standardized after the individual variables (dichotomous) were added together as the number of variables in each composite varied. The variables were added to capture conceptually similar practices. For example, within *Epistemological Instruction*, one instructor might *mention prior content* while another *mentions big ideas*, both intended to help students make connections.

I considered the relationship between instructional practices and measures of student success in terms of: (a) final grade in observed course on a four-point scale (e.g., a B+ would earn 3.7 points), (b) odds of taking a follow-on course, (c) grade in the follow-on course, and (d) odds of staying in the STEM major. By linking observational data with student-level data, I took advantage of instructional variations across sections of the same course—taught by different instructors—for a non-experimental, population-based evaluation on extent to which certain instructional practices supported improved student outcomes. Table 11 lists the instructional composites by student demographics for Years 2 and 3 to identify how observable student characteristics were correlated with each composite.

--INSERT TABLE 11 ABOUT HERE--

A student fixed effects model was developed to address selection bias. After considering student factors that predict exposure to certain instructional practices, I analyzed the relationship between instructional practices and student achievement (RQ 3). Student fixed effects models were used to assess the impact of instructional practices on grade in observed and subsequent course. Because student characteristics such as race and family background did not vary across course observations (i.e., time-invariant), those controls were only included as interactions to allow for analyses of differential effects. The parameter of interest in the models, *instruction*, estimated the extent to which an instructional practice (or composite of practices) in a particular course predicted student achievement in that course (along with subsequent course) when compared with other courses taken by that student. To be included in the student fixed effects model, students must have taken at least two observed courses; the average number of courses taken was four (*SD* = 2.80). As a result, eighty-one percent of the students contributed to the student fixed effects sample (*N* = 11,291 students; 35,998 observations) and were nearly identical—in terms of demographics and previous academics—to the analysis sample (*N* = 13,856 students; 44,981 observations), students in all similar courses that entered as freshman (*N* = 16,679 students; 78,556 observations) and students in similar courses, including transfer students (*N* = 19,134 students; 85,518 observations). Refer to Table 6, mentioned earlier, for comparisons of these four groups of students. The last two groups of students included courses offered at the same time as observed courses where the instructor declined to participate.

Regression Analyses Using Composites

Variables were standardized to facilitate ease of comparison for all continuous variables in the entire data set (*N* = 19,134 students; 85,518 observations).[9] The analysis sample (Years 2 and 3) included courses in four schools that were prerequisites for other courses and occurred in a series of two or three courses. Models clustered standard errors at the course section level, "vce(cluster coursecode),"[5] yielding 121 clusters when analyzing the full sample and 101 clusters for follow-on course. The code used for these clusters conservatively marks the same course section taught by the same instructor during the same quarter as identical whether it occurred in Year 2 or Year 3. Models using the student fixed effects sample used "vce(robust)."[10] Models including cumulative GPA at the time of the observed course yielded higher R^2 values and reduced the negative and significant coefficients for first-generation, Hispanic and low-income students. Cumulative GPA at time of course is the strongest predictor of final grade in that course. Including this covariate eliminated bias in identifying differential effects for first-generation, low-income, URM and female students' relationships with instructional practices. Time-invariant characteristics and interaction terms were included in the student fixed effects sample to analyze differential effects. After running regression analyses using the active gateway learning composites, I repeated the analysis with the individual instructional variables, discussed in greater detail separately below.

--INSERT TABLE 12 ABOUT HERE--

[9] For example, grades on a four-point scale centered on 0 as the mean, could then be compared with SAT scores on a scale of 800, once also centered on 0 as the mean.

[10] vce(robust) or vce(cluster coursecode) causes the Huber/White/sandwich variance-covariance (VCE) estimator to be calculated for the co-efficients of the regressions (Wooldridge, 2013).

Grade in current course. Models 1-3 (see Table 12) regressed grade in current course on the instructional composites. Models 1 and 2 used the full analysis sample, while Model 3 used the student fixed effects sample[11]. Interaction terms were included in Models 2 and 3. By including all three instructional composites, the models provided a more accurate representation of what occurs in the lecture hall, because observed courses included practices from each composite. This eliminates omitted variable bias. In Model 3, grade in course is positively associated with *Epistemological Instruction* (0.040; $p < .001$; $R^2 = .31$), while there is no association between grade in course and *Faculty-Student Interaction* or *Peer Interaction* for the reference group (i.e., Asian males).

Differential effects. To assess differential effects for underrepresented groups, a series of interaction terms were added (composite*female; composite * first-generation; composite*low-income; composite*Hispanic). The interaction terms estimated the extent to which association between instruction and student outcomes was different for certain groups of students, such as first-generation, when compared with peers who have greater familiarity with higher education. There was a positive association between grade in course and *Peer Interaction* for Hispanics (0.017; $p < .05$; $R^2 = .31$) and first-generation students (0.030; $p < .01$; $R^2 = .31$). However, a negative association was noted for low-income students (-0.020; $p < .05$; $R^2 = .31$). These three groups overlap; however, three-way interactions were non-significant. First-generation students experienced a negative association with *Faculty-Student Interaction* (-0.026; $p < .01$; $R^2 = .31$).[12]

Grade in follow-on course. Grade in follow-on course is contingent on successfully

[11] R^2 for the student fixed effects sample (Model 3) is the within unit R^2 and is not comparable to the OLS R^2 in Models 1 and 2.

[12] R^2 for the student fixed effects sample (Model 3) is the within unit R^2 and is not comparable to the OLS R^2 in Models 1 and 2.

completing the follow-on course, resulting in a smaller analysis sample (N = 10,507 students; 20,692 observations; Models 4 and 5) and student fixed effects sample (N = 7,653 students; 17,660 observations; Model 6). Grade in follow-on course may reveal lasting effects of instructional practices for STEM majors. These effects may include retention of knowledge to succeed in future courses and acquisition of necessary skills such as scientific thinking and reasoning. By including all three instructional composites, the models provided a more accurate representation of what occurs in the lecture hall. Results for Models 4-6 are shown in Table 13. There is no association between grade in follow-on course and three instructional composites for the reference group.

Differential effects. To assess differential effects for underrepresented groups, a series of interaction terms were added (composite*female; composite * first-generation; composite*low-income; composite*Hispanic). The interaction terms estimated the extent to which association between instruction and student outcomes was different for certain groups of students. Model 6 showed a negative association for females and *Faculty-Student Interaction* (-0.041; $p < .05$; $R^2 = .40$). Model 5 and Model 6 showed a negative association for Hispanics and Epistemological Instruction (-0.054; $p < .01$; $R^2 = .40$ and -0.067; $p < .001$; $R^2 = .14$ respectively).[13]

--INSERT TABLE 13 ABOUT HERE--

[13] R^2 for the student fixed effects sample (Model 6) is the within unit R^2 and is not comparable to the OLS R^2 in Models 1 and 2.

Logistic Regression Analyses for Student Persistence using Composites

Logistic regression was used to identify any relationships between instructional practices and student persistence, operationalized two ways: (a) successfully completing the next course and (b) staying in a STEM major[14]. The analysis sample (Years 2 and 3) included courses in four academic disciplines that occurred in a series of two or three courses. Models clustered standard errors at the course section level, yielding 121 clusters for the full sample in Models 7 and 8 (N = 13,856 students; 44,981 observations; *Pseudo* $R^2 = .04$) and 71 clusters when only including those observed courses with a specific series in Models 9 and 10 (e.g., Chem1A, 1B, 1C; N = 11,145 students; 31,282 observations; *Pseudo* $R^2 = .09$). Models 8 and 10 included interaction terms. Results for Models 7-10 are shown in Table 14. As a second measure of persistence, Models 11 and 12 used the full analysis sample (N = 13,856 students; 44,981 observations) to identify associations between the instructional composites and persistence in the STEM major. Model 12 included interaction terms. Results for Models 11 and 12 are shown in Table 15. I repeated the analysis with the individual instructional variables, rather than as composites, discussed in greater detail separately below.

--INSERT TABLES 14 & 15 ABOUT HERE--

Odds of completing follow-on course. None of the instructional composites showed significant results for the reference group. However, when interaction terms were included in Model 10 (observed courses with a specific series), first-generation students were negatively associated with completing follow-on course and *Faculty-Student Interaction* (-0.072; $p < .05$;

[14] As noted in Chapters 1 and 3, majors classified as STEM followed the National Science Foundation's Classification of Instructional Programs (CIP).

Pseudo R^2 = .09), but positively associated with completing the follow-on course and *Peer-Interaction* (0.049; $p < .05$; *Pseudo* R^2 = .09) disappeared.

Odds of staying in STEM. None of the instructional composites showed significant results for the reference group. However, for females, *Peer Interaction* provided a positive association for staying in STEM (0.066; $p < .05$; *Pseudo* R^2 = .11).

Regression Analyses using Individual Instructional Practices

I repeated the analysis with the individual instructional variables, rather than as composites. Results of the individual practices remained consistent with the composites and provided detailed information on the relationship between specific instructional practices and student achievement (e.g., the use of clickers). Results for grade in course for the student fixed effects sample are shown in Table 16 (Model 13; N = 11,291 students; 35,998 observations). Results for grade in follow-on course (Model 14; N = 10,507 students; 20,692 observations) are also shown in Table 16. As with Model 5, Model 14 included any student that successfully completed a follow-on course, whether we observed that follow-on course or not.

--INSERT TABLE 16 ABOUT HERE--

Grade in current course. Model 13 regressed the individual instructional practices on grade in observed course using the student fixed effects sample (N = 11,291 students; 35,998 observations; R^2 = .33).[15] Overall, six practices resulted in positive associations, while three practices resulted in negative associations for the reference group. For females, one practice resulted in a positive association and two practices resulted in negative associations. First-

[15] R^2 for the student fixed effects sample (Model 13) is the within unit R^2 and is not comparable to the OLS R^2 in Models 1 and 2.

generation students has no significant findings, while low-income students showed one positive association. For Hispanics, one practice resulted in a positive association, while two resulted in a negative association. While the instructional tools were not included in the composites[16]—guided by the EFA—*clickers* were positively associated with grade for the reference group.

Grade in follow-on course. Model 14 regressed individual instructional practices on grade in follow-on course (N = 10,507 students; 20,692 observations; R^2 = .39) and clustered standard errors at the course section level (67 clusters). Overall, one practice resulted in positive association, while two practices resulted in negative associations. Similarly, for females, two practices resulted in positive associations and one practice resulted in a negative association. For first-generation students, three practices resulted in a negative association. *Peer discussion following a clicker question* resulted in a positive association, yet *clickers* were negatively associated with grade in follow-on course for first-generation students. For low-income students, two practices resulted in a positive association. For Hispanic students, four practices resulted in a negative association, while *formative feedback* resulted in a positive association with grade in follow-on course, *formative feedback* is inclusive of clickers.

Logistic Regression Analyses for Student Persistence using Individual Practices

Student persistence was operationalized as successfully completing the follow-on course (series only) in Model 15 (N = 11,007 students; 31,282 observations; *Pseudo* R^2 = .15 and staying in STEM in Model 16 (N = 13,856 students; 44,981 observations; *Pseudo* R^2 = .10). Results are shown in Table 17.

---

[16] Many of the instructional tools were captured in other variables, and thus were omitted from the polychoric correlation matrix and subsequent EFA.

--INSERT TABLE 17 ABOUT HERE--

Odds of completing follow-on course. For the reference group, four practices resulted in a positive association for completing the follow-on course, while one resulted in a negative association. For females, three practices resulted in a negative association, while *deskwork* resulted in a positive association. For first-generation students, two practices resulted in a negative association and for low-income students, *interactive* resulted in a positive association while two practices resulted in a negative association, as did *clickers*. For Hispanic students, two practices resulted in negative associations and one in a positive association with successfully completing follow-on course.

Odds of staying in STEM. For the reference group, two practices resulted in a positive association for staying in STEM, while three resulted in a negative association. For females, one practice resulted in a negative association as did *clickers*. For first-generation students, four practices resulted in a negative association, while *peer discussion following a clicker question* resulted in a positive association. There were no significant associations for low-income students, and for Hispanic students, there were two practices that resulted in negative associations and *desk work* resulted in a positive association.

Results of the individual practices remained consistent with the composites. Core findings and implications for gateway STEM instructional practice are addressed in the next chapter.

CHAPTER 5: DISCUSSION

The present study extended the work of Reimer et al. (2016) by describing gateway instructional practices observed in lecture hall over three years (RQs 1 and 2) and their relationship to student outcomes for Years 2 and 3 (RQ 3). This necessitated identifying how observable student characteristics were correlated with instructional practices and how often those practices were employed by each department. Course fixed effects controlled for aspects of content and instruction that did not vary across sections of the same course. Student fixed effects controlled for student characteristics across courses including race, gender, economic and academic background, as well as invariant student characteristics such as interest and motivation. First, core findings are related to (a) faculty-student interactions, (b) epistemological instruction, and (c) peer interaction, and presented regarding the research literature from chapter 2. Second, implications for gateway STEM instructional practice are considered. Finally, the chapter concludes with reflection on study limitations, with suggested directions for future research.

Returning to core research questions, the present study found that gateway STEM courses at UCI include gateway instructional practices that go beyond independent learning norms and provide opportunity for students to work together and learn from one another in classes serving 200-550 students. One graduating senior put it well, "Five years ago I never would have encouraged my sister to major in science at a UC because it's so impersonal. Now I am trying to convince her [that] UC Irvine is different; this is where she wants to grow as a scientist and a researcher."[17] I determined the practices implemented in STEM lecture halls fell into three general categories of active gateway learning: *Faculty-Student Interaction, Epistemological*

[17] This student was a graduating 5th year senior. She began as a biology major but transferred to public health. She was the first in her family to attend college. She spent her last three years working on Project SPROUT.

Instruction, and *Peer Interaction*. Results confirmed study hypotheses that opportunities for students to work together—affirming interdependent learning norms—were positively associated with student outcomes. Discussion of departmental trends of observed instructional practices and related student outcomes are organized by instructional composites.

Active Gateway Learning: Faculty-Student Interactions

The pilot study included all interactive strategies in one composite, *Interaction*, and found that male students, Hispanic students, nonresident international students and students retaking a course self-selected into courses with higher levels of this composite (Reimer et al., 2016, p.222). The present study found that faculty-student interactions are distinct from peer interactions. This study found that a higher than average number of white, male, and continuing-generation students enrolled in courses that included higher levels of *Faculty-Student Interaction*, while a higher than average number of first-generation, low-income, Hispanic and female students enrolled in courses that included higher levels of *Peer Interaction*. Regression analyses of the composite *Faculty-Student Interaction* showed no association with overall course grade or overall follow-on course grade. However, for first-generation students, the composite was negatively associated with course grade and the likelihood of completing the follow-on course. Similarly, for female students, the composite was negatively associated with grade in follow-on course (in Model 6, the student fixed effects model). These findings seemingly contradict the initial study's positive association between *Interaction* and course grade and positive association between subsequent course grade and first-generation students (pp. 225-226). The initial study also found that students who enrolled in courses with higher levels of *Interaction* were significantly less likely to progress to the next course (p. 226). A larger sample and the inclusion of factor analysis revealed that the type of interaction matters; *Peer*

Interaction, discussed below, is positively associated with outcomes for diverse students. In contrast to the collaborative nature of *Peer-Interaction*, *Faculty-Student Interaction* is instructor-driven and models the traditional independent culture of higher education (Stephens et al., 2012).

In general, interactive instruction is believed to facilitate and define active gateway learning, resulting in positive student outcomes (Freeman et al., 2014; Smith et al., 2014; Walkington et al., 2012). The assumption is that an instructor communicates openness and accessibility by asking questions, which subsequently encourages students to respond with questions of their own. Formative feedback, a questioning strategy typically involving a clicker question, is a common method employed to increase active gateway learning (e.g., Reimer et al., 2016). Questioning strategies (e.g., Smith et al., 2014) promote independent learning: one student asks the instructor about an incongruity, potentially correcting an instructor error without concern that five hundred other students are watching and waiting. Continuing-generation students understand the culture of higher education and their role as learners who speak with confidence borne of belonging, having lived with the expectation and value of attaining a college degree (Stephens et al., 2012). These students are more likely to engage with content and perform better on tasks, assignments, and exams (Fulmer et al., 2010; Markus, 2008). Gateway learning is self-relevant; their goal is to become independent thinkers (Phillips et al., in press). For continuing-generation students, the principal mode of learning involves being individually recognized and responding in kind. Previous studies on interactive instruction did not distinguish between types of interaction nor how these different types may impact subsets of students differently (e.g., Freeman et al, 2014).

I found that *Faculty-Student Interaction* disadvantaged certain students, which implicates cultural mismatch between normative gateway STEM instructional practices and certain undergraduate groups. For first-generation students, learning is typically a collaborative, community-oriented task (Stephens et al., 2012). These students arrive at university expecting to connect with others, demonstrating sensitivity to the needs of others; their goal is to give back to the community, not become independent thinkers (Fiske & Markus, 2012). Their cultural models of self—how they understand themselves in relation to others—assumes they should "be connected to others and respond to the needs, preferences, and interests of others" (Stephens et al., 2012, p. 1180). The continuing-generation student who asks questions in a lecture hall is quietly accommodated by the first-generation student, who will respectfully listen to the instructor and not interrupt, intending to gain clarification later from peers. Interdependent learning norms have taught them to rely on their community, family and peers.

First-generation students grow up with the expectation to rely on others (e.g., family and friends) and respond to the needs of others (Fiske & Markus, 2012); social class shapes cultural models of self. The bachelor's degree is perceived as a mandate to give back to the community (Phillips et al., in press). First-generation students are also more likely to work during college, potentially worrying about the negative implications of tuition for supporting family members. They have less time and resources to participate in co-curricular activities. As such, costs associated with college and/or co-curricular resources may supersede their utility-value (Canning & Harackiewicz, 2015; Harackiewicz & Hulleman, 2010; Eccles & Wigfield, 2002; Stephens et al., 2012). First-generation students experience unseen disadvantages; gateway STEM courses with higher levels of *Faculty-Student Interaction* are a cultural mismatch.

While female students may be somewhat better at understanding the culture of higher education if they are continuing-generation students, negative associations with practices involving faculty-student interactions suggest social identity threat, which along with cultural models of self is a premise for cultural mismatch. Indeed, a recent experimental study showed that male and female professors continue to exhibit implicit bias against female students in biology, chemistry, and physics (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012). When an instructor asks if there are any questions, it may be that females might be reluctant to speak up and instructors may be less likely to respond to female students. In the context of a gateway STEM lecture hall, practices such as epistemological instruction provide greater benefit to these students.

Additional analyses of gateway practices further support the cultural mismatch interpretation. The most common gateway instructional practice involved faculty asking students if they have questions. This practice was highly correlated to other items measuring interactions between instructors and students and was negatively associated with grade in course for female, first-generation, and Hispanic students. The negative association held true for grade in follow-on course for Hispanic students, whose cultural value of respect commonly overrides any personal need to pose questions to the instructor, even if the instructor has invited questions (Calzada, Fernandez, & Cortes, 2010; Hernandez, Mobley, Coryell, Yu, & Martinez, 2013). Much like first-generation students, Hispanic students will wait to seek answers from their peers. This may also explain why epistemological instructional practices were negatively associated with grade in follow-on course for Hispanic students, as addressed below.

Active Gateway Learning: Epistemological Instruction

Regression analyses of the composite *Epistemological Instruction* showed a positive association with course grade for the reference group. However, the student fixed effects model and the less restrictive model showed a negative association between *Epistemological Instruction* and follow-on course grade for Hispanic students. These findings are different than the initial study, which found no differential effects for the composite *Epistemology*, which included both observational and coded syllabi data (Reimer et al., 2016). While both the initial and current study found a positive association between course grade and the epistemology composite for the reference group, I offer a more nuanced picture of the relationship between these practices and student outcomes. These practices are instructor-driven, formulaic and course-specific (e.g., references to exam content), thus only helpful in the immediate course for continuing generation students who understand the utility-value (usefulness of task) as related to their goals for the course (e.g., the extrinsic motivation of good grades; Eccles, 2009; Eccles & Wigfield, 2002). Instructors model for students how to succeed in their course, students respond by following their lead. However, these may not translate well to the next course taught by a different instructor and disadvantage those students less likely to ask questions if they can't follow the instructor's lead (e.g., while an instructor models algorithms to solve problems).

That *Epistemological Instruction* was negatively associated with follow-on course grade for Hispanic students suggests cultural mismatch with their learning norms. *Epistemological Instruction*—as observed—was formulaic and instructor-driven. These practices emphasize independent learning norms (e.g., what to know to do well on a test) and disadvantage those with interdependent learning expectancies (e.g., Hispanic students wanting to figure it out with their peers). Analyses of the individual practices revealed that problem solving is negatively associated with grade in follow-on course for first-generation students, while real-world examples is positively associated with follow-on course grade. Problems are somewhat different in the next course, but perhaps by relating it to the real world, students find a connection or value. Research on utility-value interventions offer a potential explanation (e.g., Harackiewicz et al., 2016). Experimenters asked students to write a brief (500 word) essay on the personal significance of course content. This intervention reduced the achievement gap for first-generation URM students by sixty-one percent, from a gap of 0.84 grade points to a gap of 0.51 grade points. In step with cultural mismatch, first-generation URM students may have linked course content with perceived obligations to communities of origin—interdependent goals—rather than independent goals (e.g., course grade).

Another explanation is what Stephen Pinker refers to as the "curse of knowledge" (2014). This has also been called "expert blindness." The instructor is providing a method to the class on how to solve a problem, but the deeper concepts behind the problem and how that translates to a novel problem is unclear to the novice student (Bassok & Novick, 2012). This type of interaction supports the transfer of knowledge rather than an opportunity to make sense of the information and apply it to a new context. The most successful models of *Epistemological Instruction* involved Problem-Based Learning (PBL), which includes high levels of student collaboration (Dochy et al., 2003; Strobel & van Barneveld, 2009). Haak and colleagues (2011) found that PBL closed the achievement gap for economically disadvantaged students in biology.

Referring back to Jaime Escalante's story in Chapter 2: (a) he established care and student concern as instructional priorities; (b) he tapped into the interdependent, collaborative ethos of his students, culminating in a collective effort to master the advanced placement calculus test; and (c) he inspired them through their cultural value of *ganas*. His students studied together and learned from one another. Escalante was a culturally responsive teacher. Perhaps more importantly, his emphasis on *Peer Interaction* was a key component.

Active Gateway Learning: Peer Interaction

The two instructional composites above include practices that are instructor-driven, focusing on the transfer of information for success in a particular course. These practices don't afford students an opportunity to interact with one another to make sense of the information, which would facilitate success in future courses (Crouch, Watkins, Fagen, & Mazur, 2007). *Peer Interaction* provides student-driven opportunities for learning (e.g., peer instruction; Crouch et al., 2007; Fagen, Crouch, & Mazur, 2002). Regression analyses of the composite *Peer Interaction* showed a positive association with course grade for first-generation and Hispanic students. While there was a small negative association between *Peer Interaction* and low-income students, we recognize that these three groups have overlap. However, three-way interactions were non-significant. *Peer Interaction* was also positively associated with the likelihood of completing the follow-on course for first-generation students and the likelihood of staying in STEM for female students. Regression analyses of the individual items showed a positive association with follow-on course grade and *deskwork* for females (0.038; $p < .05$; $R^2 = .41$). and *peer discussion following a clicker question* for first-generation students (0.043; $p < .05$; $R^2 = .41$). Desk work was associated with females being 15% more likely to complete the follow-on course. For first-generation students, *deskwork* (6%), *group-work* (9%) and *peer-discussion following a clicker question* (12%) were all positively associated with staying in STEM.

As with previous studies (Lewin et al., 2016; Smith et al., 2009), *Peer Interaction* provides students opportunities, during the lecture, to learn from one another, while they interact with course content. This composite was not used in the initial study. Analyses from the present study identified two distinct forms of interaction. While *Faculty-Student Interaction* is instructor-driven and fosters independent learning norms, *Peer Interaction* facilitates Peer Instruction (PI; Mazur, 1997), shown to improve student conceptual understanding in introductory physics courses (Fagen et al., 2002). The Force Concept Inventory is often used to measure conceptual understanding in these studies (e.g. is a pre/post measure first developed for introductory physics to identify conceptual understanding at the end of a course (Hestenes, Wells, & Swackhamer, 1992). Conceptual learning, resulting from opportunities for students to interact with one another over course content, allows students to apply their knowledge to new contexts. This may be why *Peer Interaction* was the only composite to be positively associated with grade in course and grade in follow-on course.

Instructors who maximize opportunities for *Peer Interaction* are creating a cultural match for first-generation, Hispanic, and female students. We know that students tend to sit with friends in lecture halls, so when given an opportunity to collaborate with those nearby, they likely are in a group where they have a sense of belonging. These practices are culturally responsive—they support interdependent learning norms—student-driven learning. As mentioned above, first-generation and Hispanic students grow up with expectation to rely on others (e.g., family and friends) and respond to the needs of others (Fiske & Markus, 2012). *Peer Interaction* in the gateway lecture hall allows these students to connect with content much like *Faculty-Student Interaction* practices do for continuing-generation students. The present study supports earlier findings on single courses (e.g., Smith et al., 2009) or within a single discipline (Crouch et al., 2007). Smith et al. (2009) found that when students were afforded opportunity to discuss a clicker question with peers and then given a second chance to answer the same question, conceptual understanding improved, even if no one in the group answered correctly the first time (Smith et al., 2009).

Active gateway learning inclusive of peer interaction addresses values affirmation and utility-value interventions. By way of review, values-affirmation attempts to reduce student anxiety associated with misrecognition and ameliorate personal struggles to find belonging (Lewis & Sekaquaptewa, 2016; Spencer, Logel, & Davies, 2016). First-generation, Hispanic, and female students experience recognition and belonging when interacting with peers in the same way a continuing-generation student experiences belonging when responding to an instructor's question. Utility-value interventions are designed to improve student motivation through goal identification and achievement (Eccles, 2009; Harackiewicz & Hulleman, 2010; Hulleman & Harackiewicz, 2009). The cost (personal consequence) is low and utility (usefulness of task related to goals) high for continuing-generation students responding to an instructor's questions, facilitating interaction with course content. *Peer Interaction* provides the same low-cost, high-utility opportunity for students with interdependent learning norms to interact with course content. As an added benefit, when the learning environment matches the interdependent learning norms, tasks are construed as less difficult and task performance improves (Stephens et al., 2012).

Finally, *Peer Interaction* provides student-driven opportunities for learning. In the literature, this is sometimes referred to as *Peer Instruction* and has been shown to improve conceptual understanding (Crouch et al., 2007; Fagen, Crouch, & Mazur, 2002). Rather than novices being disadvantaged by the "curse of knowledge" or "expert blindness" of the instructor, by interacting with their peers, students can make sense of the information and apply it to a new context. Peers are still close to the challenges of learning the material and thus may be better at helping each other understand (Mazur, 1997). Or as Smith and colleagues (2009) found, it's not even necessary if anyone in a group knows the answer, as long as the students have time to figure it out together.

As previously noted, providing opportunities for peers to interact with one another in a large gateway STEM lecture hall is difficult. *Peer Interaction* in this study occurred briefly during the lecture. Instructors told students to turn to a neighbor and explain or defend an answer to a clicker question in 30 seconds. Instructors displayed a new problem and asked students to work on it for a minute, followed by the resumption of lecture. Community-oriented values are affirmed when students are encouraged to work together, if only for a few minutes (e.g., Smith et al., 2009). While interactive practices have been generally associated with student success (e.g., Freeman et al., 2014), more work is needed to develop purposive *Faculty-Student Interaction* and *Peer Interaction* instructional strategies such that all student groups might benefit from active gateway learning.

Implications for Gateway STEM Instructional Practice

Because we know that students are more likely to attend lecture than discussions or co-curricular support services, instruction needs to be culturally responsive, including practices that match the learning expectancies of diverse student populations. Research has already shown that American higher education favors independent learning norms (Stephens et al., 2012). The present study has shown that large gateway STEM lecture courses, while embracing these norms, can nonetheless be modified to effectively educate diverse learners, including a range of undergraduate expectancies spanning intellectual independence to community service and obligation. First-generation, low-income, URM, and female students, to a greater or lesser extent, have been disadvantaged due to cultural mismatch with gateway STEM instruction. The present study underscores several ways that instructors can augment traditional lecture to include interdependent learning norms without detrimental effects for those students who favor independent styles of learning (e.g., continuing-generation students).

Many large STEM lecture courses use clickers to increase student participation and engagement. A crucial next step is to pair this technology with the instructor inviting students to discuss the problem with peers. As Smith et al. (2009) have shown, it does not matter whether anyone in the group knows the correct answer the first time. It simply matters that students are afforded opportunity to discuss and try again. The Smith et al. (2009) study was a controlled experiment in a single course. It did not assess the quality of gateway instructional practices, rather indexing practices that were immediately present. As the present research additionally considered the co-occurrence of instructional practices and how these impacted student outcomes, current findings constitute an exclamation mark for this line of pedagogical inquiry.

I had numerous occasions to talk with instructors during the three-year study. I was often asked about what instructors should be doing, or what they should do differently. By way of response, I would revisit observed practices in their class and suggest small adjustments. These might include providing thirty seconds for students to discuss a clicker question with their neighbor, having them talk to one another before walking through a problem on the board, having them talk to one another after solving a problem to see if they got similar results and used similar approaches, or having them sit by discussion section during lecture such that students were given four opportunities for interaction each week instead of one. Active gateway learning is enhanced with peer interaction.

The present findings highlight the importance of decomposing and interpreting research findings by student characteristics. Results will provide insight into the utility of promising instructional practices in lower-division gateway STEM courses. As such, current findings can provide guidance on gateway STEM instructional reform at large public research universities. It is my hope the present findings will stimulate critical pedagogical reflection for STEM instructors, facilitate collaborative discussion within STEM departments, and promote institutional conversation pertinent to their teaching mission (e.g., "discovering and advancing knowledge")[18]. Findings might inform orientations for new faculty and graduate students, increasing awareness of culturally responsive teaching and readily implemented strategies.

It should be noted that study findings should be qualified by the selective nature of the institution under study. Ladder rank faculty, especially at a public ivy institutions such as UCI, have extramural grants and significant obligations for scholarship. Publications and teaching activities, for these individuals, often compete for attention. At the same time, students must balance a complex mix of classes, jobs, and family obligations. Standards and expectations are very high at the University of California. In the case of gateway STEM lecture courses, small instructional changes can improve learning. While the straightforward instructional changes commended by this study evinced positive benefits at UCI, they may exert an even greater positive effect at less selective universities where competing interests are possibly somewhat less intense.

Because of its scale, study findings are potentially generalizable, with anticipated relevance to gateway STEM instruction in similar American university contexts. The tools used for this study—observation protocols, surveys, and interview protocols—were selected to maximize opportunities for replication in other settings.

[18] Mission statement from the University of California Academic Plan, 1974-1978.

Limitations

As with any study, there were limitations recognized at the outset (see Chapter 1) and others that arose as the study progressed. It became evident that some students enrolled in a course with one instructor, but attended the lecture of another. This kind of "cross-pollination" may somewhat diminish the significance of study findings. When analyzing whether a student stayed in a STEM major or not, there are many majors that one could include beyond the National Science Foundation's Classification of Instructional Programs (CIP) that also limit the findings. One can argue that leaving a STEM major for some is a good choice, and that for some students, their pathway still leads to advanced degrees. However, I find it disconcerting that certain groups of students leave a disproportionately higher rate. While the present study was comprehensive, it was limited to a single, highly selective institution. It did not evaluate how well instructional practices were implemented, only that they occurred in one or both randomly selected observations[19]. Smaller studies have shown lasting results of interventions through controlled experimental design. This is a potential asset given that students commonly begin, withdraw, and re-enter college. Students may take courses at other institutions, including online. Another limitation was the binary nature of many of the variables, captured during two observations. While the observational data includes qualitative descriptions supporting these variables, a full qualitative analysis was beyond the scope of this study. The qualitative data was used for clarification and to audit the data. Finally, attempts to estimate the likelihood of completing a follow-on course, grade in course, and likelihood of persisting in a STEM major are notoriously difficult. Cross-listed courses make identification of follow-on courses

[19] We chose a bottom-up approach to observing instructional practices to identified if practices believed to be promising were validated.

potentially problematic. Finally, observation protocols are invariably prone to human error. The present study involved five hundred observations conducted by more than thirty people over three years. Errors are inevitable.

Future Research

This study parsed differences in interactive practices in relation to students. Future research is needed regarding peer interaction in gateway STEM instruction, specifically addressing physical challenges associated with implementation. This study focused exclusively on practices in the lecture hall. Gateway instructional practices will need to be augmented by pedagogical studies in different (smaller) environments. Future research should consider how gateway instructional practices are associated with the likelihood of students pursuing co-curricular resources. This is pressing given that co-curriculars commonly target students with the least amount of time to take advantage of them. Future research questions might investigate what instructors should do to match co-curricular resources with student values and learning expectancies. Others should consider how gateway instructional practices are related to first-generation, URM, and Hispanic participation in undergraduate research. Longitudinal analyses are needed to consider gateway instructional practices in relation to degree completion, time to degree completion, and post-baccalaureate outcomes. Future research will need to expand our understanding of how these practices impact underrepresented students at other public universities, community colleges, and private colleges. And finally, my next step is to conduct a qualitative study of the data regarding *Peer Interaction*.

REFERENCES

Arnett, J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist, 55*, 469-480. DOI: http://dx.doi.org/10.1037/0003-066X.55.5.469

Baber, L. (2015). Considering the interest-convergence dilemma in higher education. *The Review of Higher Education*, *38*, 251-270.

Bailey, M., & Dynarski, S. (2011). *Gains and gaps: Changing inequality in US college entry and completion* (No. w17633). National Bureau of Economic Research.

Baillie, C., & Fitzgerald, G. (2010). Motivation and attribution in engineering students. *European Journal of Engineering Education*, *25*, 145-55.

Bassok, M., & Novick, L. (2012). Problem solving. In K. Holyoak & R. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 456-474). New York: Oxford University Press.

Berland, L., Schwarz, C., Krist, C., Kenyon, L., Lo, A., & Reiser, B. (2015). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching*. DOI: 10.1002/tea.21257

Black, P. (2013). Pedagogy in theory and in practice: Formative and summative assessments in classrooms and in systems. In D. Corrigan, R. Gunstone, & A. Dordrecht (Eds.) *Valuing assessment in science education: Pedagogy, curriculum, policy* (pp. 207-229). Dordrecht, Netherlands: Springer.

Blasco-Arcas, L., Buil, I., Hernández-Ortega, B., & Sese, F. (2013). Using clickers in class: The role of interactivity, active collaborative learning and engagement in learning performance. *Computers & Education*, *62*, 102-110.

Brint, S., Cantwell, A., & Hanneman, R. (2008). The two cultures of undergraduate academic engagement. *Research in Higher Education, 49*, 383-402.

Canning, E., & Harackiewicz, J. (2015). Teach it, don't preach it: The differential effects of directly-communicated and self-generated utility-value information. *Motivation Science, 1*, 47-71.

Carpenter, S. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science*, *21*, 279-283.

Chaplin, S. (2009). Assessment of the impact of case studies on student learning gains in an introductory biology course. *Journal of College Science Teaching, 39*, 72–79.

Chen, X. (2013). STEM attrition: College students' paths into and out of STEM fields (NCES 2014-001). *National Center for Education Statistics, Institute of Education Sciences,* U.S. Department of Education. Washington, DC.

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, *24*, 205-249.

Clotfelter, C., Ladd, H., & Vigdor, J. (2007). How and why do teacher credentials matter for student achievement? *NBER working paper 12828*.

Crouch, C. H., Watkins, J., Fagen, A. P., & Mazur, E. (2007). Peer instruction: Engaging students one-on-one, all at once. *Research-based reform of university physics*, *1*(1), 40-95.

DeLuca, V., & Lari, N. (2013). Developing students' metacognitive skills in a data-rich environment. *Journal of STEM Education*, *14*, 45–55.

De Winter, J., & Doudou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics, 39*, 695-710.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction, 13*, 533–568.

Dunbar, K. and Klahr, D. (2012). Scientific thinking and reasoning. In K. Holyoak & R. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 456-474). New York: Oxford University Press.

Durik, A., Shechter, O., Noh, M., Rozek, C., & Harackiewicz, J. (2015). What if I can't? Success expectancies moderate the effects of utility value information on situational interest and performance. *Motivation and Emotion*, DOI: 10.1007/s11031-014-9419-0

Eagan, K., Hurtado, H, Figueroa, T., and Hughes, B. (2014). *Examining STEM pathways among students who begin college at four-year institutions*. Commissioned paper prepared for the Committee on Barriers and Opportunities in Completing 2- and 4-Year STEM Degrees, National Academy of Sciences, Washington, DC. Available: http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_088834.pdf [April 2015].

Eagan, M., Stolzenber, E., Berdan Lozano, J., Aragon, M., Suchard, M., & Hurtado, S. (2014). Undergraduate teaching faculty: The 2013-2014 HERI faculty survey. Los Angeles: Higher Education Research Institute, University of California, Los Angeles.

Eccles, J. (2009). Who am I and What am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist, 44*, 78-89.

Eccles, J., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of*

Psychology, 53, 109-132.

Eddy, S., & Hogan, K. (2014). Getting under the hood: How and for whom does increasing course structure work? *CBE-Life Sciences Education*, *13*, 453-468.

Eichler, J., & Peeples, J. (2016). Flipped classroom modules for large-enrollment general chemistry courses: A low barrier approach to increase active learning and improve student grades. *Chemistry Education Research and Practice*, *17*, 197-208.

Engle, J., & Tinto, V. (2008). Moving beyond access: College success for low-income, first-generation students. *Pell Institute for the Study of Opportunity in Higher Education*.

Escalante, J., & Dirmann, J. (1990). The Jaime Escalante math program. *Journal of Negro Education, 59*, 407-423.

Fagen, A. P., Crouch, C. H., & Mazur, E. (2002). Peer instruction: Results from a range of classrooms. *The Physics Teacher*, *40*(4), 206-209.

Figlio D., Rush, M., & Yin, L. (2013). Is it live or is it internet? Experimental estimates of the effects of online instruction on student learning. *Journal of Labor Economics, 31*, 763-784. DOI:10.1086/669930

Fiske, S. T., & Markus, H. R. (2012). A wide angle lens on the psychology of social class. In S. T. Fiske & H. R. Markus (Eds.), Facing social class: Social psychology of social class. New York, NY: Russell Sage.

Flora, D., & Curran, P. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466.

Freeman, S., Eddy, S., McDonough, M., Smith, M., Okoroafor, N., Jordt, H., & Wenderoth, M. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*, 8410-8415.

Freeman, S., O'Connor, E., Parks, J., Cunningham, M., Hurley, D., Haak, D., Dirks, C. & Wenderoth, M. (2007). Prescribed active learning increases performance in introductory biology. *CBE-Life Sciences Education*, *6*, 132-139.

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

Fulmer, C., Gelfand, M., Kruglanski, A., Kim-Prieto, C., Diener, E., Pierro, A., & Higgins, E. (2010). On "feeling right" in cultural contexts: How person-culture match affects self-esteem and subjective well-being. *Psychological Science, 21*, 1563-1569. DOI: 10.1177/0956797610384742

Gasiewski, J., Eagan, M., Garcia, G., Hurtado, S., & Chang, M. (2012). From gatekeeping to engagement: A multicontextual, mixed method study of student academic engagement in introductory STEM courses. *Research in Higher Education*, *53*, 229-261.

Gauci, S., Dantas, A., Williams, D., & Kemm, R. (2009). Promoting student-centered active learning in lectures with a personal response system. *Advances in Physiology Education*, *33*, 60-71.

Gay, G. (2010). *Culturally responsive teaching: Theory, research, and practice (2nd edition).* New York: Teachers College Press.

Goldman, A. (1986). *Epistemology and Cognition.* Boston, MA: Harvard University Press.

Gorsuch, R. (1970). A comparison of biquartimin, maxplane, promax, and varimax. *Educational and Psychological Measurement*, *30*, 861-872.

Grant, B. (2008). *Practitioner research as a way of knowing: A case study of teacher learning in improving undergraduates' concept acquisition of evolution by natural selection*, NAS

Reviewed Commissioned Paper from the National Research Council Board on Science Education Workshop "Linking Evidence and Promising Practices in STEM Undergraduate Education," at the National Academies, Washington, DC (June 30, 2008).

Grantham, A., Robinson, E., & Chapman, D. (2015). "That truly meant a lot to me": A qualitative examination of meaningful faculty-student interactions. *College Teaching, 63*, 125-132. DOI:10.1080/87567555.2014.985285

Haak, D., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, *332*, 1213-1216.

Hahn, U., & Oaksford, M. (2012). Rational argument. In K. Holyoak & R. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 277-298). New York: Oxford University Press.

Hair, J., Black, W., Babin, B., & Anderson, R. (2009). *Multivariate data analysis* (7th edition). New York: Pearson.

Han, J., & Finkelstein, A. (2013). Understanding the effects of professors' pedagogical development with clicker assessment and feedback technologies and the impact on students' engagement and learning in higher education. *Computers & Education, 65,* 64–76.

Hand, V., Penuel, W.., & Gutiérrez, K. (2012). (Re) framing educational possibility: Attending to power and equity in shaping access to and within learning opportunities. *Human Development*, *55*, 250-268.

Harackiewicz, J., Canning, E., Tibbetts, Y., Priniski, S., & Hyde, J. (2016). Closing achievement gaps with a utility-value intervention: Disentangling race and social class. *Journal of Personality and Social Psychology, 111(5)*, 745-765.

Harackiewicz, J., Canning, E., Tibbetts, Y., Giffen, C., Blair, S., Rouse, D., & Hyde, J. (2014). Closing the social class achievement gap for first-generation students in undergraduate biology. *Journal of Educational Psychology, 106*, 375-389. DOI: http://dx.doi.org/10.1037/a0034679

Harackiewicz, J., & Hulleman, C. (2010). The importance of interest: The role of achievement goals and task values in promoting the development of interest. *Social and Personality Psychology Compass, 4*, 42-52.

He, W., Holton, A., Farkas, G., & Warschauer, M. (2016). The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions. *Learning and Instruction*, *45*, 61-71.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The physics teacher*, *30*(3), 141-158.

Hill, C., Corbett, C., & St. Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. American Association of University Women. 1111 Sixteenth Street NW, Washington, DC 20036.

Hora, M., & Ferrare, J. (2014). Remeasuring postsecondary teaching: How singular categories of instruction obscure the multiple dimensions of classroom practice. *Journal of College Science Teaching, 43*, 36–41.

Hulleman, C., & Harackiewicz, J. (2009). Promoting interest and performance in high school science classes. *Science, 326*, 1410-1412.

Hung, W., Jonassen, D., & Liu, R. (2008). Problem-based learning. In J. Spector, M. Merrill, J. van Merrienboer, & M. Driscoll (Eds.), H*andbook of research on educational communications and technology* (3rd ed., Vol. 1, pp. 485-506). New York: Erlbaum.

Ifill, N., Radford, A., Wu, J., Cataldi, E., Wilson, D., & Hill, J. (2016). *Persistence and attainment of 2011-2012 first-time postsecondary students after 3 years* (BPS:12/14) (NCES 2016-401). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [January 19, 2016] from http://nces.ed.gov/pubsearch/.

Inzlicht, M., & Schmader, T. (2012). *Stereotype threat: Theory, process, and application*. New York: Oxford.

Irizarry, J. (2007). "Home growing" teachers of color: Lessons learned from a town-gown partnership. *Teacher Education Quarterly, 34*, 87-102.

Jaggars, S., & Xu, D. (2016). How do online course design features influence student performance? *Computers & Education, 95*, 270-284. DOI: 10.1016/j.compedu.2016.01.014

Ke, F., Chávez, A. F., Causarano, P., & Causarano, A. (2011). Identity presence and knowledge building: Joint emergence in online learning environments? *International Journal of Computer-Supported Collaborative Learning*, *6*, 349-370.

Lewin, J., Vinson, E., Stetzer, M., & Smith, M. (2016). A campus-wide investigation of clicker implementation: The status of peer discussion in STEM classes. *CBE-Life Sciences Education*, *15*, 1-12.

Lewis, N., & Sekaquaptewa, D. (2016). Beyond test performance: A broader view of stereotype threat. *Current Opinion in Psychology*, DOI: 10.1016/j.copsyc.2016.05.002

Malcolm, S. and Feder, M. (Eds.). (2016). *Barriers and opportunities for 2-year and 4-year STEM degrees: systemic change to support diverse student pathways*. Washington, DC: The National Academies Press. DOI: 10.17226/21739

Markus, H. (2008). Pride, prejudice, and ambivalence: Toward a unified theory of race and ethnicity. *American Psychologist, 63*, 651-670.

Martell, C. (2013). Race and histories: Examining culturally relevant teaching in the US history classroom. *Theory & Research in Social Education*, *41*, 65-88.

Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ: Prentice-Hall.

McDaniel, M., Anderson, J., Dervish, M., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*, 494-513.

Mervis, J. (2010). Better intro courses seen as key to reducing attrition of STEM majors. *Science*, *330*, 306.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, *109*(41), 16474-16479.

Nasir, N., Rosebery, A., Warren, B., & Lee. C. (2006). Learning as a cultural process: Achieving equity through diversity. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 489-504). Cambridge, UK: Cambridge University Press.

National Academy of Engineering (NAE). (2005). *Educating the engineer of 2020: Adapting engineering education to the new century*. Washington, DC: The National Academies Press.

Newman, M. (2005). Problem-based learning: an introduction and overview of the key features of the approach. *Journal of Veterinary Medical Education, 32*, 12–20.

Nielsen, N. (Ed.). (2011). *Promising practices in undergraduate science, technology, engineering, and mathematics education: A summary of two workshops*. National Academies Press.

Ong, M., Wright, C., Espinosa, L., & Orfield, G. (2011). Inside the double bind: A synthesis of empirical research on undergraduate and graduate women of color in science, technology, engineering, and mathematics. *Harvard Educational Review, 81*, 172-209. DOI: http://dx.doi.org/10.17763/haer.81.2.t022245n7x4752v2

Owens, M., Seidel, S., Wong, M., Bejines, T., Lietz, S., Perez, J., Sit, S., Subedar, Z., Acker, G., Akana, S., Balukjian, B., Benton, H., Blair, J., Boaz, S., Boyer, K., Bram, J., Burrus, L., Byrd, D., Caporale, N., Carpenter, E., Chan, Y., Chen, L., Chovnick, A., Chu, D., Clarkson, B., Cooper, S., Creech, C., Crow, K., de la Torre, J., Denetclaw, W., Duncan, K., Edwards, A., Erickson, K., Fuse, M., Gorga, J., Govindan, B., Green, L., Hankamp, P., Harris, H., He, Z., Ingalls, S., Ingmire, P., Jacobs, J., Kamakea, M., Kimpo, R., Knight, J., Krause, S., Krueger, L., Light, T., Lund, L., Marquez-Magana, L., McCarthy, B., McPherson, L., Miller-Sims, V., Moffatt, C., Muick, P., Nagami, P., Nusse, G., Okimura, K., Pasion, S., Patterson, R., Pennings, P., Riggs, B., Romeo, J., Stillman, J., Swei, A., Wade, J., Waters, S., Weinstein, S., Willsie, J., Wright, D., Harrison, C., Kelley, L., Trujillo, G., Domingo, C., Schinske, J., & Tanner, K. (2017). Classroom sound can be used to classify teaching practices in college science courses. *Proceedings of the National Academies of Sciences*, DOI: 10.1073/pnas.1618693114

Pace, D., & Middendorf, J. (2004). Decoding the disciplines: Helping students learn disciplinary ways of thinking. *New Directions for Teaching and Learning, 98,* 1–12.

Pastötter, B., & Bäuml, K. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology*, *5,* 1-5.

Phillips, L., Stephens, N., & Townsend, S. (in press). Access is not enough: Cultural mismatch persists to limit first-generation students' opportunities for achievement throughout college. *Journal of Personality and Social Psychology*.

Pinker, S. (2014). *The sense of style: The thinking person's guide to writing in the 21st Century*. New York, NY: Penguin.

President's Council of Advisors on Science and Technology (PCAST). (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics.* Washington, DC: Executive Office of the President.

Price, J. (2010). The effect of instructor race and gender on student persistence in STEM fields. *Economics of Education Review*, *29*, 901-910.

Reimer, L., Nili, A., Nguyen, T., Domina, T., & Warschauer, M. (2016). Clickers in the wild: A campus-wide study of student response systems. In G. Weaver, W. Burgess, A. Childress, & L. Slakey (Eds.), *Transforming institutions: Undergraduate STEM education for the 21st century*. West Lafayette, IN: Purdue University Press.

Reimer, L., Schenke, K., Nguyen, T., O'Dowd, D., Warschauer, M., & Domina, T. (2016). Evaluating promising practices in undergraduate STEM lecture courses. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 2*, 212-230.

Ryan, M., & Reid, S. (2015). Impact of the flipped classroom on student performance and retention: A parallel controlled study in general chemistry. *Journal of Chemical Education, 93*, 13-23.

Sanchez, I., & Gunawardena, C. (1998). Understanding and supporting the culturally diverse distance learner. In C. Gibson (Ed.), *Distance learners in higher education* (pp. 47-64). Madison, WI: Atwood Publishing.

Sawada, D., Piburn, M., Judson, E., Turley, J., Falconer, K., Benford, R., and Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics 102*, 245–53.

Singer, S., Nielsen, N., & Schweingruber, H. (Eds.). (2012). *Discipline-based education research: understanding and improving learning in undergraduate science and engineering*. National Academies Press.

Smith, D., & Ayers, D. (2006). Culturally responsive pedagogy and online learning: Implications for the globalized community college. *Community College Journal of Research and Practice, 30*, 401-415. DOI:10.1080/10668920500442125

Smith, M., Vinson, E., Smith, J., Lewin, J., & Stetzer, M. (2014). A Campus-Wide Study of STEM Courses: New Perspectives on Teaching Practices and Perceptions. *CBE-Life Sciences Education*, *13*, 624-635.

Smith, M., Wood, W., Adams, W., Wieman, C., Knight, J., Guild, N., & Su, T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science*, *323*, 122-124.

Soria, K., & Stebleton, M. (2012). First-generation students' academic engagement and retention. *Teaching in Higher Education*, *17*, 673-685.

Spencer, S., Logel, C., & Davies, P. (2016). Stereotype threat. *Annual Review of Psychology, 67*, 415-437.

Stage, F., & Kinzie, J. 2009. Reform in undergraduate science, technology, engineering, and mathematics: The classroom context. *The Journal of General Education, 58*, 85-105.

Steele, C. (2011). *Whistling Vivaldi: How stereotypes affect us and what we can do*. New York: Norton.

Steele, C., Spencer, S., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology*, *34*, 379-440.

Stephens, N., Fryberg, S., Markus, H., Johnson, C., & Covarrubias, R. (2012). Unseen disadvantage: how American universities' focus on independence undermines the academic performance of first-generation college students. *Journal of Personality and Social Psychology*, *102*, 1178.

Strobel, J., & van Barneveld, A. (2009). When is PBL more effective? A meta-synthesis of meta-analyses comparing PBL to conventional classroom. *Interdisciplinary Journal of Problem-based Learning, 3*, 44–58.

Tekian, A., & Hruska, L. (2004). A review of medical school records to investigate the effectiveness of enrichment programs for “at risk” students. *Teaching and Learning in Medicine*, *16*, 28-33.

Tibbets, Y., Harackiewicz, J., Priniski, S., & Canning, E. (2016). Broadening participation in the life sciences with social-psychological interventions. *CBE—Life Sciences Education, 15*, 1-10. DOI: 10.1187/cbe.16-01-0001.

Tinto, V. (2012). *Completing college: Rethinking institutional action*. Chicago: University of Chicago Press.

U.S. News and World Report. (2010). Best colleges. *U.S. News and World Report*. Retrieved from http://colleges.usnews.rankingsandreviews.com/best-colleges

Walkington, C., Arora, P., Ihorn, S., Gordon, J., Walker, G., Abraham, L., and Marder. M. (2012). *Development of the UTeach observation protocol: A classroom observation instrument to evaluate mathematics and science teachers from the UTeach preparation* *program*. Unpublished paper. Southern Methodist University.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach*. *5th ed*. Mason, OH: South-Western.

Wolter, B., Lundeberg, M., Kang, H., & Herreld, C. (2011). Students' perceptions of using personal response systems ("Clicker") with cases in science. *Journal of College Science Teaching, 40*, 14–19.

Xu, Z., Hannaway, J., & Taylor, C. (2011). Making a difference? The effects of Teach for America in high school. *Journal of Policy Analysis and Management, 30*, 447–469.

APPENDIX A

Table 1
Courses Observed, Years 1-3

Course	Title	Sections
BioSci93	DNA to Organisms	8
BioSci94	Organisms to Ecosystems	7
BioSci97	Genetics	1
BioSci98	Biochemistry	4
BioSci99	Molecular Biology	5
EECS10	Computational Methods in Electrical and Computer Engineering	1
EECS12	Introduction to Programming	1
Engr7A	Introduction to Engineering 1	2
Engr7B	Introduction to Engineering 2	1
Engr30	Statics	3
EngrCEE20	Introduction to Computational Problem Solving	1
EngrMAE10	Introduction to Engineering Computations	2
EngrMAE91	Introduction to Thermodynamics	1
I&CSci6B	Boolean Algebra & Logic	2
I&CSci6D	Discrete Math for Computer Science	1
I&CSci31	Intro to Programming	7
I&CSci32	Programming with Software Libraries	2
I&CSci33	Intermediate Programming	4
Stats7	Basic Statistics	11
Stats8	Introduction to Biological Statistics	5
Chem1A	General Chemistry 1	17
Chem1B	General Chemistry 2	13
Chem1C	General Chemistry 3	9
Chem51A	Organic Chemistry 1	11
Chem51B	Organic Chemistry 2	6
Chem51C	Organic Chemistry 3	9
Math2A	Calculus 1	8
Math2B	Calculus 2	8
Math2D	Multivariable Calculus 1	6
Physics3A	Basic Physics 1	7
Physics3B	Basic Physics 2	4
Physics3C	Basic Physics 3	4
Physics7C	Classical Physics 1	12
Physics7D	Classical Physics 2	6
Physics7E	Classical Physics 3	1
Total		190

Table 2
Courses Observed, Years 2 and 3 (Analysis Sample)

Course	Title	Sections	Including Duplicate Sections
BioSci93	DNA to Organisms	5	10
BioSci94	Organisms to Ecosystems	3	4
BioSci97	Genetics	1	3
BioSci98	Biochemistry	4	4
BioSci99	Molecular Biology	5	5
Engr7A	Intro to Engineering 1	2	2
Engr7B	Intro to Engineering 2	1	1
EngrCEE20	Intro to Computational Problem Solving	1	1
EngrMAE10	Intro to Engineering Computations	2	2
Engr30[+]	Statics	1	1
I&CSci6B	Boolean Algebra and Logic	2	2
I&CSci6D	Discrete Math for Computer Science	1	1
I&CSci31[*]	Intro to Programming	5	7
I&CSci32[*]	Programming with Software Libraries	2	3
I&CSci33[*]	Intermediate Programming	4	4
Stats 7[#]	Basic Statistics	9	12
Stats 8[#]	Intro to Bio Statistics	5	5
Chem1A	General Chemistry 1	11	12
Chem1B	General Chemistry 2	9	13
Chem1C	General Chemistry 3	7	9
Chem51A	Organic Chemistry 1	8	10
Chem51B	Organic Chemistry 2	3	3
Chem51C	Organic Chemistry 3	5	5
Math2A[**]	Calculus	4	6
Math2B	Calculus	5	7
Math2D	Multivariate Calculus	4	4
Physics3A	Basic Physics 1	7	8
Physics3B	Basic Physics 2	4	5
Physics3C	Basic Physics 3	4	4
Physics7C	Classical Physics 1	8	8
Physics7D	Classical Physics 2	6	6
Physics7E	Classical Physics 3	1	1
Total		139	168

Note. [+]Course is cross-listed as EngrCEE30 and EngrMAE30. [*]Courses are cross-listed in Computer Science and Engineering. [#]Students take either Stats7 or Stats8. [**]Most STEM majors are only required to take Math2A.

Table 3
Course Participation Rates by Quarter, Years 1-3

	Observed Courses		Including Duplicate Sections	
	Participation	Percent	Participation	Percent
Pilot Year				
Spring 2013	8/8	100%	10/10	100%
Year One				
Fall 2013	23/24	96%	29/30	97%
Winter 2014	20/24	83%	23/27	85%
Year Two				
Fall 2014	30/32	94%	39/41	95%
Winter 2015	27/32	84%	32/37	86%
Spring 2015	22/24	92%	24/26	92%
Year Three				
Fall 2015	18/23	78%	25/30	83%
Winter 2016	24/28	86%	28/32	88%
Spring 2016	18/25	72%	20/27	74%
Total	190/220	86%	230/260	88%

Note. All three years of observational data were used for latent variable analysis of instructional variables. Some courses were co-taught. Some courses had duplicate sections taught by the same instructor.

Table 4
Instructor Characteristics

	Years 1 - 3		Years 2 and 3	
	Number	Percent	Number	Percent
Title				
Graduate Student	4	4%	1	1%
Post-Doc	2	2%	2	3%
Research Scientist	3	3%	3	4%
Adjunct	1	1%	1	1%
Lecturer	5	5%	3	4%
Lecturer PSOE	6	7%	6	8%
Lecturer SOE	10	11%	10	14%
Assistant Professor	8	9%	7	10%
Associate Professor	13	14%	9	12%
Professor	38	42%	30	42%
Emeritus	1	1%	0	0%
Gender				
Female	27	30%	23	32%
Male	64	70%	49	68%
School				
Biological Sciences	23	25%	19	26%
Engineering/ICS*	14	15%	12	17%
Physical Sciences	54	60%	41	57%
Discipline				
Biology	23	25%	19	26%
General Chemistry	11	12%	10	14%
Organic Chemistry	11	12%	9	12%
Engineering/ICS*	14	21%	10	14%
Physics	17	19%	15	21%
Math	13	14%	7	10%
Statistics	2	2%	2	3%
Total	91		72	

Note. *Courses in ICS are cross-listed in Engineering.

Table 5
Number of Courses Taken by Students, Years 2 and 3 (Analysis Sample[*])

Courses	Students	Percent
1	2,770	19.99%
2	2,257	16.29%
3	1,950	14.07%
4	1,668	12.04%
5	1,388	10.02%
6	1,103	7.96%
7	841	6.07%
8	611	4.41%
9	467	3.37%
10	338	2.44%
11	222	1.60%
12	141	1.02%
13	72	0.52%
14	22	0.16%
15	4	0.03%
16	2	0.00%
N	13,856	100%

Note. Mean: 4.13 (*SD* = 2.83). *Does not include transfer students (n = 3,577) and students with missing SAT scores (n = 17).

Table 6
Student Demographics, Years 2 and 3

	All Students*	Entered as Freshman	Analysis Sample	Student Fixed Effects Sample
Asian	43%	43%	45%	46%
Hispanic	22%	22%	22%	21%
White	11%	11%	11%	11%
Non-resident	16%	16%	13%	13%
Declined to state	4%	4%	5%	5%
Other**	4%	4%	5%	5%
Female	50%	50%	53%	50%
Male	50%	50%	47%	50%
First-generation	49%	49%	50%	50%
Low-income	35%	35%	36%	36%
Entered as freshman	92%	100%	100%	100%
Average SAT Math Score	631	632	629	630
SD	(86)	(86)	(86)	(85)
Average SAT Verbal Score	562	562	565	566
SD	(91)	(91)	(90)	(90)
Average Entering GPA	3.92	3.96	3.98	3.98
SD	(0.28)	(0.23)	(0.23)	(0.23)
N students	19,134	16,679	13,856	11,291
N observations	85,518	78,556	44,981	35,998

Note. *Includes students enrolled in similar courses, but sections that were not observed. **Other includes two or more races, Black, Pacific Islander, and American Indian. Not all categories equal 100% due to rounding. Standard deviations in parentheses.

Table 7
Observed Instructor Practices by Department/Discipline, Years 1-3

	Biology	Gen Chem	Organic Chem	Engr/ ICS	Math/ Stats	Physics
Faculty-Student Interactions						
Interactive	55%	36%	56%	48%	49%	58%
Instructor display questions	71%	81%	73%	75%	89%	72%
Instructor open-ended questions	37%	14%	19%	27%	27%	16%
Instructor checks for questions	69%	71%	79%	77%	64%	71%
Formative feedback	76%	29%	23%	25%	52%	65%
Student corrects instructor	4%	18%	31%	18%	22%	19%
Student environmental questions	6%	17%	21%	11%	12%	12%
Student administrative questions	18%	19%	21%	32%	11%	12%
Student conceptual questions	61%	62%	73%	73%	51%	57%
Epistemology and Metacognition						
Problem solving	55%	96%	87%	75%	100%	91%
Real-world examples	86%	85%	65%	61%	45%	71%
Mentions prior content	59%	74%	81%	66%	63%	70%
Mentions exam content	35%	37%	54%	16%	23%	20%
Summarizes big ideas	47%	36%	48%	46%	38%	29%
Handwritten visuals	33%	78%	100%	66%	99%	90%
Peer Interaction						
Desk work	12%	19%	4%	5%	3%	7%
Group work	51%	8%	13%	5%	8%	26%
Peer discussion w/clickers	(no data for year 1)					
Instructional Tools						
Pre-made visuals	100%	99%	73%	68%	51%	87%
Digital slides	98%	73%	465	50%	42%	80%
Pointer	76%	64%	63%	45%	40%	62%
Chalkboard or dry erase	6%	9%	19%	13%	27%	35%
Projector (i.e. document camera)	16%	55%	56%	27%	55%	42%
Instructor shows a video clip	41%	12%	4%	16%	0%	7%
Computer simulation	6%	9%	2%	30%	12%	7%
Displays a website	6%	19%	6%	34%	14%	9%
Physical demonstration	8%	24%	12%	16%	8%	55%
Clickers (Student Response Sys)	82%	12%	6%	2%	42%	58%

Table 8

Observed Instructor Practices by Department/Discipline, Years 2 and 3

	Biology	Gen Chem	Organic Chem	Engr/ ICS	Math/ Stats	Physics
Faculty-Student Interactions						
Interactive	54%	28%	44%	40%	47%	56%
Instructor display questions	77%	80%	72%	76%	94%	74%
Instructor open-ended questions	34%	13%	22%	29%	24%	13%
Instructor checks for questions	77%	65%	75%	79%	59%	69%
Formative feedback	71%	28%	31%	29%	51%	66%
Student corrects instructor	3%	15%	28%	17%	12%	20%
Student environmental questions	5%	11%	6%	7%	12%	7%
Student administrative questions	14%	13%	16%	31%	6%	8%
Student conceptual questions	69%	59%	66%	74%	49%	56%
Epistemology and Metacognition						
Problem solving	69%	96%	94%	76%	100%	90%
Real-world examples	80%	83%	59%	55%	49%	70%
Mentions prior content	51%	74%	78%	64%	63%	67%
Mentions exam content	40%	37%	50%	14%	24%	21%
Summarizes big ideas	57%	43%	47%	48%	41%	31%
Handwritten visuals	29%	78%	100%	64%	100%	90%
Peer Interaction						
Desk work	11%	19%	6%	5%	4%	8%
Group work	49%	7%	19%	7%	12%	26%
Peer discussion w/clickers	70%	2%	0%	10%	17%	39%
Instructional Tools						
Pre-made visuals	100%	98%	78%	69%	61%	85%
Digital slides	97%	72%	56%	55%	53%	77%
Pointer	77%	70%	72%	48%	47%	66%
Chalkboard or dry erase	3%	2%	19%	7%	16%	38%
Projector (i.e. document camera)	20%	59%	44%	24%	65%	44%
Instructor shows a video clip	43%	9%	6%	19%	0%	5%
Computer simulation	9%	4%	0%	26%	18%	7%
Displays a website	6%	13%	6%	33%	20%	8%
Physical demonstration	9%	19%	13%	2%	4%	59%
Clickers (Student Response Sys)	74%	6%	9%	0%	49%	52%

Table 9
Exploratory Factor Model of SPROUT Variables

Variable	Factor Loading		
	Factor 1	Factor 2	Factor 3
Instructor checks for questions	0.76		
Student conceptual questions	0.65		
Interactive (back and forth)	0.63		
Student environmental questions	0.63		
Student administrative questions	0.54		
Problem solving		0.81	
Handwritten visuals		0.79	
Mentions prior content		0.51	
Premade visuals			0.85
Pointer			0.57
Deskwork			0.48

Table 10

Enrollment Trends by Student Demographics and Factor, Years 2 and 3 (Analysis Sample)

	Factor 1	Factor 2	Factor 3
Asian	4.05	4.49	**3.12**
Hispanic	3.99	4.47	**3.15**
White	**4.08**	**4.59**	3.05
Non-resident	**4.15**	**4.78**	2.84
Declined to state	4.05	4.32	**3.24**
Other	3.97	**4.55**	**3.14**
Male	**4.14**	**4.59**	2.96
Female	3.97	4.48	**3.20**
First-generation	4.02	4.50	**3.12**
Low-income	4.00	4.49	**3.12**
Average	4.05	4.53	3.09

Note. Numbers exceeding overall average are in bold.

Table 11
Enrollment Trends by Student Demographics and Instructional Composite, Years 2 and 3 (Analysis Sample)

	Faculty-Student Interaction	Epistemological Instruction	Peer Interaction
Asian	**7.23**	7.44	**0.97**
Hispanic	7.09	7.38	**1.04**
White	**7.25**	7.44	0.89
Non-resident	**7.28**	**7.60**	0.74
Declined to state	**7.21**	7.40	**1.00**
Other	7.09	**7.53**	**0.97**
Male	**7.30**	7.41	0.83
Female	7.12	**7.49**	**1.05**
First-generation	7.15	7.44	**0.99**
Low-income	7.13	7.43	**1.00**
Average	7.20	7.45	0.95

Note. Numbers exceeding overall average are in bold.

Table 12

Association between Instructional Composites and Grade in Observed Course

	Model 1 Course Fixed Effects	Model 2 + Interaction Terms	Model 3 Student Fixed Effects
	b/se	b/se	b/se
Faculty-Student Interaction	-0.008	0.008	0.011
	(0.014)	(0.013)	(0.007)
Epistemological Instruction	0.013	0.029	0.040***
	(0.015)	(0.016)	(0.007)
Peer Interaction	-0.003	-0.004	0.009
	(0.019)	(0.021)	(0.008)
Female	-0.120***	-0.120***	0.000
	(0.009)	(0.009)	(.)
First-generation	-0.005	-0.005	0.000
	(0.007)	(0.007)	(.)
Low-income	-0.003	-0.004	0.000
	(0.007)	(0.007)	(.)
Hispanic	-0.052***	-0.053***	0.000
	(0.012)	(0.012)	(.)
White	-0.037**	-0.037**	0.313
	(0.013)	(0.013)	(0.291)
Non-resident	0.188***	0.187***	0.000
	(0.024)	(0.024)	(.)
Other*	-0.024*	-0.024*	0.194
	(0.011)	(0.011)	(0.285)
High school GPA	0.040***	0.039***	0.000
	(0.006)	(0.006)	(.)
SAT total score	0.027**	0.026**	0.000
	(0.008)	(0.008)	(.)
Student took STEM AP exam	0.040***	0.040***	0.000
	(0.012)	(0.012)	(.)
Cumulative GPA at time of observed course**	0.666***	0.666***	0.909***
	(0.019)	(0.019)	(0.017)
Female*Faculty-Student Interaction		-0.003	-0.008
		(0.009)	(0.008)
First-generation*Faculty-Student Interaction		-0.015	-0.026**
		(0.008)	(0.009)
Low-income*Faculty-Student Interaction		-0.015	-0.013

		(0.008)	(0.009)
Hispanic*Faculty-Student Interaction		-0.006	-0.018
		(0.013)	(0.010)
Female*Epistemological Instruction		-0.000	-0.006
		(0.009)	(0.008)
First-generation*Epistemological Instruction		-0.002	0.004
		(0.007)	(0.008)
Low-income*Epistemological Instruction		0.002	-0.009
		(0.009)	(0.007)
Hispanic*Epistemological Instruction		-0.014	-0.015
		(0.016)	(0.010)
Female*Peer Interaction		0.002	-0.009
		(0.009)	(0.007)
First-generation*Peer Interaction		0.011	0.017*
		(0.007)	(0.008)
Low-income*Peer Interaction		-0.014*	-0.020*
		(0.006)	(0.008)
Hispanic*Peer Interaction		0.006	0.030**
		(0.011)	(0.010)
N	44,981	44,981	35,998
R^2	0.55	0.55	0.31***

Note: * Other includes two or more races, Black, Pacific Islander, and American Indian. Asian was used as the reference group.

***Student fixed effects R^2 is the within R^2 and not comparable to an OLS R^2.

**When the observed course was taken in the first term, the student's entering GPA was substituted.

*$p < .05$. **$p < .01$. ***$p < .001$.

Table 13

Association between Instructional Composites and Grade in Follow-on Course

	Model 4 Course Fixed Effects	Model 5 +Interaction Terms	Model 6 Student Fixed Effects
	b/se	b/se	b/se
Faculty-Student Interaction	-0.010	0.003	0.016
	(0.021)	(0.027)	(0.016)
Epistemological Instruction	-0.010	-0.010	-0.009
	(0.016)	(0.021)	(0.011)
Peer Interaction	-0.006	-0.021	-0.029
	(0.019)	(0.024)	(0.016)
Female	-0.132***	-0.133***	0.000
	(0.018)	(0.018)	(.)
First-generation	-0.018	-0.018	0.000
	(0.015)	(0.015)	(.)
Low-income	0.014	0.015	0.000
	(0.014)	(0.013)	(.)
Hispanic	-0.058*	-0.062**	0.000
	(0.023)	(0.022)	(.)
Non-resident	-0.004	-0.004	0.158
	(0.019)	(0.019)	(0.089)
Other*	0.196***	0.195***	0.000
	(0.036)	(0.037)	(.)
High school GPA	0.017	0.017	0.000
	(0.019)	(0.019)	(.)
SAT total score	0.069***	0.068***	0.000
	(0.009)	(0.009)	(.)
Student took STEM AP exam	0.041**	0.041**	0.000
	(0.015)	(0.015)	(.)
Cumulative GPA at time of observed course**	-0.009	-0.009	0.000
	(0.014)	(0.014)	(.)
Female*Faculty-Student Interaction		-0.003	-0.041*
		(0.020)	(0.018)
First-generation*Faculty-Student Interaction		-0.023	-0.036
		(0.019)	(0.020)
Low-income*Faculty-Student Interaction		-0.008	-0.010
		(0.013)	(0.021)

Hispanic*Faculty-Student Interaction		0.006	0.038
		(0.020)	(0.024)
Female* Epistemological Instruction		0.004	0.005
		(0.016)	(0.011)
First-generation* Epistemological Instruction		0.013	-0.001
		(0.017)	(0.012)
Low-income* Epistemological Instruction		0.008	-0.001
		(0.014)	(0.013)
Hispanic*Epistemological Instruction		-0.054**	-0.067***
		(0.020)	(0.015)
Female*Peer Interaction		0.017	0.020
		(0.022)	(0.015)
First-generation*Peer Interaction		0.005	0.010
		(0.016)	(0.017)
Low-income*Peer Interaction		-0.017	-0.019
		(0.012)	(0.018)
Hispanic*Peer Interaction		0.032	0.024
		(0.019)	(0.022)
N	20,692	20,692	17,660
R^2	0.40	0.40	0.14***

Note: *Other includes two or more races, Black, Pacific Islander, and American Indian. Asian was used as the reference group.
***Student fixed effects R^2 is the within R^2 and not comparable to an OLS R^2.
**When the observed course was taken in the first term, the student's entering GPA was substituted.
*p < .05. **p < .01. ***p < .001.

Table 14
Association between Instructional Composites and Student Persistence

	Likelihood of Completing the Follow-on Course			
	Model 7	Model 8	Model 9	Model 10
	Analysis Sample		Series Only	
	Course Fixed Effects	+Interaction Terms	Course Fixed Effects	+Interaction Terms
	b/se	b/se	b/se	b/se
Faculty-Student Interaction	0.024	-0.027	0.196	0.164
	(0.152)	(0.143)	(0.202)	(0.185)
Epistemological Instruction	-0.005	0.052	-0.077	-0.076
	(0.109)	(0.113)	(0.080)	(0.086)
Peer Interaction	-0.025	-0.035	-0.239	-0.265
	(0.194)	(0.188)	(0.219)	(0.210)
Female	-0.010	-0.010	0.055	0.053
	(0.068)	(0.069)	(0.065)	(0.068)
First-generation	0.019	0.021	0.003	-0.004
	(0.027)	(0.025)	(0.031)	(0.027)
Low-income	-0.036	-0.035	-0.055	-0.055
	(0.026)	(0.026)	(0.029)	(0.029)
Hispanic	-0.037	-0.041	-0.161**	-0.153**
	(0.057)	(0.056)	(0.061)	(0.055)
White	-0.119*	-0.119*	-0.161*	-0.163*
	(0.060)	(0.060)	(0.070)	(0.070)
Non-resident	-0.302**	-0.307**	-0.746***	-0.751***
	(0.100)	(0.100)	(0.107)	(0.106)
Other*	0.088	0.088	0.010	0.010
	(0.053)	(0.053)	(0.057)	(0.058)
High school GPA	0.111***	0.111***	0.096***	0.097***
	(0.021)	(0.021)	(0.025)	(0.025)
Student took STEM AP exam	-0.023	-0.023	-0.057	-0.058*
	(0.029)	(0.029)	(0.030)	(0.030)
Cumulative GPA at time of observed course**	0.202***	0.203***	0.167***	0.167***
	(0.043)	(0.044)	(0.038)	(0.038)
Female*Faculty-Student Interaction		0.095		0.070
		(0.074)		(0.075)

First-generation*Faculty-Student Interaction		-0.026		-0.072*
		(0.025)		(0.032)
Low-income*Faculty-Student Interaction		0.045		0.056
		(0.027)		(0.036)
Hispanic*Faculty-Student Interaction		-0.019		0.032
		(0.052)		(0.061)
Female* Epistemological Instruction		-0.044		0.034
		(0.063)		(0.047)
First-generation* Epistemological Instruction		-0.038		-0.026
		(0.023)		(0.022)
Low-income* Epistemological Instruction		0.011		0.027
		(0.024)		(0.026)
Hispanic*Epistemological Instruction		-0.081		-0.078
		(0.051)		(0.046)
Female*Peer Interaction		-0.014		0.007
		(0.067)		(0.072)
First-generation*Peer Interaction		0.025		0.049*
		(0.025)		(0.021)
Low-income*Peer Interaction		-0.004		-0.000
		(0.023)		(0.028)
Hispanic*Peer Interaction		0.016		-0.021
		(0.047)		(0.050)
N	44,981	44,981	31,282	31,282
Pseudo R^2	0.04	0.04	0.09	0.09

Note: *Other includes two or more races, Black, Pacific Islander, and American Indian. Asian was used as the reference group.
**When the observed course was taken in the first term, the student's entering GPA was substituted.
*p < .05. **p < .01. ***p < .001.

Table 15
Association between Instructional Composites and Student Persistence

	Staying in STEM	
	Model 11	Model 12
	Course Fixed Effects	+ Interaction Terms
	b/se	b/se
Faculty-Student Interaction	0.053	0.013
	(0.045)	(0.054)
Epistemological Instruction	-0.049	-0.071
	(0.037)	(0.045)
Peer Interaction	0.009	-0.018
	(0.047)	(0.053)
Female	-0.184***	-0.185***
	(0.036)	(0.036)
First-generation	0.077**	0.078***
	(0.024)	(0.023)
Hispanic	-0.005	-0.003
	(0.024)	(0.024)
White	-0.280***	-0.277***
	(0.035)	(0.035)
Non-resident	-0.225***	-0.226***
	(0.033)	(0.033)
Other	-0.193*	-0.191*
	(0.076)	(0.076)
High school GPA	-0.123**	-0.125**
	(0.041)	(0.041)
SAT total score	0.363***	0.363***
	(0.018)	(0.018)
Student took STEM AP exam	0.416***	0.416***
	(0.020)	(0.020)
Cumulative GPA at time of observed course**	-0.038*	-0.038*
	(0.029)	(0.029)
Female*Faculty-Student Interaction		0.019
		(0.027)
First-generation*Faculty-Student Interaction		0.025
		(0.025)
Low-income*Faculty-Student Interaction		0.019
		(0.027)

Hispanic*Faculty-Student Interaction		0.042
		(0.036)
Female* Epistemological Instruction		0.048
		(0.032)
First-generation* Epistemological Instruction		-0.019
		(0.027)
Low-income* Epistemological Instruction		0.021
		(0.026)
Hispanic*Epistemological Instruction		-0.015
		(0.040)
Female*Peer Interaction		0.066*
		(0.032)
First-generation*Peer Interaction		-0.000
		(0.027)
Low-income*Peer Interaction		-0.020
		(0.025)
Hispanic*Peer Interaction		-0.026
		(0.032)
N	44,981	44,981
Pseudo R^2	0.11	0.11

Note: *Other includes two or more races, Black, Pacific Islander, and American Indian. Asian was used as the reference group.
**When the observed course was taken in the first term, the student's entering GPA was substituted.
$*p < .05$. $**p < .01$. $***p < .001$.

Table 16
Association between Individual Instructional Practices and Grades

	Model 13 Grade in Course Student Fixed Effects	Model 14 Grade in Follow-on Course Course Fixed Effects + Interaction Terms
	b/se	b/se
Faculty-Student Interactions		
Interactive	0.030***	-0.001
	(0.008)	(0.019)
Instructor display questions	0.012	0.047**
	(0.008)	(0.018)
Instructor open-ended questions	-0.003	0.002
	(0.009)	(0.015)
Instructor checks for questions	-0.043***	0.006
	(0.009)	(0.023)
Formative Feedback	0.040***	-0.011
	(0.011)	(0.022)
Student corrects instructor	0.017*	-0.018
	(0.008)	(0.015)
Student environmental questions	-0.019**	-0.025
	(0.007)	(0.016)
Student administrative questions	-0.006	0.009
	(0.008)	(0.017)
Student conceptual questions	0.008	0.002
	(0.009)	(0.019)
Epistemological Instruction		
Problem solving	0.034***	-0.005
	(0.010)	(0.019)
Real-world examples	0.050***	-0.026
	(0.009)	(0.018)
Mentions prior content	-0.008	0.016
	(0.010)	(0.021)
Mentions exam content	0.001	0.018
	(0.008)	(0.016)
Summarizes big ideas	0.008	-0.049*
	(0.008)	(0.019)
Handwritten visuals	0.028*	-0.004
	(0.011)	(0.025)

Peer Interaction		
Desk work	-0.042***	-0.022
	(0.008)	(0.014)
Group work	0.044***	-0.016
	(0.012)	(0.020)
Peer discussion w/clickers	-0.036**	-0.014
	(0.013)	(0.034)
*Instructional Tools*****		
Pre-made visuals	0.048***	0.003
	(0.010)	(0.020)
Chalkboard or dry erase	0.005	0.022
	(0.009)	(0.020)
Projector (i.e., document camera)	-0.017	0.009
	(0.009)	(0.018)
Instructor shows a video clip	0.024**	-0.033
	(0.009)	(0.017)
Computer simulation	-0.028*	-0.046*
	(0.012)	(0.018)
Displays a website	0.003	0.030
	(0.009)	(0.017)
Physical demonstration	-0.012	-0.029
	(0.010)	(0.019)
Clickers (Student Response Sys)	0.042**	0.056*
	(0.014)	(0.027)
Female	0.000	-0.158***
	(.)	(0.015)
Faculty-Student Interactions		
Interactive *Female	0.015	-0.012
	(0.008)	(0.015)
Instructor display questions*Female	0.026**	-0.031
	(0.009)	(0.016)
Instructor open-ended questions*Female	-0.006	-0.005
	(0.009)	(0.015)
Instructor checks for questions*Female	-0.029**	0.015
	(0.009)	(0.021)
Formative Feedback*Female	0.002	-0.028
	(0.011)	(0.017)
Student corrects instructor*Female	0.008	-0.021
	(0.008)	(0.017)
Student administrative questions*Female	-0.015	-0.036*
	(0.009)	(0.018)
Student environmental questions*Female	-0.000	-0.005
	(0.008)	(0.013)

Student conceptual questions*Female	-0.018	0.006
	(0.010)	(0.019)
Epistemological Instruction		
Problem solving*Female	0.010	-0.043
	(0.010)	(0.022)
Real-world examples*Female	-0.027**	-0.011
	(0.009)	(0.015)
Mentions prior content*Female	-0.005	0.028
	(0.011)	(0.017)
Mentions exam content*Female	0.004	0.033*
	(0.009)	(0.014)
Summarizes big ideas*Female	0.002	0.029
	(0.009)	(0.019)
Handwritten visuals*Female	-0.018	0.013
	(0.011)	(0.023)
Peer Interaction		
Desk work*Female	-0.001	0.038**
	(0.008)	(0.014)
Group work*Female	-0.017	-0.008
	(0.011)	(0.018)
Peer discussion w/clickers*Female	0.008	0.027
	(0.014)	(0.024)
Instructional Tools****		
Pre-made visuals*Female	0.000	0.030
	(0.012)	(0.018)
Chalkboard or dry erase*Female	-0.006	-0.004
	(0.009)	(0.016)
Projector (i.e., document camera)*Female	-0.005	0.009
	(0.010)	(0.018)
Instructor shows a video clip*Female	-0.004	-0.005
	(0.009)	(0.013)
Computer simulation*Female	0.028*	-0.005
	(0.012)	(0.015)
Displays a website*Female	0.007	-0.047**
	(0.009)	(0.014)
Physical demonstration*Female	-0.021*	-0.003
	(0.009)	(0.015)
Clickers (Student Response Sys)*Female	-0.003	0.006
	(0.014)	(0.022)
First-generation	0.000	-0.048***
	(.)	(0.012)
Faculty-Student Interactions		
Interactive*First-generation	0.010	0.010
	(0.009)	(0.011)

Instructor display questions*First-generation	0.014	-0.060***
	(0.010)	(0.012)
Instructor open-ended questions*First-generation	-0.006	-0.014
	(0.010)	(0.014)
Instructor checks for questions*First-generation	-0.026*	0.029
	(0.010)	(0.016)
Formative Feedback*First-generation	-0.010	-0.004
	(0.013)	(0.016)
Student corrects instructor*First-generation	0.003	-0.018
	(0.010)	(0.014)
Student administrative questions*First-generation	-0.015	-0.029*
	(0.010)	(0.012)
Student environmental questions*First-generation	0.012	-0.014
	(0.008)	(0.009)
Student conceptual questions*First-generation	-0.027*	-0.018
	(0.011)	(0.012)
Epistemological Instruction		
Problem solving*First-generation	-0.006	-0.028*
	(0.011)	(0.012)
Real-world examples*First-generation	0.005	0.027*
	(0.010)	(0.012)
Mentions prior content*First-generation	-0.002	0.016
	(0.012)	(0.013)
Mentions exam content*First-generation	-0.006	0.006
	(0.010)	(0.013)
Summarizes big ideas*First-generation	0.017	0.023
	(0.010)	(0.015)
Handwritten visuals*First-generation	-0.009	-0.028
	(0.012)	(0.016)
Peer Interaction		
Desk work*First-generation	0.009	0.019
	(0.009)	(0.011)
Group work*First-generation	0.016	0.018
	(0.013)	(0.014)
Peer discussion w/clickers*First-generation	0.002	0.043*
	(0.016)	(0.017)
Instructional Tools[****]		
Pre-made visuals*First-generation	-0.046***	-0.012
	(0.013)	(0.016)
Chalkboard or dry erase*First-generation	-0.009	0.026
	(0.010)	(0.022)
Projector (i.e., document camera)*First-generation	-0.001	0.030**
	(0.011)	(0.011)
Instructor shows a video clip*First-generation	0.003	0.030**
	(0.011)	(0.011)

Computer simulation*First-generation	-0.011	0.044**
	(0.013)	(0.016)
Displays a website*First-generation	0.020	0.002
	(0.011)	(0.012)
Physical demonstration*First-generation	0.013	0.020
	(0.011)	(0.013)
Clickers (Student Response Sys)*First-generation	-0.011	-0.079***
	(0.016)	(0.021)
Low-income	0.000	-0.009
	(.)	(0.013)
Faculty-Student Interactions		
Interactive*Low-income	-0.013	-0.028
	(0.009)	(0.014)
Instructor display questions*Low-income	-0.007	0.033*
	(0.010)	(0.015)
Instructor open-ended questions*Low-income	-0.012	0.009
	(0.011)	(0.012)
Instructor checks for questions*Low-income	-0.008	-0.018
	(0.011)	(0.015)
Formative Feedback*Low-income	0.016	-0.017
	(0.013)	(0.017)
Student corrects instructor*Low-income	0.007	0.025
	(0.010)	(0.018)
Student administrative questions*Low-income	0.009	0.018
	(0.011)	(0.015)
Student environmental questions*Low-income	0.001	0.023*
	(0.009)	(0.011)
Student conceptual questions*Low-income	0.026*	-0.007
	(0.012)	(0.014)
Epistemological Instruction		
Problem solving*Low-income	-0.013	-0.017
	(0.011)	(0.012)
Real-world examples*Low-income	0.001	-0.006
	(0.010)	(0.015)
Mentions prior content*Low-income	-0.015	0.022
	(0.013)	(0.016)
Mentions exam content*Low-income	0.013	-0.003
	(0.010)	(0.014)
Summarizes big ideas*Low-income	-0.008	-0.015
	(0.010)	(0.016)
Handwritten visuals*Low-income	0.021	-0.010
	(0.012)	(0.017)

Peer Interaction		
Desk work*Low-income	-0.006	-0.013
	(0.010)	(0.012)
Group work*Low-income	0.001	-0.022
	(0.013)	(0.017)
Peer discussion w/clickers*Low-income	-0.015	0.004
	(0.016)	(0.020)
*Instructional Tools*****		
Pre-made visuals*Low-income	-0.005	0.038
	(0.014)	(0.023)
Chalkboard or dry erase*Low-income	-0.017	-0.034*
	(0.011)	(0.017)
Projector (i.e., document camera)*Low-income	-0.004	-0.024
	(0.011)	(0.016)
Instructor shows a video clip*Low-income	-0.003	-0.026*
	(0.011)	(0.013)
Computer simulation*Low-income	0.013	-0.017
	(0.013)	(0.014)
Displays a website*Low-income	0.015	0.011
	(0.011)	(0.020)
Physical demonstration*Low-income	0.001	0.007
	(0.011)	(0.016)
Clickers (Student Response Sys)*Low-income	-0.017	0.006
	(0.016)	(0.023)
Hispanic	0.000	-0.094***
	(.)	(0.019)
Faculty-Student Interactions		
Interactive*Hispanic	0.013	0.026
	(0.011)	(0.018)
Instructor display questions*Hispanic	0.013	0.007
	(0.012)	(0.017)
Instructor open-ended questions*Hispanic	-0.005	0.026
	(0.012)	(0.016)
Instructor checks for questions*Hispanic	-0.032**	-0.049*
	(0.011)	(0.020)
Formative Feedback*Hispanic	-0.023	0.063**
	(0.015)	(0.021)
Student corrects instructor*Hispanic	0.007	-0.003
	(0.011)	(0.021)
Student administrative questions*Hispanic	0.013	-0.063**
	(0.012)	(0.020)
Student environmental questions*Hispanic	0.009	-0.043***
	(0.010)	(0.012)

Student conceptual questions*Hispanic	-0.021	-0.031
	(0.012)	(0.016)
Epistemological Instruction		
Problem solving*Hispanic	-0.022	0.010
	(0.013)	(0.016)
Real-world examples*Hispanic	0.008	0.006
	(0.011)	(0.018)
Mentions prior content*Hispanic	0.031*	0.041
	(0.015)	(0.024)
Mentions exam content*Hispanic	-0.015	-0.008
	(0.012)	(0.017)
Summarizes big ideas*Hispanic	-0.002	-0.011
	(0.012)	(0.015)
Handwritten visuals*Hispanic	-0.034*	-0.059*
	(0.014)	(0.024)
Peer Interaction		
Desk work*Hispanic	0.009	-0.003
	(0.011)	(0.012)
Group work*Hispanic	0.021	0.003
	(0.015)	(0.022)
Peer discussion w/clickers*Hispanic	0.005	0.002
	(0.018)	(0.027)
Instructional Tools[****]		
Pre-made visuals*Hispanic	-0.003	-0.042*
	(0.016)	(0.020)
Chalkboard or dry erase*Hispanic	0.002	-0.045
	(0.012)	(0.031)
Projector (i.e., document camera)*Hispanic	0.015	0.045*
	(0.012)	(0.018)
Instructor shows a video clip*Hispanic	-0.000	0.012
	(0.012)	(0.023)
Computer simulation*Hispanic	-0.003	0.000
	(0.014)	(0.019)
Displays a website*Hispanic	-0.006	-0.039
	(0.012)	(0.024)
Physical demonstration*Hispanic	-0.019	-0.008
	(0.012)	(0.020)
Clickers (Student Response Sys)*Hispanic	0.023	-0.008
	(0.018)	(0.025)
High school GPA	0.000	0.055***
	(.)	(0.009)
SAT total score	0.000	0.034*
	(.)	(0.014)

Student took STEM AP exam	0.000	-0.049***
	(.)	(0.014)
Cumulative GPA at time of observed course**	0.887***	0.631***
	(0.016)	(0.031)
N	35,998	21,060
R^2	0.33***	0.41

Note: *Other includes two or more races, Black, Pacific Islander, and American Indian. Asian was used as the reference group.
**When the observed course was taken in the first term, the student's entering GPA was substituted.
***Student fixed effects R^2 is the within R^2 and not comparable to an OLS R^2.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 17
Association between Individual Instructional Practices and Student Persistence

	Model 15	Model 16
	Completing Follow-on Course Series Only	Staying in STEM
	Course Fixed Effects + Interaction Terms	
	b/se	b/se
Faculty-Student Interactions		
Interactive	-0.234*	-0.079
	(0.114)	(0.052)
Instructor display questions	-0.156	0.016
	(0.130)	(0.051)
Instructor open-ended questions	0.013	-0.046
	(0.130)	(0.053)
Instructor checks for questions	0.417*	0.059
	(0.200)	(0.054)
Formative Feedback	0.146	-0.024
	(0.175)	(0.060)
Student corrects instructor	-0.025	0.045
	(0.103)	(0.046)
Student environmental questions	0.357*	0.078
	(0.152)	(0.042)
Student administrative questions	0.112	-0.018
	(0.105)	(0.046)
Student conceptual questions	0.341*	0.041
	(0.135)	(0.055)
Epistemological Instruction		
Problem solving	-0.198	-0.223***
	(0.151)	(0.051)
Real-world examples	0.133	-0.059
	(0.130)	(0.051)
Mentions prior content	0.115	0.008
	(0.158)	(0.057)
Mentions exam content	0.268*	0.090
	(0.121)	(0.056)
Summarizes big ideas	0.067	-0.078
	(0.125)	(0.047)
Handwritten visuals	-0.201	0.127*
	(0.171)	(0.059)

Peer Interaction		
Desk work	0.014	0.028
	(0.115)	(0.042)
Group work	-0.159	0.033
	(0.171)	(0.052)
Peer discussion w/clickers	-0.389	-0.065
	(0.200)	(0.060)
*Instructional Tools*****		
Pre-made visuals	0.191	0.127*
	(0.157)	(0.062)
Chalkboard or dry erase	0.205	-0.056
	(0.141)	(0.043)
Projector (i.e., document camera)	-0.009	-0.066
	(0.129)	(0.058)
Instructor shows a video clip	0.042	0.102*
	(0.132)	(0.046)
Computer simulation	-0.268*	-0.127**
	(0.126)	(0.041)
Displays a website	-0.058	-0.132**
	(0.126)	(0.043)
Physical demonstration	0.099	0.028
	(0.114)	(0.044)
Clickers (Student Response Sys)	0.062	0.065
	(0.188)	(0.073)
Female	-0.002	-0.190***
	(0.056)	(0.029)
Faculty-Student Interactions		
Interactive *Female	-0.002	-0.098**
	(0.053)	(0.034)
Instructor display questions*Female	0.027	0.032
	(0.072)	(0.039)
Instructor open-ended questions*Female	-0.032	0.040
	(0.070)	(0.034)
Instructor checks for questions*Female	0.065	0.031
	(0.069)	(0.032)
Formative Feedback*Female	-0.106	0.045
	(0.073)	(0.040)
Student corrects instructor*Female	0.052	-0.013
	(0.064)	(0.035)
Student administrative questions*Female	-0.153**	0.025
	(0.058)	(0.036)
Student environmental questions*Female	-0.162***	0.035
	(0.049)	(0.030)

Student conceptual questions*Female	-0.080	0.056
	(0.080)	(0.035)
Epistemological Instruction		
Problem solving*Female	0.097	0.057
	(0.092)	(0.041)
Real-world examples*Female	0.050	0.045
	(0.080)	(0.032)
Mentions prior content*Female	-0.110	-0.010
	(0.092)	(0.040)
Mentions exam content*Female	-0.096	-0.044
	(0.061)	(0.032)
Summarizes big ideas*Female	-0.013	-0.050
	(0.078)	(0.039)
Handwritten visuals*Female	-0.255**	-0.034
	(0.094)	(0.046)
Peer Interaction		
Desk work*Female	0.150*	0.034
	(0.069)	(0.028)
Group work*Female	0.050	0.022
	(0.082)	(0.038)
Peer discussion w/clickers*Female	0.145	0.070
	(0.106)	(0.048)
Instructional Tools[****]		
Pre-made visuals*Female	0.037	0.001
	(0.095)	(0.051)
Chalkboard or dry erase*Female	0.121	0.013
	(0.089)	(0.036)
Projector (i.e., document camera)*Female	0.270**	0.002
	(0.085)	(0.038)
Instructor shows a video clip*Female	-0.152*	-0.029
	(0.060)	(0.038)
Computer simulation*Female	0.083	-0.026
	(0.068)	(0.034)
Displays a website*Female	0.023	0.006
	(0.081)	(0.036)
Physical demonstration*Female	0.014	0.096**
	(0.068)	(0.031)
Clickers (Student Response Sys)*Female	-0.188	-0.129**
	(0.102)	(0.048)
First-generation	0.037	0.090***
	(0.027)	(0.023)
Faculty-Student Interactions		
Interactive*First-generation	-0.031	0.013
	(0.033)	(0.027)

Instructor display questions*First-generation	-0.034	-0.006
	(0.046)	(0.028)
Instructor open-ended questions*First-generation	-0.035	-0.068*
	(0.029)	(0.034)
Instructor checks for questions*First-generation	-0.035	-0.012
	(0.045)	(0.027)
Formative Feedback*First-generation	-0.047	0.017
	(0.042)	(0.038)
Student corrects instructor*First-generation	0.070	0.036
	(0.049)	(0.026)
Student administrative questions*First-generation	-0.021	0.026
	(0.042)	(0.034)
Student environmental questions*First-generation	-0.032	0.056
	(0.029)	(0.032)
Student conceptual questions*First-generation	-0.041	0.036
	(0.036)	(0.029)
Epistemological Instruction		
Problem solving*First-generation	0.038	-0.001
	(0.045)	(0.036)
Real-world examples*First-generation	-0.069*	-0.045
	(0.030)	(0.026)
Mentions prior content*First-generation	-0.005	0.004
	(0.048)	(0.030)
Mentions exam content*First-generation	-0.009	0.019
	(0.033)	(0.036)
Summarizes big ideas*First-generation	-0.063	-0.067*
	(0.037)	(0.032)
Handwritten visuals*First-generation	-0.116*	-0.107**
	(0.045)	(0.036)
Peer Interaction		
Desk work*First-generation	0.045	0.056*
	(0.047)	(0.024)
Group work*First-generation	-0.031	-0.086***
	(0.030)	(0.022)
Peer discussion w/clickers*First-generation	0.067	0.121***
	(0.047)	(0.033)
Instructional Tools[****]		
Pre-made visuals*First-generation	0.055	-0.002
	(0.046)	(0.035)
Chalkboard or dry erase*First-generation	0.018	-0.027
	(0.046)	(0.026)
Projector (i.e., document camera)*First-generation	0.007	-0.013
	(0.048)	(0.030)
Instructor shows a video clip*First-generation	-0.011	-0.054
	0.043	0.030

Computer simulation*First-generation	(0.027)	(0.026)
	0.081*	0.079*
Displays a website*First-generation	(0.041)	(0.032)
	0.010	0.003
Physical demonstration*First-generation	(0.032)	(0.024)
	0.052	0.004
Clickers (Student Response Sys)*First-generation	(0.056)	(0.050)
	0.043	0.030
Low-income	0.032	0.029
	(0.026)	(0.022)
Faculty-Student Interactions		
Interactive*Low-income	0.087**	0.050
	(0.029)	(0.028)
Instructor display questions*Low-income	-0.069*	-0.012
	(0.029)	(0.029)
Instructor open-ended questions*Low-income	0.024	0.009
	(0.035)	(0.027)
Instructor checks for questions*Low-income	-0.001	0.013
	(0.037)	(0.028)
Formative Feedback*Low-income	0.033	-0.015
	(0.039)	(0.040)
Student corrects instructor*Low-income	-0.050	-0.014
	(0.032)	(0.027)
Student administrative questions*Low-income	-0.046	-0.002
	(0.036)	(0.030)
Student environmental questions*Low-income	0.030	-0.033
	(0.028)	(0.022)
Student conceptual questions*Low-income	-0.027	-0.024
	(0.043)	(0.034)
Epistemological Instruction		
Problem solving*Low-income	-0.044	0.060
	(0.046)	(0.031)
Real-world examples*Low-income	0.038	0.002
	(0.039)	(0.028)
Mentions prior content*Low-income	0.069	0.023
	(0.045)	(0.034)
Mentions exam content*Low-income	-0.026	-0.015
	(0.030)	(0.031)
Summarizes big ideas*Low-income	-0.079*	-0.016
	(0.035)	(0.029)
Handwritten visuals*Low-income	-0.000	-0.040
	(0.040)	(0.037)

Peer Interaction		
Desk work*Low-income	0.013	-0.044
	(0.037)	(0.023)
Group work*Low-income	-0.012	0.000
	(0.036)	(0.037)
Peer discussion w/clickers*Low-income	0.076	-0.036
	(0.063)	(0.046)
*Instructional Tools*****		
Pre-made visuals*Low-income	0.008	0.033
	(0.036)	(0.026)
Chalkboard or dry erase*Low-income	-0.001	0.067
	(0.044)	(0.039)
Projector (i.e., document camera)*Low-income	-0.145***	-0.054
	(0.036)	(0.036)
Instructor shows a video clip*Low-income	0.023	-0.009
	(0.033)	(0.027)
Computer simulation*Low-income	0.044	0.019
	(0.041)	(0.027)
Displays a website*Low-income	0.045	-0.008
	(0.034)	(0.023)
Physical demonstration*Low-income	0.105*	0.095**
	(0.042)	(0.035)
Clickers (Student Response Sys)*Low-income	-0.143***	0.026
	(0.042)	(0.042)
Hispanic	-0.010	-0.196***
	(0.047)	(0.030)
Faculty-Student Interactions		
Interactive*Hispanic	0.163**	0.038
	(0.057)	(0.037)
Instructor display questions*Hispanic	0.073	-0.009
	(0.060)	(0.038)
Instructor open-ended questions*Hispanic	0.002	-0.058
	(0.038)	(0.034)
Instructor checks for questions*Hispanic	-0.209***	-0.054
	(0.062)	(0.038)
Formative Feedback*Hispanic	-0.006	0.078
	(0.061)	(0.046)
Student corrects instructor*Hispanic	-0.073	0.051
	(0.056)	(0.039)
Student administrative questions*Hispanic	-0.102	0.031
	(0.060)	(0.035)
Student environmental questions*Hispanic	0.040	-0.013
	(0.042)	(0.033)

Student conceptual questions*Hispanic	-0.033	-0.033
	(0.059)	(0.042)
Epistemological Instruction		
Problem solving*Hispanic	-0.158*	-0.104*
	(0.065)	(0.044)
Real-world examples*Hispanic	0.037	-0.027
	(0.058)	(0.035)
Mentions prior content*Hispanic	0.054	0.077
	(0.069)	(0.041)
Mentions exam content*Hispanic	-0.060	0.008
	(0.045)	(0.039)
Summarizes big ideas*Hispanic	-0.011	-0.076*
	(0.055)	(0.037)
Handwritten visuals*Hispanic	0.041	-0.043
	(0.068)	(0.042)
Peer Interaction		
Desk work*Hispanic	-0.044	0.090*
	(0.051)	(0.036)
Group work*Hispanic	-0.002	-0.042
	(0.051)	(0.046)
Peer discussion w/clickers*Hispanic	0.041	0.056
	(0.084)	(0.059)
Instructional Tools[****]		
Pre-made visuals*Hispanic	-0.135*	-0.023
	(0.053)	(0.042)
Chalkboard or dry erase*Hispanic	-0.143*	-0.022
	(0.071)	(0.037)
Projector (i.e., document camera)*Hispanic	-0.091	-0.015
	(0.062)	(0.036)
Instructor shows a video clip*Hispanic	-0.023	-0.087*
	(0.047)	(0.040)
Computer simulation*Hispanic	0.119*	0.025
	(0.050)	(0.028)
Displays a website*Hispanic	0.058	0.065
	(0.059)	(0.034)
Physical demonstration*Hispanic	0.077	0.076*
	(0.056)	(0.037)
Clickers (Student Response Sys)*Hispanic	-0.040	-0.064
	(0.093)	(0.064)
High school GPA	0.161***	0.385***
	(0.023)	(0.018)
SAT total score	-0.018	0.424***
	(0.028)	(0.020)

Student took STEM AP exam	0.326***	0.468***
	(0.043)	(0.034)
Cumulative GPA at time of observed course**	0.657***	-0.037*
	(0.047)	(0.018)
N	31,282	44,981
Pseudo R^2	0.15	0.10

Note: *Other includes two or more races, Black, Pacific Islander, and American Indian. Asian was used as the reference group.
**When the observed course was taken in the first term, the student's entering GPA was substituted.
*$p < .05$. **$p < .01$. ***$p < .001$.

APPENDIX B

			AY 2014-2015					AY 2015-2016				
			Fall 2014	Winter 2015	Spring 2015		**Complete and Re-checked**	Fall 2015	Winter 2016	Spring 2016		**Complete and Re-checked**
Ord Term			202	203	204			206	207	208		
OIR Current Qtr Data	Imported and Cleaned	tab 1	done	done	done		**3.23.17**	done	done	done		**3.25.17**
		tab 2	done	done	done		**3.23.17**	done	done	done		**3.25.17**
		tab 3	done	done	done		**3.23.17**	done	done	done		**3.25.17**
		tab 4	done	done	done		**3.23.17**	done	done	done		**3.25.17**
fcref Do File	Create Focal Course Reference Number		done	done	done		**1.7.17**	done	done	done		**1.7.17**
Next Courses	All Identified, including cross-listed		done	done	done		**1.21.17**	done	done	done		**1.21.17**
fc_next Do File	Create Next Course Reference Number to merge		done	done	done		**1.30.17**	done	done	done		**1.30.17**
fc_seq_LR	Prioritize Next Course		done	done	done		**1.30.17**	done	done	done		**1.30.17**
OIR Next Qtr Data	Imported and Cleaned	tab 4	done	done	done		**3.23.17**	done	done	done		**3.25.17**
OIR Next Qtr Data	Imported and Cleaned from previous cohorts not currently observed	tab 5	done	done	done		**3.23.17**	done	done	done		**3.25.17**
Course & Instructor Audit	Complete by Qtr, including mult. Sections and correctly labeling each section		done	done	done		**3.17.17**	done	done	done		**3.17.17**
WIDER Course Code Protocol	Protocol updated with additional insructors, courses, and courses taught by mult. Instructors		done	done	done		**3.17.17**	done	done	done		**3.17.17**
WIDER_coursecode Do File	Complete by Qtr		done	done	done		**3.17.17**	done	done	done		**3.17.17**
STEM Majors Do File	Complete by Qtr		done	done	done		**2.5.17**	done	done	done		**2.5.17**
LR_gradepts Do File			done	done	done			done	done	done		**3.23.17**
Master Raw OIR Data File	Complete by QTR		done	done	done		**3.23.17**	done	done	done		**3.25.17**
OP Data File	Clean and Complete by Qtr		done	done	done		**3.23.17**	done	done	done		**3.28.17**
OIR_OP_LR Do File	Complete by Qtr		done	done	done		**3.23.17**	done	done	done		**3.28.17**
	OIR and OP merged and cleaned		done	done	done		**3.23.17**	done	done	done		**3.28.17**
Append	Year 2						**3.28.17**					
	Year 3											**3.28.17**
	Year 2 and 3 together											**3.28.17**
Check numbers (audit data)	Years 2 and 3											**3.31.17**
Clean combined data	Years 2 and 3											**3.31.17**
Create necessary variables	Years 2 and 3											**3.31.17**
Run Descriptives	Years 2 and 3											**3.31.17**
Run Analyses	Years 2 and 3											**3.31.17**
Data Tables	Years 2 and 3											**3.31.17**
Master OIR_OP Data File	**Years 2 and 3 Complete**											**3.31.17**

APPENDIX C

Simple Protocol for Observing Undergraduate Teaching (SPROUT)	
SPROUT is designed to give an objective measure of instructional practices in undergraduate STEM courses. It contains a series of dichotomous questions (yes or no) that note the presence or absence of certain occurrences. SPROUT also includes opportunities for the researcher to include qualitative evidence supporting the observation of instructional practices. This material is based upon work supported by the National Science Foundation under Grant Number 1256500.	
I. BACKGROUND INFORMATION	
Lecturer:	
Date of Observation:	
Start & End Time of Observation:	
Wave:	
Course Name:	
Course Code:	
Observer:	
Location:	
Current Total Enrollment:	
Approx. Attendance at Lecture:	
Seat Location of Observation:	
II. LESSON OVERVIEW	
(a) Lesson Description:	
Evidence:	

(b) Describe Faculty-Student Interaction:				
Evidence:				
(c) Describe Peer Interactions (If Any):				
Evidence:				
(d) Describe Problem-Solving (If Any):				
Evidence:				
(e) Describe the resources used by the instructor and by the students (including technology):				
Evidence:				
III. TEACHING DIMENSIONS				
(a) Teaching Methods				
1. Lecture without Visuals:				
		Yes		No
Evidence:				
2. Lecture with Pre-Made Visuals:				
		Yes		No
Evidence:				
3. Lecture with Handwritten Visuals:				
		Yes		No
Evidence:				
4. Lecturing with Demonstration of Topic or Phenomena:				
		Yes		No
Evidence:				
5. Interactive Lecture:				
		Yes		No

Evidence:

a. If yes, how many instances?	

6. Deskwork:

	Yes		No

Evidence:

7. Groupwork:

	Yes		No

Evidence:

8. Student Presentation:

	Yes		No

Evidence:

9. Does the instructor solve problems in front of the class?

	Yes		No

Evidence:

10. Does the instructor warn the class about common mistakes/misconceptions?

	Yes		No

Evidence:

a. If yes, how many times?	

(b) Pedagogical Moves Observed:

1. Illustration with the real world:

	Yes		No

Evidence:
2. References prior course content: ☐ Yes ☐ No
Evidence:
3. Assessment: ☐ Yes ☐ No
Evidence:
4. Formative feedback (iClickers, handwritten, or definite show of hands): ☐ Yes ☐ No
Evidence:
a. If yes, did the instructor modify the lesson accordingly as a result of this feedback? ☐ Yes ☐ No ☐ N/A
Evidence:
5. Does the instructor mention specifically what students need to know for the test or exam? ☐ Yes ☐ No
Evidence:
6. Does the instructor summarize the ideas presented in lecture? ☐ Yes ☐ No
Evidence:
(c) Instructor/Student Interactions Observed (Types of Q & A)

1. Instructor display question:

	Yes		No

If yes, note which of the following types of student responses to instructor occurred:

	Fixed Values
	Yes One Student Responds
	Yes Collectively Students Respond
	Yes No One Responds

Evidence:

2. Instructor open-ended question:

	Yes		No

If yes, note which of the following types of student responses to instructor occurred:

	Fixed Values
	Yes One Student Responds
	Yes Collectively Students Respond
	Yes No One Responds

Evidence:

3. Instructor checks for student understanding:

	Yes		No

If yes, note which of the following types of student responses to instructor occurred:

	Fixed Values
	Yes One Student Responds
	Yes Collectively Students Respond
	Yes No One Responds

Evidence:

4. Student corrects instructor:

	Yes		No

Evidence:

5. Student administrative question:

	Yes		No

Evidence:

6. Student classroom environmental question:

	Yes		No

Evidence:

7. Student conceptual question:

	Yes		No

Evidence:

(d) Observed Student Behavior at Two Time Points (observers randomly)
After 20 Minutes
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
After 40 Minutes
1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
(e) Instructional Technology
1. Book(s): ☐ Yes ☐ No

2. Pointer: ☐ Yes ☐ No
3. Chalk-board/White-board: ☐ Yes ☐ No
4. Overhead: ☐ Yes ☐ No
5. PowerPoint or Other Digital Slides: ☐ Yes ☐ No
6. Clickers: ☐ Yes ☐ No

If yes:	
a. How many questions?	
b. What was the format of the questions? (Multiple Choice, True/False, or Both)	
c. What were the types of questions? (Content-Recall, Calculation Based, or Both)	
d. Were students given a second chance to answer the question?	

	Yes		No		N/A

e. Were students encouraged to discuss the clicker question with their peers?

	Yes		No		N/A

7. Demonstration Equipment:

	Yes		No

8. Movie, Documentary, Video Clips, or YouTube Video:

	Yes		No

9. Calculator:

	Yes		No

10. Simulations:

	Yes		No

11. Website:

	Yes		No

IV. SUMMARY COMMENTS