Making Undergraduate STEM Education more Inclusive, Interpersonal, and Interdisciplinary through Challenge-Based Learning

By

Haider Ali Bhatti

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

Science & Math Education

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Robert J. Full, Co-Chair
Professor Marcia C. Linn, Co-Chair
Professor Mark Wilson

Summer 2024

Abstract

Making Undergraduate STEM Education more
Inclusive, Interpersonal, and Interdisciplinary
through Challenge-Based Learning

By

Haider Ali Bhatti

Doctor of Philosophy in Science & Math Education

University of California, Berkeley

Professor Robert J. Full, Co-Chair
Professor Marcia C. Linn, Co-Chair

The increasing complexity of global challenges demands a STEM-enriched approach to learning for all students, regardless of their future career paths. Challenge-Based Learning (CBL) is a pedagogical method to foster a STEM-enriched education, engaging students in the design of societally impactful, interdisciplinary solutions. To investigate the potential of CBL, specifically in the context of Undergraduate STEM Education (USE), it is crucial to assess students' affective development such as their attitudes, beliefs, and self-perceptions related to STEM. This dissertation explores the impact of CBL on student affect through three interconnected studies centered on a large-enrollment *Bioinspired Design* course. Chapter 1 explores overall growth in measures of science connection—Science Identity (SciID), Science Self-Efficacy (Eff), and Internalization of Scientific Community Values (Val)—using the Tripartite Integration Model of Social Influence (TIMSI) framework. Results demonstrated significant pre/post increases in SciID and Eff across five semesters, with Val remaining stable. Item level analyses revealed specific impacts of CBL activities on these affective measures, particularly in developing students' confidence in creating novel technologies. Chapter 2 investigates the equity of these affective growth outcomes across seven demographic variables. Results indicated that the observed increases in science connection were largely equitable across diverse student populations, with differences in SciID development based on STEM major status and class status. Chapter 3 introduces and validates a novel affective construct: *Innovation Skills* self-efficacy. Developed using the Berkeley Evaluation & Assessment Research (BEAR) Assessment System, this construct provides a more targeted measure of self-efficacy aligned with the Innovation Skills needed for the future STEM-enriched workforce. Results showed approximately one standard deviation of pre/post growth, with a large effect size in the context of educational interventions. Collectively, this dissertation showcases the potential of CBL approaches in USE to foster equitable development of science connection and Innovation Skills self-efficacy across diverse student populations through comprehensive, psychometrically robust assessments of student affect. This research underscores the importance of holistic approaches to STEM education that cultivate not only knowledge and skills, but also the attitudes and beliefs necessary for success in the known and unknown STEM-enriched careers of the future.

Table of Contents

Acknowledgments

I am guided by Islam in all parts of my life. Thus, I begin by acknowledging Allah (God) through the phrase الحمد لله (*alhamdulillah*), meaning “All praises be to Allah.”

This dissertation is dedicated to Ami, my late mother (Nabila Bhatti). I thank her for all the lessons she taught me and making me into the person I am today. I love you, Ami.

I thank my wife, Zarra, for being the most loving, caring, supportive, and fun life partner I could ever ask for. You are the love of my life, Zarra.

I thank Abu Jaan, my dad (Obaid Bhatti), for his love and support throughout my life, especially after Ami passed away. He is the strongest man I know. I also thank my sisters, Api and Fatima, for always being by my side. I simply cannot put into words the impact my family has had on me. I love and thank you all so much.

Thank you to all my friends for the hangouts, group chats, debates, sports talk, and gaming that got me through all the tough days after losing Ami.

Thank you to my advisor, Bob Full. I’ll never forget when he called me before I committed to Berkeley and asked, “Why haven’t you accepted our offer yet?” In that moment, he made it clear to me that I would be supported and that I belonged at Berkeley. He was so right. I am a better thinker, researcher, and teacher because of him.

Thank you to Marcia Linn, my first connection to Berkeley. While on a work trip to the Bay Area, she invited me to join her research group. Her knowledge, kindness, and welcoming persona all made me think, “I want to be here.” Through her mentorship, I’ve learned what it means to be someone truly devoted to improving science education for all students.

Thank you to Mark Wilson, a mentor who helped me find my identity as an education researcher through his classes and ongoing guidance. He helped me craft a research path that I am so deeply passionate about and look forward to further developing.

Thank you to Mica Estrada and Perman Gochyyev for all their expertise, feedback, and guidance. This research would not have been possible without them.

My doctoral journey and the research in this dissertation was funded by a Howard Hughes Medical Institute grant (GT1047) awarded to my advisor. I also received support through fellowships from the Berkeley School of Education, the Gardner Seminar Higher Education Fellowship, and the Diversity & Community Fellowship through the Office for Graduate Diversity. I am grateful for all the financial support I received.

Positionality Statement

I am a practicing Muslim and firmly believe in the mutually reinforcing nature of Islam and science. I come from an immigrant and first-generation background. I was born in Pakistan to parents without college degrees. I come from a background in which no one I ever knew pursued a PhD. As an undergraduate, pursuing a biology degree felt easy for me (aside from organic chemistry) because I had been so well prepared by my high school, Bergen County Academies. But the same could not be said for so many of my peers. I saw students drop out of introductory STEM courses like Gen Bio, Gen Chem, and Gen Physics as part of what seemed like an embedded “weed out” process. These same peers often left their pursuit of a STEM degree altogether. This prompted my initial motivation to investigate pedagogical practices within undergraduate STEM education. From there, I became interested in improving student learning through disciplinary-based education research.

Eventually, I found my way to the Science and Math Education (SESAME) program at Berkeley. I was excited by the interdisciplinary nature of the program that emphasized both STEM and the learning/teaching of STEM through evidence-based practices. As a SESAME student, I was introduced to the interdisciplinary subject of *Bioinspired Design*. I saw firsthand just how impactful an undergraduate course based on this unique context can be for all students. Barriers of STEM versus non-STEM were erased. Learning was a collaborative endeavor. Learning was built on authentic discovery. Learning involved making something that was personally meaningful and societally relevant.

After seeing the promise of this innovative course, I personally became interested in researching the assessment of student learning within this course. I began to consider questions such as: How do we know this course is actually working? Is it working for all students? What can we measure to support this? Why should we measure that and not something else? How should we measure this? Based on my interests in these questions, I present the following dissertation on how we can make undergraduate STEM education more inclusive, interpersonal, and interdisciplinary through Challenge-Based Learning.

Dissertation Introduction

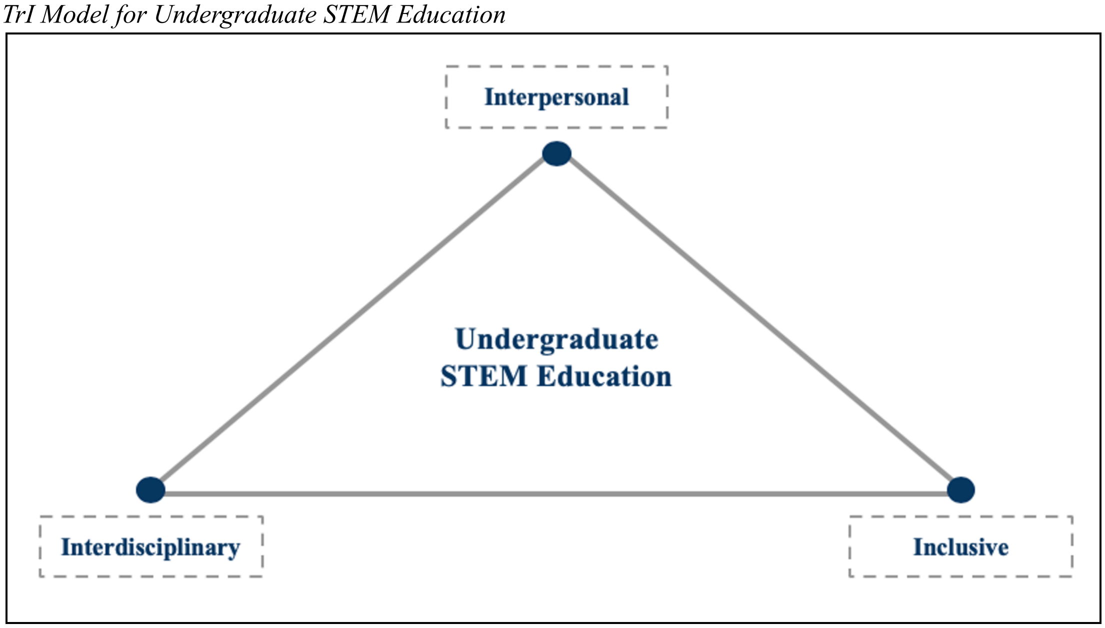
Reimagining Undergraduate STEM Education through the TrI Model

The demand for innovative, interdisciplinary solutions to global challenges is rapidly changing the landscape of science, technology, engineering, and mathematics (STEM). To meet these challenges, we must fundamentally reimagine undergraduate STEM education (USE) to better prepare the next generation of informed citizens. This reimagination requires a shift towards a *STEM-enriched* approach to learning, where scientific knowledge and practices are applied not just within traditional STEM fields, but across all areas of society. This outlook recognizes that all students, regardless of their major, benefit from developing STEM-enriched perspectives to make informed decisions as part of their civic duties in a democratic society.

Considering the evolving demands of an ever-changing world, many of the jobs of the future do not currently exist. This poses a unique challenge—how can we prepare undergraduates for an uncertain future? Moreover, how can we ensure that USE meets the needs of an increasingly diverse student population while fostering the skills necessary to tackle pressing global issues? These questions underscore the need for a USE system that is not confined to STEM majors alone, but rather enriches all students with STEM-related skills and knowledge applicable to areas both within and outside of STEM.

To meet these challenges, the research in this dissertation is centered on a reimagination of USE based on the Tr**I** model (Figure 1). The future of USE must be **I**nclusive, **I**nterpersonal, and **I**nterdisciplinary. These three principles form the foundation of a STEM education that is effective in developing technical skills while also being more equitable, collaborative, and reflective of scientific practices that promote innovation. By embracing these principles within a STEM-enriched framework, we can create a USE system that truly serves all students, empowering them to apply scientific thinking and problem-solving skills across diverse contexts, thereby advancing the frontiers of scientific knowledge and societal progress.

Figure 1
TrI Model for Undergraduate STEM Education



The future of USE must be Inclusive. A student cannot adequately learn if they do not feel like they belong, and historically, STEM education has not been a place where marginalized groups have felt welcome. Scientific investigations require the voices of diverse stakeholders to develop creative solutions to complex problems. Therefore, it is our responsibility to develop a USE system that is not only diverse, but also inclusive, ensuring that those diverse perspectives can flourish. This inclusivity is central to a STEM-enriched education, recognizing that all students, regardless of their major or background, can contribute valuable insights to scientific problem-solving. Fostering this environment requires culturally-sustaining research practices that better integrate how students' backgrounds and lived experiences impact their learning. This dissertation addresses this crucial need for inclusivity through a Challenge-Based Learning (CBL) approach implemented in a *Bioinspired Design* course. By examining the development of science connection across diverse student populations, this research demonstrates how CBL can cultivate an inclusive learning environment that promotes equitable growth in Science Identity (SciID), Science Self-Efficacy (Eff), and Internalization of Scientific Community Values (Val) for all students, regardless of their demographic identity or disciplinary background.

The future of USE must be Interpersonal. STEM is a deeply interpersonal endeavor. Nearly all landmark publications of the past decade contain multiple authors conducting research as an agile team. Collaboration is no longer a luxury—it is a necessity. USE must be the same. Learners need to work on the STEM problems of today and tomorrow *together*, developing skills that are transferable across various domains in our STEM-enriched society. The prevailing image of an isolated student completing solitary assignments distorts students' perceptions of STEM. Thus, interpersonal collaboration needs to start as early as possible. Courses can capture this reality with opportunities for collaborative CBL. The *Bioinspired Design* course studied in this dissertation embodies this principle by engaging students in interdisciplinary team projects. These projects mirror real-world scientific practice, requiring students to work together to translate scientific discoveries into novel, societally beneficial designs. The course not only develops students' collaborative skills but also provides them with a more authentic understanding of how science works, benefiting both future STEM professionals and those who will apply STEM thinking in other fields.

The future of USE must be Interdisciplinary. The cutting-edge discoveries of today originate from interdisciplinary publications. Seemingly disparate research fields are becoming more integrated because the world's most pressing problems require interdisciplinary solutions. USE of the future must reflect this reality, preparing students for a STEM-enriched landscape where scientific thinking integrates into various sectors of society. This dissertation explores how a course open to all majors, all years, with no prerequisites, can break down traditional disciplinary barriers. Our *Bioinspired Design* course brings together students from over 40 different majors, fostering collaborative learning between diverse disciplinary perspectives. This interdisciplinary principle is exemplified in Chapter 3 by the *Innovation Skills* self-efficacy construct which measured students' confidence in their abilities related to scientific discovery and translation, interdisciplinary thinking, and interdisciplinary collaboration. Increasing confidence in these skills is a critical part of equipping all students with the ability to apply STEM-enriched thinking to all fields, enhancing their capacity to tackle complex societal issues regardless of their chosen profession.

Dissertation Overview

This dissertation explores the implementation and impact of a CBL approach in USE through three interconnected studies. Chapter 1 examines the overall growth in science connection measures (SciID, Eff, and Val) observed in students participating in a *Bioinspired Design* course, providing insight into how CBL can foster meaningful engagement with science across diverse student populations. Building on these findings, Chapter 2 investigates the equity of these outcomes across various demographic groups, demonstrating the potential of CBL to promote inclusive USE. Finally, Chapter 3 advances our understanding of student self-efficacy in the unique CBL environment by developing and validating a novel construct called Innovation Skills self-efficacy. This progression from established measures to new, context-specific assessments reflects the evolving nature of USE and the need for more comprehensive methods to measure student confidence in carrying out 21st century skills.

Chapter 1

A Challenge-Based Learning Approach to Foster Science Connection in a *Bioinspired Design* course

Abstract

Fostering science connection among undergraduate students is crucial for developing a STEM-enriched citizenry capable of addressing complex global challenges. This chapter examines the impact of a Challenge-Based Learning (CBL) course, *Bioinspired Design*, on promoting science connection among 180 undergraduate students from over 40 majors in a single semester across five distinct years. The course engaged students in collaborative identification of biological principles to develop solutions to societal challenges. Using the Tripartite Integration Model of Social Influence (TIMSI) framework, we conceptualized science connection development based on changes in Science Identity (SciID), Science Self-Efficacy (Eff), and Internalization of Scientific Community Values (Val). We utilized a pre/post survey design, employing repeated measures ANOVAs and paired *t*-tests to evaluate individual and collective changes in the science connection constructs. We found significant increases in science connection over five years of course iterations, with notable increases in SciID and Eff specifically, while Val remained stable. Significant connection increases were observed across all semester iterations of course, despite varying instructional modalities due to the COVID-19 pandemic. Item level analyses revealed specific impacts of course activities on TIMSI measures, particularly in developing students' Eff, related to their confidence to create novel technological solutions. Findings demonstrated that CBL can effectively promote science connection in interdisciplinary classroom settings within a single semester, even under challenging social disruptions. Results showcase the potential of CBL as a scalable and adaptable model to support reform in undergraduate STEM education. By creating pathways for students to meaningfully connect with science, CBL experiences contribute to building a more STEM-enriched society crucial for navigating unprecedented rates of change in the world today. The subsequent chapter follows up on these growth outcomes in science connection by investigating potential differential impacts across diverse demographic groups to further understand the inclusive nature of this CBL approach.

Background

Reimagining Undergraduate STEM Education through Challenge-Based Learning

Student learning trajectories through STEM education need to be reimagined to meet the evolving demands of the future (Batchelor et al., 2021; National Science Board, 2015). For example, the "braided river" model proposed by Batchelor et al. (2021) challenges the notion that STEM education is confined to a narrow focus of preparing students for careers in STEM fields alone. Instead, this model captures the interconnected and dynamic nature of STEM education and careers, envisioning the educational landscape as a braided river with multiple interweaving channels representing the diverse paths students can take, both within and beyond traditional STEM boundaries. This model better aligns with future workforce requirements expanding the skills needed for STEM success beyond single-domain expertise (National Research Council, 2014). Importantly, embracing a perspective that includes students from more diverse paths can also lead to the competitive advantages that diversity brings to STEM, empowering a STEM-enriched workforce that can propel future innovation (AlShebli et al., 2018; Hernandez et al., 2013; Hofstra et al., 2020; National Academy of Engineering, 2015).

Interventions at the undergraduate level present a critical opportunity to enhance diversity in STEM fields, as this stage offers a wide variety of entry points and pathways into STEM careers (Miller & Wai, 2015). Implementing effective change strategies during undergraduate education can lead to a more inclusive and interdisciplinary approach to STEM that recognizes the diverse routes students can take and values transferable STEM skills across various contexts. This approach is particularly important given the need to develop a STEM-enriched workforce capable of applying scientific understanding in *both* STEM and Non-STEM fields (National Science Board, 2015). To achieve this goal, educational models that promote interdisciplinary connections among students are essential. One promising framework for designing impactful interdisciplinary learning experiences in undergraduate education is Challenge-Based Learning (CBL).

CBL is as an interdisciplinary, collaborative learning approach where students work to develop solutions to real-world, open-ended challenges by prioritizing global issues and involving the use of technology and input from external stakeholders (Gallagher & Savage, 2020; Johnson et al., 2009; Malmqvist et al., 2015; Membrillo-Hernández et al., 2019). CBL differs from other approaches like problem-based learning (PBL) in its open-endedness and reduced emphasis on prescribed rules of engagement. In PBL, students are given a specific problem with clear parameters and a structured process for developing a solution (Savery, 2006). In contrast, CBL highlights broader, less defined challenges that require students to identify and frame the problem themselves before developing solutions (Malmqvist et al., 2015). CBL is inherently multidisciplinary, drawing on knowledge and skills from multiple fields to develop solutions to complex problems (Malmqvist et al., 2015). Students collaborate in teams, leveraging technology to enable the learning process and develop solutions (Johnson et al., 2009). The learning process in CBL centers itself on problem formulation and solution design, not just acquiring content knowledge (Membrillo-Hernández et al., 2019). By engaging students in authentic, real-world challenges, CBL aims to both educate students and have real-world impact (Gallagher & Savage, 2020; Apple, 2008).

Implementing Challenge-Based Learning through Bioinspired Design

At the University of California, Berkeley, we developed a large-enrollment (180 students) semester course called *Bioinspired Design* (Full et al., 2021) exemplifying key features of CBL (Table 1). The course was situated within our campus makerspace at the Jacobs Institute for Design Innovation, capitalizing on principles of the Maker Movement to provide hands-on, creative learning experiences. Through this partnership, each student received a maker pass, comprehensive safety training, and access to a wide array of makerspace equipment, enabling them to bring their designs to life. Throughout the course, students engaged in the *bioinspired design process* by identifying biological principles from original, published, scientific discoveries, analyzing their potential for translation, and designing sustainable solutions to societal problems. For example, in one design project, students proposed designs for a safer infant car seat activated by the unique frictional adhesive mechanism of gecko toes. In another design project based on insect exoskeletons, students constructed an origami-based legged robot and developed modified prototypes for societally relevant contexts such as agricultural soil monitoring, humanitarian demining, and search-and-rescue operations. In each of these guided design projects, students demonstrated the identification and translation of biological principles into designs that solved open-ended societal challenges.

These projects, and the course as a whole, represented the CBL feature of interdisciplinary thinking. Our course was open to all majors, all years, with no prerequisites, leading to enrollment of students from over 40 different majors. This disciplinarily diverse mix of students, from both STEM and Non-STEM backgrounds, represented the interdisciplinary thinking needed to solve real-world challenges. The course also included the CBL feature of collaborative learning, with students working in selected teams that were balanced across major, class year, and prior design experience. Throughout the course, students also engaged in various team-building exercises and received training on effective teaming strategies to ensure inclusive collaboration.

The course gave students the opportunity to develop solutions to open-ended challenges they themselves defined and deemed most urgent. This included multiple scaffolded assignments where students deconstructed research papers, distinguished between primary and secondary sources, evaluated credibility, used analogical reasoning to propose novel designs, and communicated their ideas to diverse audiences. Through this iterative process of discovery, analysis, and invention, students learned to frame problems, ideate solutions, and refine their designs based on feedback. This culminated in a variety of team-based final design projects derived from start to finish by students, showcasing the interdisciplinary and sociotechnical nature of the challenges they aimed to address. Student teams collaboratively designed sustainable solutions that considered environmental, social, and economic factors. Final design projects have included a flexible cast to reduce muscle atrophy based on the skeleton of seahorses, a voice restoration system for throat cancer patients based on a songbird's syrinx, and a compliant novel suturing device derived from porcupine spines (Full et al., 2021). Many of these final design projects can be seen at https://www.behance.net/berkeleybiodesign.

Table 1
Bioinspired Design Course Connections with Challenge-Based Learning (CBL)

CBL Features	Course Connection	Explanation/Example
Real-world, open-ended challenges	Bioinspired design process; sustainable solutions to societal problems	Guided team design projects; final team design project
Interdisciplinary	40+ majors; interdisciplinary student teams	Criteria selected teams balanced across major, class year, and prior design experience
Collaborative learning	Team-building exercises; training on effective collaboration strategies	Seed dispersal (marshmallow) challenge; collaborative plan
Problem formulation and solution design	Scaffolded assignments; iterative process of discovery, analysis, and invention	Discovery Decomposition and Analogy Check (See Figures 4 and 5)

From Full et al. (2021)

Assessing Social Influence Constructs in a CBL Scientific Community

Considering the distinctive CBL features in our *Bioinspired Design* course (Table 1), we became interested in assessing the correspondingly unique scientific learning community that formed over the semester. In a professional context, a scientific community is a network of scientists and researchers who share a common field of study, interact with one another, exchange ideas and information, and work together to advance scientific knowledge within their domain of expertise (Börner et al., 2010). This community is characterized by shared knowledge and methodologies, communication and collaboration, peer review, consensus-building, and association with specific institutions and organizations (Fortunato et al., 2018; Mukherjee et al., 2017). This professional community also consists of university faculty and researchers who act as agents of *social influence*, attempting to socialize and integrate students into the community through courses, training programs, and mentorship (Estrada et al., 2011). Social influence refers to the ways in which individuals' attitudes, beliefs, and behaviors are shaped by their interactions with others and their social environment to align with the norms and expectations of the community (Estrada et al., 2011; Kelman, 1958).

Our *Bioinspired Design* course represents what we envision as the science learning community of the future. By engaging both STEM and Non-STEM students in authentic socioscientific issues, we hypothesize that the course cultivates a scientific community which equips all students with proscience attitudes, scientific literacy skills, and evidence-based decision-making skills, empowering them to contribute to society as scientifically literate citizens (Ballen et al., 2017; Gormally & Heil, 2022; Sadler, 2009). This community could mirror the collaborative, interdisciplinary nature of professional scientific communities, allowing students to experience authentic aspects of impactful scientific work. As students participated in our unique CBL-based scientific community, we aimed to assess the social influence constructs shaping their connection to the field of science. Students in scientific communities can become more connected to science by developing greater *Science Identity* (SciID), *Science Self-Efficacy*

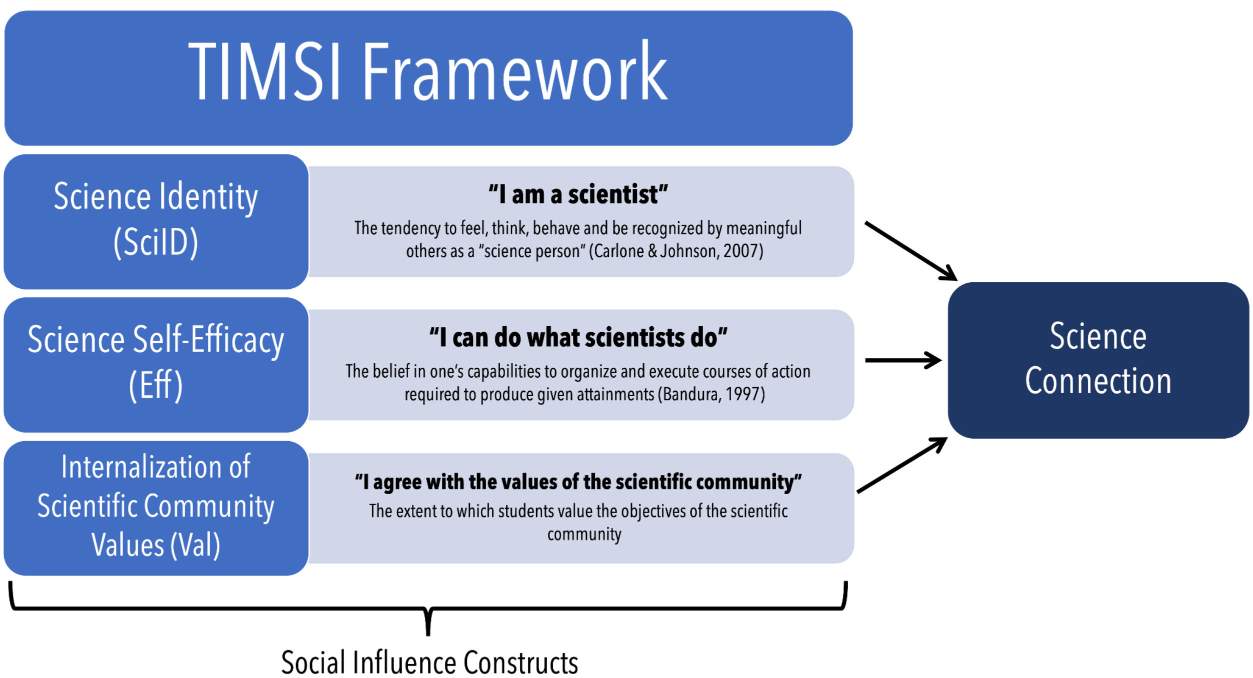
(Eff), and *Internalization of Scientific Community Values* (Val), three social influence constructs that make up the *Tripartite Integration Model of Social Influence* (TIMSI) (Estrada et al., 2011).

Using TIMSI as a Framework to Promote Science Connection for All Students

TIMSI offers a valuable framework for examining social influence constructs that contribute to students' integration into the scientific community (Estrada et al., 2011). In this framework, three constructs individually and collectively contribute to student integration—SciID, Eff, and Val (Figure 1). SciID refers to the extent to which a person sees themselves as a "science person" (Carlone & Johnson, 2007; Chemers et al., 2011; Estrada et al., 2011). Eff is defined as a person's belief in their ability to successfully perform scientific tasks and achieve scientific goals (Bandura, 1997; Estrada et al., 2011). Val involves endorsing and accepting the objectives, goals, and values of the scientific community as one's own (Estrada et al., 2011; Kelman, 1958). Put simply, each of these social influence constructs can be characterized by three belief statements: "I am a scientist" (SciID), "I can do what scientists do" (Eff), and "I agree with the values of the scientific community" (Val).

Figure 1

Conceptual Diagram of the Tripartite Integration Model of Social Influence (TIMSI) Framework



Note. The framework illustrates how Science Identity (SciID), Science Self-Efficacy (Eff), and Internalization of Scientific Community Values (Val) collectively contribute to Science Connection. Each construct is accompanied by a brief definition and a representative statement reflecting its core concept. Arrows indicate how these social influence constructs collectively lead to science connection. Adapted from Estrada et al., (2011).

In the TIMSI framework, these social influence constructs impact student integration into the scientific community. Importantly, in disciplinarily diverse settings like our course, the overarching outcome of integration (i.e., persistence in STEM) may not be equally relevant or desirable for all students. To account for this breadth of interests, we reconceptualized use of the

TIMSI framework (and its social influence constructs) to measure *science connection*. Conceptually, science connection acknowledges the diverse ways in which both STEM and Non-STEM students can develop a connection to science without necessarily pursuing a STEM career (i.e., integrating into a formalized STEM community). These students, just like their STEM counterparts, participate in science contexts within and outside of the classroom. We recognize that nearly all students participate in proximal science contexts such as STEM course communities (e.g., science breadth courses), but these same students could benefit from science connections after graduating, such as participation in science-related legislation, public health decision-making, and consumer decision-making. These contexts represent moments in which high levels of SciID, Eff, and Val can lead to informed participation in the democratic process, pro-science behaviors, and critical evaluation of scientific claims (Estrada et al., 2017; Ballen et al., 2017; Gormally & Heil, 2022). By fostering science connection, all students can contribute to a more scientifically literate society and make informed decisions across diverse aspects of their lives. Thus, this reconceptualization of the TIMSI framework allows us to examine the development of key social influence constructs—SciID, Eff, and Val—among all students, regardless of their disciplinary background. Here and henceforth, "development" refers to growth in these constructs, particularly after an intervention. By demonstrating that different types of students can develop in these constructs, we can provide evidence for the transformative potential of CBL-based interventions for a broader array of students.

Hypotheses for Science Connection Development in a CBL Environment

The unique student composition of our *Bioinspired Design* course offers a critical opportunity to investigate the development of science connection in a unique CBL environment open to all students. Broadly speaking, measures of science connection have been shown to decrease or remain stable in students as they progress through traditional undergraduate STEM courses (Estrada et al., 2019; Seymour & Hunter, 2019). Cole & Beck (2022) found that while Eff and SciID increased over a year-long introductory biology course sequence, Val did not, suggesting that different aspects of science connection may develop at different rates and in response to different factors. Other studies examining changes in science connection constructs during a single semester have yielded mixed results, with increases in Eff in introductory biology (Ainscough et al., 2016) and varied trajectories of SciID in introductory chemistry, including rapid negative shifts for some students (Robinson et al., 2019). Given the trend of decreasing or stable measures of science connection in other studies, maintaining baseline levels of these constructs can be considered a positive outcome, while any significant increases would be particularly noteworthy. We predicted that our *Bioinspired Design* course would cultivate statistically significant gains in social influence indicators of science connection (SciID, Eff, and Val) and tested the following hypothesis:

> **Overall growth hypothesis; H_1.** The *Bioinspired Design* course will lead to significant pre/post increases in science connection, as measured by individual and collective SciID, Eff, and Val development.

Moreover, the semester in which students took the course may have influenced their science connection due to variations in instructional modality instigated by the COVID-19 pandemic. Our data span five Spring semesters (Spring 2019-Spring 2023), each with varying instructional modalities pre-, mid-, and post-pandemic. The unprecedented transition to online learning in Spring 2020 had significant impacts on students' experiences and outcomes, with many facing challenges related to technology access, motivation, and course engagement (Means & Neisler, 2020). These impacts were not uniform and continued to evolve as the pandemic progressed, with institutions adapting their instructional approaches over time (National Academies of Sciences, Engineering, and Medicine [NASEM], 2021). Given the systemic effects of the pandemic on higher education, it is crucial to assess whether semester specific contexts influenced students' development of science connection. Our dataset spanning five different years provides a unique opportunity to examine the course's effectiveness and adaptability in promoting science connection under diverse and evolving circumstances. Thus, we also tested:

> **Semester growth hypothesis; H_2**. The *Bioinspired Design* course will lead to pre/post improvements in science connection across five different semester iterations (Spring 2019-Spring 2023) with varying modalities of instruction.

Methods

Study Design

We employed a pre/post survey design to measure the change in each of the TIMSI constructs in students as a result of participating in the *Bioinspired Design* course. This research design allowed us to conduct within-subjects comparisons that assessed the same set of students before and after the course. The survey data comes from independent iterations of the course over five semesters (Spring 2019-Spring 2023). We analyzed response data from students who completed both the pre- and post-survey, enabling us to conduct the subsequent statistical analyses with matched pre/post pairs (ranging from $N = 494$ to $N = 529$ depending on the statistical test). We also collected student demographic data and present those results in Chapter 2 alongside the demographic-based data analyses assessing equitable course outcomes. Students voluntarily completed both the pre- and post-survey online via Qualtrics for one point of extra credit. The survey included an informed consent notice and assurance of confidentiality. All experimental protocols were IRB approved (Protocol ID: 2017-12-10602).

Statistical Analysis

We used a statistical package for all analyses (IBM SPSS Statistics for Macintosh, Version 28.0, IBM Corp., 2021). Complementary visualizations of the SPSS results were created using R (version 4.4.1; R Core Team, 2024) with the *ggplot2* and *dplyr* packages (Wickham, 2016) to enhance the interpretability of the statistical findings. Descriptive statistics were computed for each survey item (pre and post) and the overall constructs of SciID, Eff, and Val. Normality was examined through histograms, skewness and kurtosis values. The subsequent parametric tests (e.g., ANOVA) are considered robust to violations of normality, particularly with large sample sizes such as ours (Blanca et al., 2017; Blanca et al., 2023). Bonferroni corrections were applied to all applicable statistical analyses to control for Type I error when making multiple comparisons.

We conducted repeated measures ANOVA (RM ANOVA) to compare pre- and post-survey scores. Given a repeated measures design with only two time points (pre and post), both assumptions of sphericity and homogeneity of variance were inherently met. We first conducted a set of RM ANOVAs labeled "Overall RM ANOVAs." We used a within-subjects design with *time* (pre/post) as the within-subjects factor. These RM ANOVAs tested H_1 (overall growth hypothesis) by measuring overall pre to post changes in survey scores. Multivariate tests assessed overall pre-to-post changes across the combined constructs and univariate tests assessed pre-to-post changes for each individual construct. The next set, labeled "Semester RM ANOVAs," tested the impact of the *Bioinspired Design* course on pre to post changes in science connections within the context of each unique semester. This allowed us to test H_2 (semester growth hypothesis), or the effectiveness of the course in promoting science connection under the specific circumstances of each semester (e.g., varying instructional modalities necessitated by the COVID-19 pandemic).

We then supplemented the RM ANOVA analysis with two sets of paired *t*-tests analyses at the item and overall construct levels to further assess pre/post differences. The item level analyses investigated pre-to-post changes at a more granular level and were used to explore underlying factors potentially driving the overall results. Based on Maher et al.'s (2013) recommendation for quantitative discipline-based education research, we also obtained effect sizes (partial eta squared [η_p^2] values and Cohen's *d* values) to evaluate the practical significance of any pre/post changes. In other words, we sought to determine how meaningful the magnitude of change was in the specific context of our course-based intervention, beyond just statistical significance. For partial eta squared (η_p^2), effect size values of .01 (small), .06 (medium), and .14 (large) are referenced as general guidelines, while for Cohen's *d*, values of 0.2 (small), 0.5 (medium), and 0.8 (large) are commonly cited benchmarks (Cohen, 1988). Being mindful of Cohen's (1988) caution against using these benchmarks as rigid defaults and in accordance with recent discussions in educational research (Kraft, 2020), we interpreted our effect sizes within the specific context of our semester-based educational intervention (e.g., adapted Cohen's *d* value interpretations to < 0.1 = small, 0.1 to < 0.3 = medium, and ≥ 0.3 = large). We further explain the details motivating our effect size interpretations in Supplement 1.

Survey Design

We administered an adapted version of a previously validated pre/post survey measuring student integration into the scientific community based on the TIMSI framework (Estrada et al., 2011). Our survey measured the aforementioned constructs—SciID, Eff, Val—through the same three sets of Likert scale items as Estrada et al. (2011) with the exception of two new Eff items (see Table 2). This included a 5-item scale evaluating SciID, with statements such as "*I have come to think of myself as a scientist*" rated from 1 (strongly disagree) to 7 (strongly agree). SciID items connected to course activities such as the final design project, where students proposed and communicated their own bioinspired designs, potentially promoting a view of themselves as capable of contributing to scientific discourse and innovation (Full et al., 2021). Eff was measured through an 8-item scale assessing confidence to complete core science tasks on a scale from 1 (not at all confident) to 5 (absolutely confident). Items 12 and 13 were specifically added for our unique course context. Item 12 assessed confidence in designing experiments to test hypotheses, which aligned with the course's scaffolded assignments involving primary literature decomposition and analogical reasoning to propose novel designs (Full et al., 2021). Item 13 evaluated confidence in developing novel technologies, which connected to several course activities like the final design project where student teams created a novel bioinspired invention. Lastly, Val was evaluated with a 6-item scale including endorsements of priorities like building scientific knowledge and identifying truths using the scientific method on a scale from 1 (not like me at all) to 6 (very much like me). For example, the final design project involved communicating discoveries in a public design showcase, reflecting a scientific community value of making knowledge accessible and usable by society (Full et al., 2021).

By outlining these connections between the SciID, Eff, and Val items and key course activities, we establish a validity argument for employing the TIMSI framework to measure the development of science connection in the unique CBL context of our course. The alignment between the TIMSI survey items and the activities in the *Bioinspired Design* course supports the use of this adapted instrument as an applicable tool for evaluating outcomes related to science connection.

Table 2
TIMSI Survey Items

Item	TIMSI Construct	Scale
1. I have a strong sense of belonging to the community of scientists.	SciID	Strongly Disagree (1) to Strongly Agree (7)
2. I derive great personal satisfaction from working on a science team that is doing important work.	SciID	Strongly Disagree (1) to Strongly Agree (7)
3. I have come to think of myself as a scientist.	SciID	Strongly Disagree (1) to Strongly Agree (7)
4. I feel like I belong in the field of science.	SciID	Strongly Disagree (1) to Strongly Agree (7)
5. The daily work of a scientist is appealing to me.	SciID	Strongly Disagree (1) to Strongly Agree (7)
6. Use technical skills (use of tools, instruments, and/or techniques of your field of study).	Eff	Not at all confident (1) to Absolutely confident (5)
7. Generate a research question to answer.	Eff	Not at all confident (1) to Absolutely confident (5)
8. Determine what data/observations to collect and how to collect them.	Eff	Not at all confident (1) to Absolutely confident (5)
9. Create explanations for the results of the study.	Eff	Not at all confident (1) to Absolutely confident (5)
10. Use academic literature and/or reports to guide your research.	Eff	Not at all confident (1) to Absolutely confident (5)
11. Develop theories (integrate and coordinate results from multiple studies and/or theories).	Eff	Not at all confident (1) to Absolutely confident (5)
12. Designs experiments to test hypotheses.	Eff	Not at all confident (1) to Absolutely confident (5)

13. Develop novel technologies.	Eff	Not at all confident (1) to Absolutely confident (5)
14. A person who thinks discussing new theories and ideas between scientists is important.	Val	Not like me at all (1) to Very much like me (6)
15. A person who believes writing up research results to be published in a leading scientific journal is a good use of time.	Val	Not like me at all (1) to Very much like me (6)
16. A person who thinks it is valuable to conduct research that builds the world's scientific knowledge.	Val	Not like me at all (1) to Very much like me (6)
17. A person who thinks that scientific research can solve many of today's world challenges.	Val	Not like me at all (1) to Very much like me (6)
18. A person who feels discovering something new in the sciences is thrilling.	Val	Not like me at all (1) to Very much like me (6)
19. A person who thinks it is important work to identify truths using the scientific method.	Val	Not like me at all (1) to Very much like me (6)

Reliability

Reliability analyses were performed to measure the internal consistency (Cronbach's alpha [α]) of the pre- and post-survey. The instrument demonstrated sufficient internal consistency across each set of pre/post items (SciID, 5 items; Eff, 8 items; Val, 6 items) and the overall pre/post survey (19 items). Alpha values ranged from good to excellent internal consistency ($\alpha > 0.8$) as shown in Table 3. In response to Maric et al.'s (2023) call for expanding reliability evidence in science education research, Supplement 2 contains further discussion of these results, including contextualized interpretations and an additional analysis of reliability based on McDonald's omega (ω). The results in the supplementary analysis matched the results presented here, supporting the overall reliability of the instrument.

Table 3
Reliability of Survey Constructs (Cronbach's Alpha [α])

Construct	N of Items	N (pre)	N (post)	α (pre)	α (post)
SciID	5	966	554	0.925	0.927
Eff	8	958	546	0.907	0.910
Val	6	956	540	0.894	0.906
Overall	19	944	529	0.916	0.925

Results

Significant Pre/Post Growth in Science Connection Based on Overall RM ANOVAs

The descriptive statistics (means and standard deviations) for each construct in our sample of $N = 494$ matched pre/post surveys are shown in Table 4. For the Overall RM ANOVAs, pre-post differences were analyzed through both a multivariate test combining all constructs alongside univariate tests analyzing each individual construct (SciID, Eff, and Val).

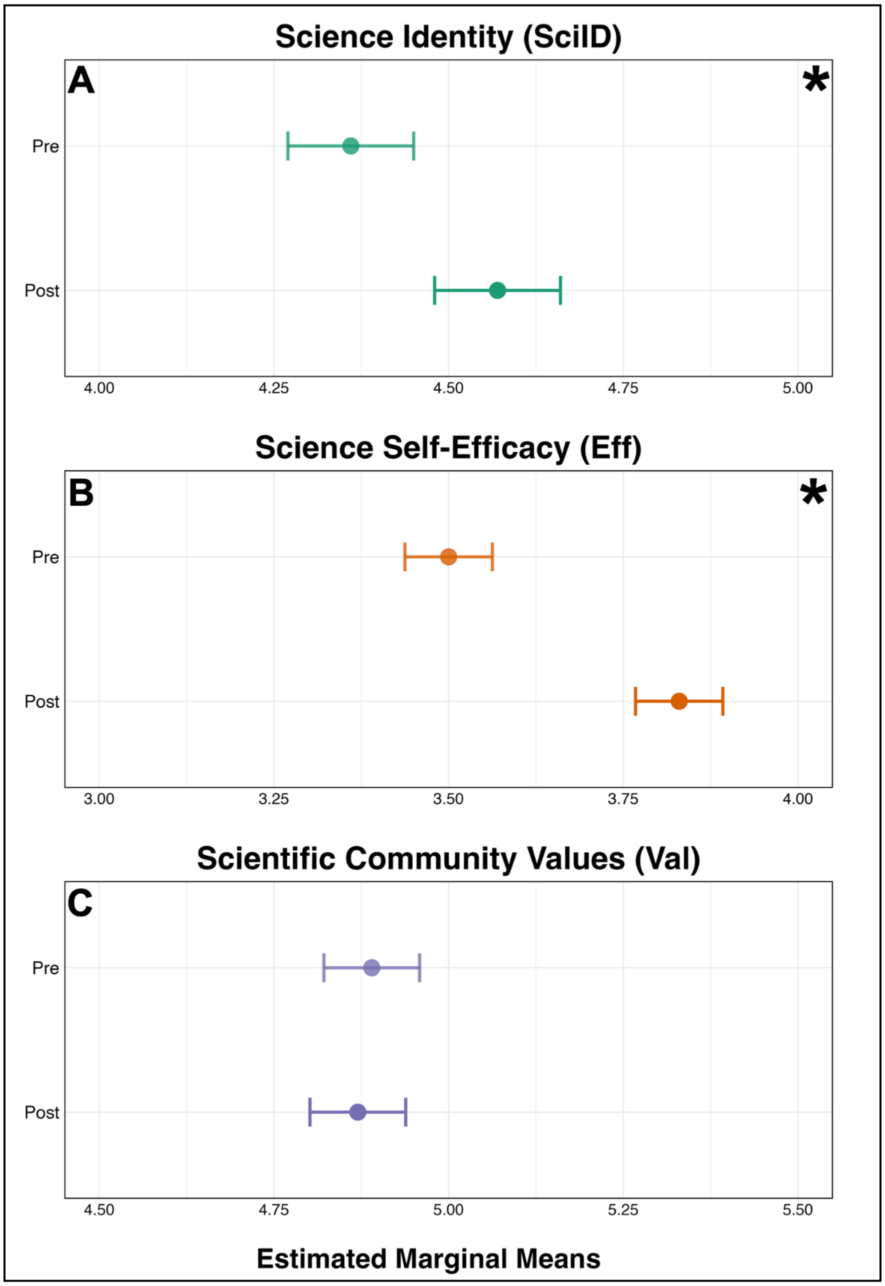
Table 4
Descriptive Statistics of RM ANOVA Constructs

Construct	Mean	SD	*N*
SciID_mean_pre	4.363	1.500	494
SciID_mean_post	4.569	1.477	494
Eff_mean_pre	3.503	.735	494
Eff_mean_post	3.830	.705	494
Val_mean_pre	4.886	.834	494
Val_mean_post	4.872	.926	494

Based on the multivariate test, the impact of completing the *Bioinspired Design* course on the change in pre-to-post survey scores was statistically significant at $p < .05$ level: $F(3, 491) = 43.766$, $p < .001$, $\eta_p^2 = .211$. This represents a large effect size ($\eta_p^2 > 0.14$; Cohen, 1988) when comparing the change in pre- versus post-survey scores in the combined constructs. Based on the univariate tests (Figure 2; Table 5), the impact of completing the *Bioinspired Design* course on the change in pre-to-post survey scores was statistically significant at the $p < .05$ level for SciID with small-medium effect size: $F(1, 493) = 20.086$, $p < .001$, $\eta_p^2 = .039$ (Figure 2A). Post-hoc pairwise comparisons with a Bonferroni adjustment indicated a statistically significant mean score gain (+0.206; 95% CI [.116, .296], $p < .001$) in the SciID post-survey (M = 4.57, SE = .046) versus the pre-survey (M = 4.36, SE = .046). There was also a statistically significant difference at the $p < .05$ level for the change in Eff with large effect size: $F(1, 493) = 105.470$, $p < .001$, $\eta_p^2 = .176$ (Figure 2B). Post-hoc pairwise comparisons with a Bonferroni adjustment indicated a statistically significant mean score gain (+0.327; 95% CI [.264, .389], $p < .001$) in the Eff post-survey (M = 3.83, SE = .032) versus the pre-survey (M = 3.50, SE = .032). Lastly, there was no statistically significant difference from pre to post in Val: $F(1, 493) = .173$, $p = .677$, $\eta_p^2 = .000$ (Figure 2C). Post-hoc pairwise comparisons with a Bonferroni adjustment indicated that there was non-significant mean difference of -.015 (95% CI [-.083, .054], $p = .667$) in the Val post-survey (M = 4.87, SE = .035) versus the pre-survey (M = 4.89, SE = .035).

Figure 2

Dot Plots of Overall RM ANOVA Univariate Tests



Note. Dot plots show change in estimated marginal means from pre-to-post in A) SciID [1 (strongly disagree) to 7 (strongly agree) Likert scale], B) Eff [1 (not at all confident) to 5 (absolutely confident) Likert scale], and C) Val [1 (not like me at all) to 6 (very much like me) Likert scale] constructs. Error bars represent 95% confidence intervals. Asterisks indicate statistically significant differences between pre and post with Bonferroni correction applied ($p < .05/3$).

Table 5
Overall RM ANOVA - Tests of Within-Subjects Effects - Univariate

Source	Construct	Type III Sum of Squares	df	Mean Square	F	Sig.*	Partial Eta Squared	Effect Size
PrePost	**SciID**	**10.489**	**1**	**10.489**	**20.086**	**<.001**	**.039**	**Small-med**
	Eff	**26.399**	**1**	**26.399**	**105.470**	**<.001**	**.176**	**Large**
	Val	.052	1	.052	.173	.677	.000	N/A
Error (PrePost)	SciID	257.451	493	.522				
	Eff	123.398	493	.250				
	Val	147.906	493	.300				

*Bold indicates significant difference between pre and post; Bonferroni correction applied ($p < .05/3$)

Development of Science Connection in All Semesters Based on RM ANOVAs

A second set of RM ANOVAs were conducted for each individual semester (Spring 2019, $N = 127$; Spring 2020, $N = 127$; Spring 2021, $N = 97$; Spring 2022, $N = 72$; Spring 2023, $N = 71$). In the multivariate RM ANOVA combining all constructs, we observed statistically significant pre to post growth in each individual semester (Table 6). The effect sizes varied across semesters, with Spring 2020 showing the highest effect size ($\eta_p^2 = .084$; medium-large effect), followed by Spring 2023 ($\eta_p^2 = .072$; medium-large effect), Spring 2022 ($\eta_p^2 = .047$; small-medium effect), Spring 2021 ($\eta_p^2 = .036$; small-medium effect), and Spring 2019 (partial $\eta^2 = .043$; small-medium effect).

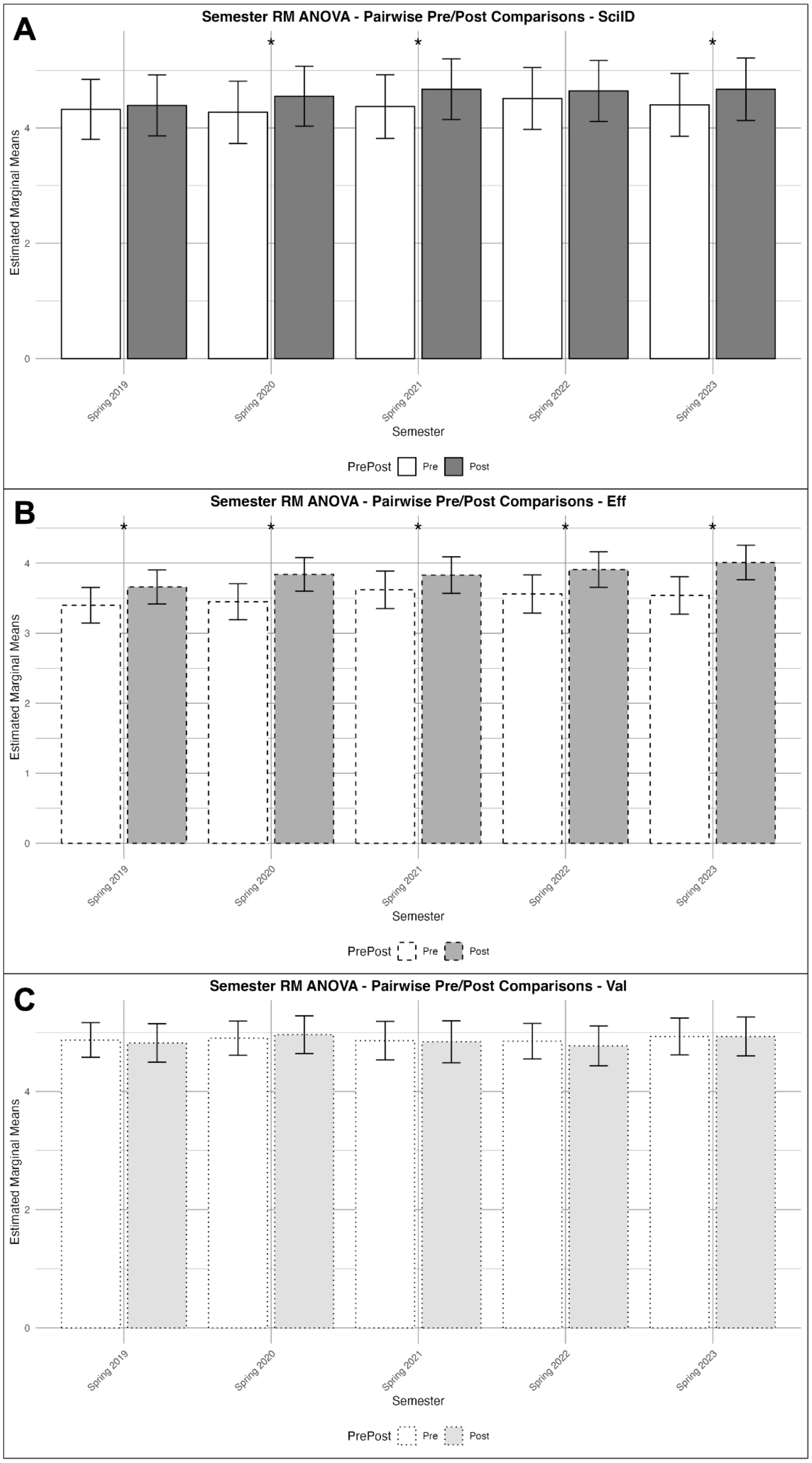
Table 6
Semester RM ANOVA - Tests of Within-Subjects Effects - Multivariate

Semester		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Effect Size
Spring 2019	Pillai's Trace	.043	7.251	3.000	487.000	<.001	.043	Small-med
	Wilks' Lambda	..957	7.251	3.000	487.000	<.001	.043	Small-med
Spring 2020	Pillai's Trace	.084	14.809	3.000	487.000	<.001	.084	Med-large
	Wilks' Lambda	.916	14.809	3.000	487.000	<.001	.084	Med-large
Spring 2021	Pillai's Trace	.036	5.987	3.000	487.000	<.001	.036	Small-med
	Wilks' Lambda	.964	5.987	3.000	487.000	<.001	.036	Small-med
Spring 2022	Pillai's Trace	.047	7.988	3.000	487.000	<.001	.047	Med-large
	Wilks' Lambda	.953	7.988	3.000	487.000	<.001	.047	Med-large
Spring 2023	Pillai's Trace	.072	12.520	3.000	487.000	<.001	.072	Med-large
	Wilks' Lambda	.928	12.520	3.000	487.000	<.001	.072	Med-large

In the univariate RM ANOVA (Figure 3; Table 7), pairwise pre/post comparisons for each semester revealed that for SciID (Figure 3A), statistically significant increases were observed in all semesters except Spring 2019 and Spring 2022, though these semesters still showed non-significant positive trends. Eff (Figure 3B) demonstrated statistically significant increases across all semesters, indicating consistent growth in students' scientific self-efficacy regardless of the instructional modality. Val (Figure 3C) showed no statistically significant increases or decreases in any semester, mirroring the stability observed in this construct from the Overall RM ANOVA analysis.

Figure 3

Bar Graphs of Semester RM ANOVA Pairwise Pre/Post Comparisons



Note. Bar graphs show change in estimated marginal means from pre-to-post in A) SciID [1 (strongly disagree) to 7 (strongly agree) Likert scale], B) Eff [1 (not at all confident) to 5 (absolutely confident) Likert scale], and C) Val [1 (not like me at all) to 6 (very much like me) Likert scale] constructs across five semesters (Spring 2019 to Spring 2023). Error bars represent 95% confidence intervals. Asterisks indicate statistically significant differences between pre and post with Bonferroni correction applied ($p < .05/5$).

Table 7
Semester RM ANOVA - Pairwise Pre/Post Comparisons - Univariate

Semester	Construct	Pre Mean (SD)	Post Mean (SD)	Mean Difference	Sig*
Spring 2019	SciID	4.32 (1.45)	4.39 (1.48)	0.07	.455
	Eff	**3.40 (0.71)**	**3.66 (0.68)**	**0.26**	**<.001**
	Val	4.87 (0.82)	4.82 (0.91)	-0.05	.435
Spring 2020	**SciID**	**4.27 (1.51)**	**4.55 (1.45)**	**0.28**	**.002**
	Eff	**3.45 (0.72)**	**3.84 (0.67)**	**0.39**	**<.001**
	Val	4.90 (0.81)	4.96 (0.89)	0.06	.371
Spring 2021	**SciID**	**4.37 (1.54)**	**4.67 (1.47)**	**0.30**	**.004**
	Eff	**3.62 (0.75)**	**3.83 (0.73)**	**0.21**	**.003**
	Val	4.86 (0.91)	4.84 (0.99)	-0.02	.777
Spring 2022	SciID	4.51 (1.50)	4.64 (1.48)	0.13	.300
	Eff	**3.56 (0.76)**	**3.91 (0.71)**	**0.35**	**<.001**
	Val	4.85 (0.84)	4.77 (0.94)	-0.08	.363
Spring 2023	**SciID**	**4.40 (1.52)**	**4.67 (1.51)**	**0.27**	**.026**
	Eff	**3.54 (0.75)**	**4.01 (0.69)**	**0.47**	**<.001**
	Val	4.93 (0.87)	4.93 (0.92)	0.00	1.000

*Bold indicates significant difference between pre and post; Bonferroni correction applied

Science Connection Development Based on Paired t-Tests, Significant Increases in All Eff Items

Overall construct paired *t*-tests (Table 8) indicated statistically significant gains in mean score from pre to post in the SciID (+0.193, $t(524) = 4.283$, $p < .001$) and Eff (+0.329, $t(516) = 10.613$, $p < .001$) constructs. Based on an effect size interpretation of Cohen's *d* contextualized to our specific educational context (Kraft, 2020; see Supplement 1), overall effect sizes were medium and large for the pre/post growth in SciID and Eff, respectively (< 0.1 = small, 0.1 to < 0.3 = medium, and ≥ 0.3 = large). Val had a -0.001 decrease in mean score that was not statistically significant ($t(506) = -.037$, $p = 0.970$).

Table 8
Overall Construct Paired t-tests

Construct (Post-Pre)	Mean Difference	SD	Std. Error Mean	95% CI Lower	95% CI Upper	t	df	Sig.*	Cohen's *d*
SciID	.193	1.037	.045	.104	.282	4.283	524	**<.001**	.187
Eff	.329	.705	.031	.268	.390	10.613	516	**<.001**	.467
Val	-.001	.790	.035	-.070	.067	-.037	506	.970	

*Bold indicates significant difference between pre and post; Bonferroni correction applied ($p < .05/3$)

We also compared item means from the pre-survey to the post-survey (Table 9). Sample sizes ranged from $N = 512$ to $N = 529$. Increases from pre to post were observed in 5/5 SciID items, 8/8 Eff items, and 2/6 Val items. The remaining 4 Val items showed no difference in 1 item (Item 14) and decreases in Items 16-18. To assess the statistical significance of these differences, item level paired *t*-tests were conducted. These tests showed statistically significant pre-to-post increases in 2/5 SciID items, 8/8 Eff items, and 1/6 Val items after Bonferroni correction ($p < .05/19$). Mean differences in Item 4 and Item 18 were initially significant, but not after Bonferroni correction. Effect sizes of these differences ranged from $d = 0.136$ (medium) to $d = 0.620$ (large) (Table 10).

Table 9
Item Level Paired t-Tests

Item (Post-Pre)	Mean Difference	SD	Std. Error Mean	95% CI Lower	95% CI Upper	t	df	Sig.* (2-tailed)
Item1_SciID	**.308**	**1.390**	**.060**	**.189**	**.427**	**5.097**	**528**	**<.001**
Item2_SciID	.100	1.500	.065	-.028	.229	1.538	527	.125
Item3_SciID	**.302**	**1.372**	**.060**	**.185**	**.420**	**5.051**	**525**	**<.001**
Item4_SciID	.159	1.366	.059	.042	.275	2.674	528	.008
Item5_SciID	.109	1.421	.062	-.012	.231	1.773	529	.077
Item6_Eff	**.235**	**.956**	**.042**	**.153**	**.317**	**5.650**	**526**	**<.001**
Item7_Eff	**.294**	**1.046**	**.046**	**.204**	**.384**	**6.429**	**523**	**<.001**
Item8_Eff	**.276**	**.983**	**.043**	**.191**	**.360**	**6.433**	**525**	**<.001**
Item9_Eff	**.238**	**.942**	**.041**	**.157**	**.319**	**5.790**	**524**	**<.001**
Item10_Eff	**.245**	**1.010**	**.044**	**.159**	**.332**	**5.570**	**525**	**<.001**
Item11_Eff	**.316**	**1.033**	**.045**	**.228**	**.405**	**7.011**	**524**	**<.001**
Item12_Eff	**.312**	**1.008**	**.044**	**.226**	**.399**	**7.098**	**524**	**<.001**
Item13_Eff	**.699**	**1.127**	**.049**	**.603**	**.795**	**14.252**	**527**	**<.001**
Item14_Val	.000	1.174	.052	-.101	.101	.000	518	1.000
Item15_Val	**.174**	**1.286**	**.057**	**.063**	**.286**	**3.080**	**515**	**.002**
Item16_Val	-.050	1.050	.046	-.141	.040	-1.091	514	.276
Item17_Val	-.060	1.024	.045	-.149	.028	-1.336	512	.182
Item18_Val	-.113	1.078	.048	-.206	-.019	-2.371	514	.018
Item19_Val	.004	1.139	.050	-.095	.102	.077	515	.938

*Bold indicates significant difference between post item mean and pre item mean; Bonferroni correction applied ($p < .05/19$)

Note: Item4_SciID (0.008 > .05/19) and Item18_Val (0.018 > .05/19) not statistically significant after Bonferroni correction

Table 10
Effect Sizes of Significant Item Level t-Tests

Item (Post-Pre)	Mean Difference	Sig.* (2-tailed)	Cohen's *d*	95% CI Lower	95% CI Upper
Item1_SciID	.308	<.001	.222	.135	.308
Item3_SciID	.302	<.001	.220	.134	.307
Item6_Eff	.235	<.001	.246	.159	.333
Item7_Eff	.294	<.001	.281	.193	.368
Item8_Eff	.276	<.001	.281	.193	.368
Item9_Eff	.238	<.001	.253	.166	.339
Item10_Eff	.245	<.001	.243	.156	.329
Item11_Eff	.316	<.001	.306	.218	.393
Item12_Eff	.312	<.001	.310	.222	.397
Item13_Eff	.699	<.001	.620	.527	.713
Item15_Val	.174	.002	.136	.049	.222

*Bonferroni correction applied ($p < .05/19$)

Discussion

In this study, we presented a CBL course context that expanded the use of TIMSI constructs (i.e. social influence constructs; SciID, Eff, and Val) to an undergraduate breadth course open to all students. Previous research has examined these constructs in longitudinal science training programs (Estrada et al., 2021), investigations centered on social influence agents like mentor networks and research experiences (Hernandez et al., 2020), and various formalized research experiences such as course-based undergraduate research experiences (CUREs) (Newell & Ulrich, 2022; Ramírez-Lugo et al., 2021; Shuster et al., 2019). Our *Bioinspired Design* course represented a new type of learning environment that diverged from the longitudinal nature of previous TIMSI studies (e.g., Estrada et al., 2018; Robnett et al., 2015), the smaller class sizes of CUREs (e.g., Miller et al., 2023; Rodenbusch et al., 2017), while overcoming some of the known challenges to CURE implementation (e.g., scalability, instructor training, and resource availability) (Bakshi et al., 2016; Linn et al., 2015; Spell et al., 2014). In this course, we built upon the strengths of CUREs in facilitating discovery-based inquiry, while making key adaptations for a breadth course format that was scalable and accessible to a broad student population. This breadth course context accommodated large enrollments and did not require extensive experimental methods training yet remained effective at equitably promoting science connection in students based on the adapted TIMSI framework. We discuss the possible mechanisms behind the key results below.

Overall Growth Hypothesis (H_1): Significant Gains in SciID, Eff, and stability in Val Demonstrate Course Impact

Our analysis of pre/post survey data revealed significant overall gains in science connection, supporting H_1 (overall growth hypothesis). The large effect size observed in the multivariate (combined constructs) RM ANOVA ($\eta_p^2 = .211$) suggests that our CBL course had a meaningful impact on students' overall science connection after they completed the course. For example, as compared to before the course, greater science connection may have developed through students completing various course activities such as engagement with scientific literature, team collaborations, and multiple design projects. On the univariate (individual construct) level, we found statistically significant increases with meaningful effect sizes in both SciID and Eff, but not in Val (Figure 2; Table 5). We saw similar results in the overall construct paired *t*-tests with the highest increases in Eff, followed by growth in SciID, and no statistically significant changes in Val (Table 8). The large effect sizes for Eff in both analyses ($\eta_p^2 = .176$; $d = 0.467$) suggests that our course was particularly effective in enhancing this social influence construct. The small/medium effect sizes for SciID ($\eta_p^2 = .039$; $d = 0.187$) indicates a notable but smaller impact compared to Eff.

These results may be attributed to the inherent nature of SciID as a construct that takes longer to develop than Eff. As students gain confidence in their ability to perform scientific tasks (Eff), they may then gradually incorporate these experiences into their self-concept as a scientist (SciID). In the research context of a single semester, SciID has been shown to remain mostly stable or even decline rapidly for some students in introductory science courses (Robinson et al., 2019). Our results, gains in both Eff and SciID, are consistent with studies suggesting that increases in Eff often precede and contribute to the development of a stronger SciID (Cole & Beck, 2022; Robnett et al., 2015). This also corresponds with qualitative research suggesting that increased confidence in science abilities contributes to students' sense of becoming a scientist (Seymour et al., 2004; Aschbacher et al., 2010). In contrast to Eff and SciID, the Val construct showed no significant change from pre-to-post, suggesting that students' internalization of scientific values remained relatively stable over the course of the semester. Our discussion of the item level results provides more insight on Val outcomes.

Semester Growth Hypothesis (H_2): A CBL Approach Demonstrates Resilient Development of Science Connection

The semester-by-semester analysis of our CBL course revealed significant growth in science connection across all five semesters, regardless of instructional modality (Figure 3; Tables 6 and 7). Given the varied circumstances under which the course was delivered in each semester, a particularly notable result came from the Spring 2020 multivariate RM ANOVA (Table 6). This semester involved a sudden shift from in-person to remote learning due to the COVID-19 pandemic, yet we observed the highest growth based on effect size ($\eta_p^2 = .084$; medium-large effect) in this semester compared to other semesters. Despite the abrupt transition to remote learning at the onset of a global pandemic, students demonstrated significant gains in science connection. This finding is particularly significant when considered alongside other studies that found decreased satisfaction and engagement among STEM students during the transition to online learning (Means & Neisler, 2020).

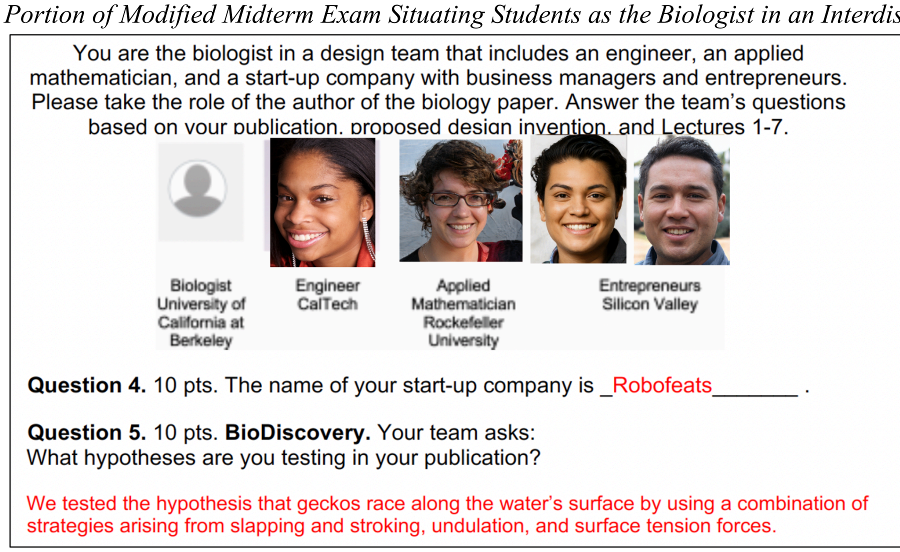
We suspect that important structural adaptations to the course potentially enhanced certain aspects of student learning and subsequently, the social influence constructs underlying their science connection. For example, implementing flexible deadlines, modified assessment formats, and the option for both asynchronous and synchronous learning may have provided students with greater autonomy and reduced stress, potentially facilitating deeper engagement with the course material. These adaptations matched with recommendations from Camfield et al. (2020) for addressing pandemic-induced inequalities in higher education by offering more flexible and inclusive assessment methods. Such modifications may have helped mitigate some of the negative impacts on student learning reported in other studies during the pandemic (NASEM, 2021). This approach seems to have been particularly effective in sustaining and enhancing students' science connection, contrasting with the challenges in student engagement and learning reported by Morán-Soto et al. (2022) during the transition to online learning. Based on overwhelmingly positive student feedback, we retained these adaptations in subsequent iterations of the course, leading to sustained positive impacts as exemplified by the growth in science connection across all semesters of analysis even after Spring 2020.

Focusing on the specific social influence construct level (univariate Semester RM ANOVAs), SciID and Eff showed increases in every semester of the analysis (Figure 3; Table 7). Increases in SciID were statistically significant in three out of five semesters (Figure 3A), while Eff significantly increased in every semester (Figure 3B). This consistent growth in Eff is especially notable, as it stands in contrast to findings from Forakis & March (2023), who observed decreased Eff in chemistry students during the pandemic. Similarly, Means & Neisler (2020) reported that STEM students faced significant challenges in maintaining confidence in their abilities during remote learning. Our results suggest that the CBL approach may be particularly effective in advancing students' confidence in their scientific abilities (Eff), even during unprecedented educational contexts.

We theorize that the revised midterm exam format (Figure 4) may have contributed to the specific growth we saw in these social influence constructs. The midterm exam situated learning in a real-world context, tasking students to take on the role of an expert biologist in an interdisciplinary design team, explain relevant concepts from lectures to their hypothetical team, apply class methods to extract fundamental principles from authentic discoveries, and ultimately propose their own novel bioinspired design based on extracted principles. Students were given a 14-day window to complete the exam, reducing stress and promoting more equitable evaluation practices. This creative application of course concepts and extended timeframe for submission may have contributed to increased Eff, as students had more opportunity to engage deeply with the material and demonstrate their understanding in a novel "show what you know" assessment. The real-world context and role-playing aspect may have supported SciID, as students envisioned themselves as part of a diverse scientific team.

Figure 4

Portion of Modified Midterm Exam Situating Students as the Biologist in an Interdisciplinary Design Team


You are the biologist in a design team that includes an engineer, an applied mathematician, and a start-up company with business managers and entrepreneurs. Please take the role of the author of the biology paper. Answer the team's questions based on your publication, proposed design invention, and Lectures 1-7.

Biologist University of California at Berkeley

Engineer CalTech

Applied Mathematician Rockefeller University

Entrepreneurs Silicon Valley

Question 4. 10 pts. The name of your start-up company is _Robofeats_______ .

Question 5. 10 pts. **BioDiscovery.** Your team asks:
What hypotheses are you testing in your publication?

We tested the hypothesis that geckos race along the water's surface by using a combination of strategies arising from slapping and stroking, undulation, and surface tension forces.

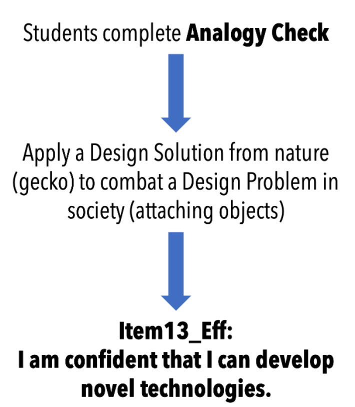
Considering these outcomes, our semester-by-semester analysis reveals the robustness of the CBL approach in fostering science connection across unprecedented learning contexts necessitated by a global pandemic. This approach was resilient across modalities and speaks to the adaptability and effectiveness of CBL as a framework for promoting science connection. While other studies have highlighted the challenges and disruptions faced by STEM students during the pandemic (NASEM, 2021), our findings further highlight the potential of CBL as an effective pedagogical approach, capable of fostering science connection even in challenging and rapidly changing educational environments.

Item Level Analysis Reveals Specific CBL Impacts on Science Connection Measures

Item level paired *t*-tests (Table 9) provided further insight on the trends observed in each of the individual constructs. All eight Eff items showed statistically significant increases. These gains closely corresponded with key course activities. The largest observed effect size was for Item 13 ($d = 0.620$; mean difference = +0.699), which assessed confidence in developing novel technologies. Throughout the course, students engaged in multiple design projects that were carefully scaffolded to build both their skills and confidence. A key principle of each design project was the development of novel technologies with potential for societal impact. Integrated into each of these projects was an *Analogy Check* exercise (Figure 5), which required students to use analogical reasoning to systematically translate biological principles into innovative design solutions (Full et al., 2021). Multiple iterations of this activity may have been particularly instrumental to promoting students' confidence in developing novel technologies.

Figure 5

Linking Analogy Check Exercise to Item13_Eff



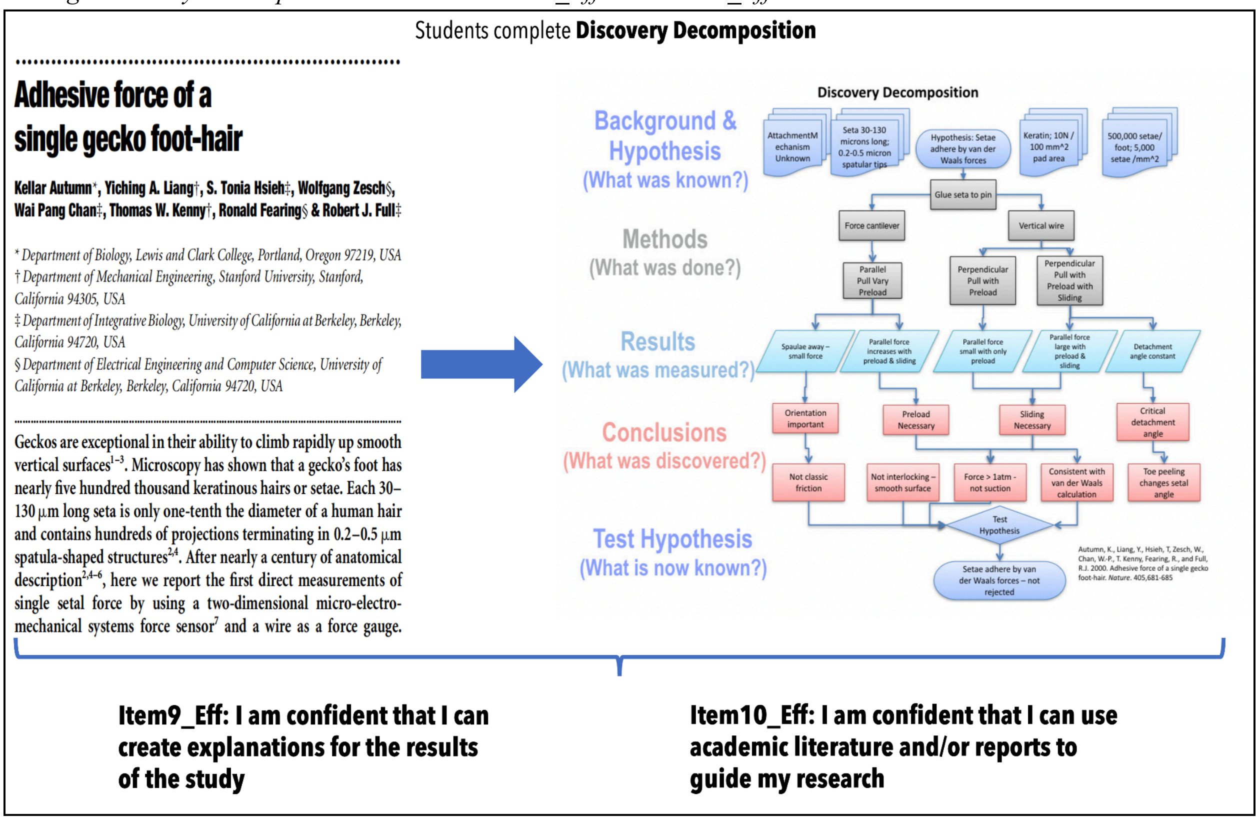
Design Solution	Analogy Check	Design Problem
Tokay Gecko (Gekko gecko)		**Hang and LeGECKO: Reusable Wall Hangers**
Function (What does system or organism do?)		**Function** (What do you want system to do?)
Attach gecko to any surface	Similar	Adhere objects to any surface
Structural Components (What is the structure or organization of system?)		**Structural Components** (What can the structure be?)
Array (about 1 billion) small structures (spatulae)	Uncertain	Comprised of multiple small projections (not as precise shape as spatulae)
Made of keratin	Similar	Made of silicone rubber (PDMS)
Operating Environment (Where?)		**Operating Environment** (Where?)
Polarizable semiconductor wafers (made of hydrophilic oxidized silicon or hydrophobic gallium arsenide)	Different	Walls of homes (plaster, drywall, stone, etc) that are probably non-polar
Smooth surface	Different but shouldn't matter	Smooth or rough
Size (What is size?)		**Size** (What size needed?)
Spatulae radii are 0.13-0.16μm	Different	Projections about 7000μm
Geckos are 20-40 cm in length and about 0.31-0.88 lbs in wieght	Different	One patch hold objects 61-91 cm and weigh up to 16 lbs
Mechanisms (How does system work?)		**Mechanisms** (How do you want the system to work?)
van der Waals forces	Similar	van der Waals forces
Perpendicular force as preload and sliding necessary for max adhesion	Similar	Push product onto wall and push thing to hang onto wall (perpendicular) then let hang (slide)
Peel toes to reach critical detachment angle to detach	Similar	Peel product off wall/surface to remove
Characteristics/Specification (Which are distinguishing?)		**Characteristics/Specification** (What are your specifications?)
Repeatedly adhere and detach to surfaces	Similar	Repeatedly adhere and detach to surfaces
Stick to smooth surfaces	Similar	Stick to smooth surfaces
Surface hydrophilicity does not matter	Similar	Surfaces have differing hydrophilicity
Dry adhesive	Similar	Dry adhesive
Performance Criteria (How well does system work?)		**Performance Criteria** (How well must the system work?)
~40 μN adhesion force per seta (~567,000 N/m^2)	Different but less weight	Hold ~4,600 N/m^2
Constraints (What compromises system?)		**Constraints** (Can compromises be removed?)
Historical/evolutionary constraint – made of keratin	Uncertain	Material can be varied
Historical/evolutionary constraint – no chemical adhesion	Different	Can use chemical adhesion
Multifunctional – used for walking so must attach and detach infinite number of times	Different	Need to be reattachable a magnitudinally fewer number of times

Note. Students analogize a Design Solution (left hand column) from nature to a societally relevant Design Problem (right hand column). Students check their analogy (compare and contrast design solution and problem) in the middle column by labeling it "similar," "different," or "uncertain." Adapted from S2 in Full et al. (2021).

Other Eff items of interest included those that assessed confidence in creating explanations for study results (Item 9; *d* = 0.306; mean difference = +0.238), using academic literature to guide research (Item 10; *d* = 0.243; mean difference = +0.245), and designing experiments to test hypotheses (Item 12; *d* = 0.310; mean difference = +0.312). The guided team design projects, such as the gecko-inspired adhesive and insect-inspired robot activities, provided opportunities for students to design and conduct experiments, potentially enhancing their self-efficacy in this area. Students' confidence in creating explanations for study results and examining academic literature may have improved due to the *Discovery Decomposition* exercise (Figure 6). In this activity, students extracted principles from scientific papers to visualize a logic flow that connected the study results back to the initial hypothesis (Full et al., 2021).

Figure 6

Linking Discovery Decomposition Exercise to Item9_Eff and Item10_Eff



Note. Students read original, published, scientific discoveries and formulate a logic flow that extracts fundamental principles with potential for translation into bioinspired designs. Adapted from S2 in Full et al. (2021).

For SciID, two out of five items showed statistically significant increases with small-to-medium effect sizes. Item 1 (*I have a strong sense of belonging to the community of scientists*; *d* = 0.222; mean difference = +0.308) and Item 3 (*I have come to think of myself as a scientist*; *d* = 0.220; mean difference = +0.302) demonstrated the most substantial changes. These results suggest that our CBL course was particularly effective in students developing a sense of belonging and identity within the scientific community. This may be attributed to the course's emphasis on engaging with real scientific literature, participating in societally relevant team-based projects, and presenting work in a public showcase, all of which mirror practices embodied by professional scientists.

The lack of significant change, or stability, in the Val construct is noteworthy. Results from the univariate Overall RM ANOVA (Figure 2C; Table 5) and the overall construct paired *t*-tests (Table 8) suggest that the course did not significantly impact pre/post development of Val. This finding aligned with previous work from Cole and Beck (2022), who also observed no significant changes in Val (despite increases in SciID and Eff) after a year-long introductory biology sequence. Additionally, at the item level (Table 9), 5/6 Val items showed no significant increases, with three items (Items 16-18) showing non-significant decreases. Considering these results, we suspect that the Val construct measured more stable, trait-like attitudes that were less susceptible to change within the timeframe of a single semester. In science education research, values have been considered more enduring than other constructs, requiring more time to substantially shift (Koballa & Glynn, 2013; Trenholm, 1989). This is supported by other TIMSI research showcasing the longitudinal development of Val (Estrada et al., 2021). Future longitudinal studies could provide insight into how Val evolves as students continue to engage with CBL communities beyond a single course. While a single course may not produce major shifts in Val, we suspect that the cumulative effect of multiple CBL experiences over time has the potential to yield more substantial changes.

From another perspective, the fact that Val scores did not decline suggests that our CBL approach effectively supported students' existing alignment with scientific community values. This maintenance of Val, especially for the many Non-STEM majors in our course engaging in STEM-based coursework, highlights the course's ability to sustain students' connection to scientific values while developing other aspects of their science connection (SciID and Eff). Additionally, while the overall Val construct remained stable, one Val item did show a statistically significant increase: *A person who believes writing up research results to be published in a leading scientific journal is a good use of time* (Item 15; $d = 0.136$; mean difference = +0.174). The growth in this specific item may be attributed to dedicated course activities that involved evaluating sources, learning about the publication process that leads to primary literature, and consistently referring to primary literature for design inspiration. Internalizing this value suggests students became more aware of the origin of facts, the need for discovery dissemination, and the overall significance of primary literature.

Chapter 1 Conclusion: Limitations and Future Research

This study examined the impact of a *Bioinspired Design* CBL course on students' science connection based on an application of the TIMSI framework. We observed statistically significant increases in SciID and Eff, while Val remained stable. We also saw significant improvements in science connection across all five semester iterations of the course spanning pre-, mid-, and post-COVID-19 pandemic timelines. While results from this study were encouraging, they represented only the first step in understanding the full impact of our CBL course. Limitations of this study and their connection to future research prospects are discussed below.

This study was conducted within a single course at one institution. Replication in other settings could enhance generalizability. Additionally, self-report survey data are subject to biases such as social desirability. Importantly, science connection is a subjective experience based on affective social influence constructs (SciID, Eff, and Val). Thus, in our assessment context, claims made based on affective self-report data are valid. Investigating affective aspects of student experiences is critical for understanding and improving learning in undergraduate science classrooms (Trujillo & Tanner, 2014). Nonetheless, we plan on conducting future research cross-validating the self-report data with analysis of student products. For example, the most notable Eff item—*I am confident that I can develop novel technologies*—can be further assessed in the context of student performance on course assignments that require this skill, such as the final project or midterm assessment. This follow-up analysis of student work would provide a more comprehensive understanding of science connection rooted in both subjective and objective measures.

Additionally, this study focused on overall outcomes without considering potential variations across different demographic populations of students. Given the inclusive nature of our course—open to all majors, all years, with no prerequisites—there is a critical opportunity to investigate how the diverse student population of the course responded to our CBL approach. The next chapter further investigates science connection development by assessing whether the observed growth from this chapter was promoted equitably across various demographic groups. By examining potential differential impacts, we can gain insights into how CBL might differentially affect students based on gender, underrepresented minority status, first-generation status, major, or class year. This exploration is needed to not only provide a better understanding of our course's impact, but also as a contribution to the broader need for creating inclusive undergraduate STEM learning environments that benefit all students.

Chapter 2

Using Challenge-Based Learning to Promote Equitable Science Connection in a *Bioinspired Design* Course

Abstract

Building on the overall growth in science connection observed in Chapter 1, this chapter investigates the equity of these outcomes across diverse student populations in a Challenge-Based Learning (CBL) course. We examined whether a CBL approach could promote equitable outcomes in science connection measures for 180 undergraduate students from over 40 majors across five semester iterations of our Bioinspired Design course. Using the Tripartite Integration Model of Social Influence (TIMSI) framework, we assessed changes in Science Identity (SciID), Science Self-Efficacy (Eff), and Internalization of Scientific Community Values (Val) across seven demographic variables (gender, underrepresented minority status, first-generation status, STEM/Non-STEM major, Biology/Not Biology major, class status, and term). We expanded our statistical approach from Chapter 1 by employing repeated measures ANOVAs with demographic variables as between-subjects factors and introducing ANCOVAs to control for pre-survey scores. Results revealed that the significant increases in SciID/Eff and the stability in Val observed in Chapter 1 were equitable across all demographic groups, with exceptions in SciID development based on STEM/Non-STEM status and class status. Findings from this chapter suggest that our CBL approach effectively promoted science connection development across diverse student populations, challenging the notion that such development requires extensive time or is limited to specific demographic groups. The chapter highlights the potential of CBL as an inclusive pedagogical model that can foster equitable science connection for all students within a single semester. In conjunction with the previous chapter, this research supports a view of science connection as a continuous process that can be cultivated through participation in various types of CBL communities, such as large enrollment breadth courses. Overall, this work supports creating more inclusive STEM education experiences by emphasizing the importance of promoting science connection for all students in the ongoing effort to build a more scientifically literate society capable of addressing complex global challenges.

Background

In Chapter 1, we explored the impact of our Challenge-Based Learning (CBL) course, *Bioinspired Design* (Full et al., 2021), on fostering science connection among undergraduate students. Using the Tripartite Integration Model of Social Influence (TIMSI) framework, we observed significant increases in Science Identity (SciID) and Science Self-Efficacy (Eff), while Internalization of Scientific Community Values (Val) remained stable. These gains persisted across varying instructional modalities necessitated by the COVID-19 pandemic, demonstrating the resilience of the CBL approach. However, these overall findings did not address potential variations in outcomes across different demographic groups—a critical consideration for promoting STEM-enriched learning for all students. To capitalize on this research opportunity, the current chapter investigated whether the observed growth in science connection was promoted equitably across diverse student populations.

To further assess the key developmental results from Chapter 1 and contribute to our understanding of equitable science education, Chapter 2 focuses on the following aims:

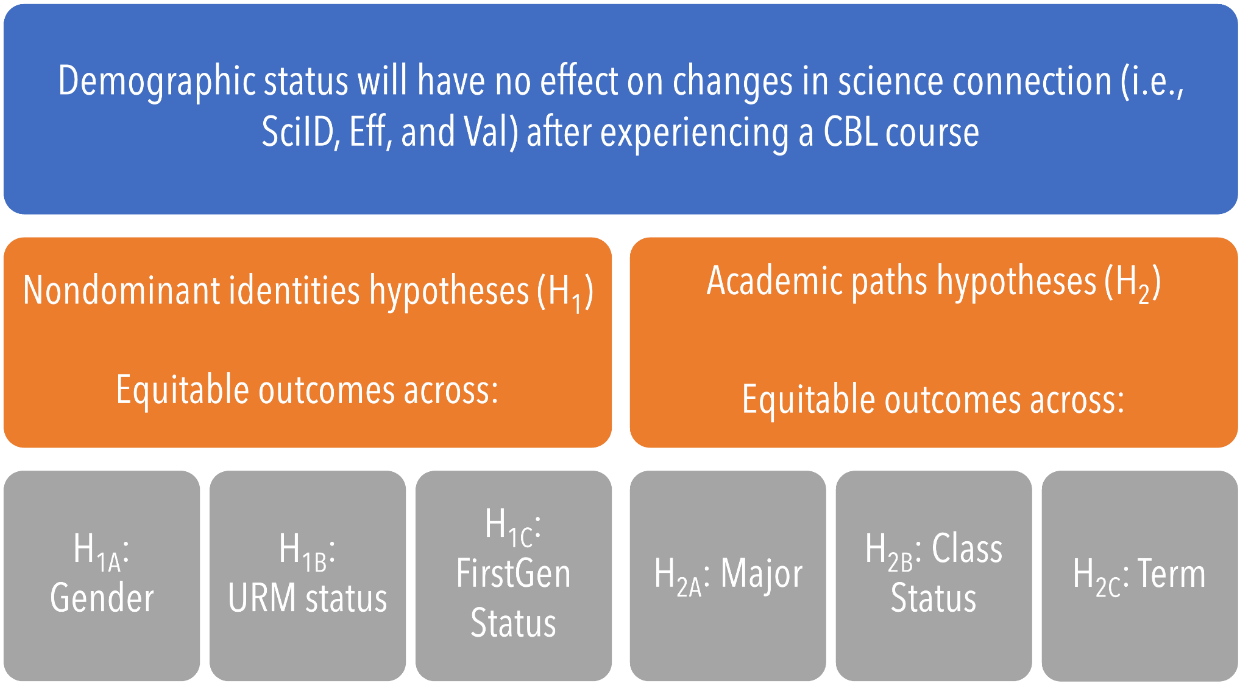
> Aim 1: Examine the effect of demographic groups on the development of science connection (SciID, Eff, and Val) in our CBL course.
> Aim 2: Evaluate the effectiveness of our CBL approach in promoting equitable science connection outcomes across seven key demographic variables.

By centering these aims, we strive to provide a more comprehensive understanding of the impact of our CBL approach on diverse student populations and contribute to the broader discourse on equity in STEM education.

Hypotheses for Equitable Science Connection Development in a Diverse CBL Environment

The unique student composition of our *Bioinspired Design* course offers a critical opportunity to investigate the development (i.e., change, or growth) of science connection in a unique CBL environment open to all students. Given that students from diverse backgrounds may respond differently to evidence-based teaching methods (Ballen et al., 2017b), we recognized the crucial need to assess our CBL course-based intervention for potentially differential impacts across various demographic identities. Therefore, we aimed to explore how key demographic variables may have impacted science connection development in our students. The specific demographic variables of interest in this study had previously documented impacts in the context of SciID, Eff, and Val development (e.g., Cole & Beck, 2022). We intentionally chose these variables to connect with the prior literature and predicted equitable growth outcomes for students in the course regardless of demographic classification. Thus, we investigated the following **Equity hypothesis**: *Demographic status will have no effect on changes in science connection (i.e., SciID, Eff, and Val) after experiencing a CBL course.* Below, we break down this Equity hypothesis into two nested hypotheses differentiated by their thematic grouping of seven key demographic variables (Figure 1).

Figure 1
Visual Representation of the Equity Hypothesis and its Sub-Hypotheses



Note. Equity hypothesis and its sub-hypotheses for analyzing science connection development across diverse student populations in a Challenge-Based Learning (CBL) course. The Equity hypothesis is broken down into Nondominant identities hypotheses (H_1) and Academic paths hypotheses (H_2), each examining specific demographic variables.

> **Nondominant identities hypothesis; H_1.** Our CBL course will promote equitable outcomes in SciID, Eff, and Val development regardless of students' gender, underrepresented minority (URM) status, or first-generation status (Figure 1). Specifically, we tested three distinct statistical sub-hypotheses within H_1: H_{1A}-Gender (Female/Male), H_{1B}-URM status (URM/Not URM), and H_{1C}-First-generation status (FirstGen/Not FirstGen) will have no effect on changes in science connection after experiencing a CBL course.

Our goal of implementing equitable and inclusive pedagogy for all students results in a need to investigate learning interventions for potential differential impacts on students with nondominant[1] identities (Ballen et al., 2017b; Cole & Beck, 2022; Shortlidge et al., 2024; Vincent-Ruz et al., 2018). Historically, students from nondominant backgrounds, especially based on gender, URM status, and first-generation status, have faced barriers to integration, and by extension, science connection (Estrada et al., 2011; Estrada et al., 2016; Hazari et al., 2013; Li et al., 2024; Shortlidge et al., 2024). These barriers negatively impact key social influence constructs of science connection, such as SciID, Eff, and Val (Cole & Beck, 2022; Estrada et al., 2021). Previous studies have found that perceptions of gender stereotypes and racial stigmas

[1] Our choice in using this specific term aligns with the reasoning outlined by Gutiérrez et al. (2009): "We use the term *nondominant* rather than terms such as *minority*, *students of color*, and so on, given that the central issue is the power relations between those who are in power and those who, despite their growing census numbers, are not."

within the sciences negatively influence students' SciID (Hazari et al., 2013; Shortlidge et al., 2024). Women and URM students, even those intending to pursue STEM careers, are less likely to see themselves as a "science person" compared to their male and non-URM counterparts (Hazari et al., 2013; Vincent-Ruz et al., 2018). Females in STEM also report lower Eff than their male counterparts, even when actual academic performance is held constant (Bloodhart et al., 2020; Marshman et al., 2018; Robnett & Thoman, 2017). In a study by Cole and Beck (2022), URM and female students entered an introductory biology course sequence with lower SciID and Eff compared to their non-URM and male counterparts and despite experiencing similar rates of growth, these initial disparities persisted. This result aligned with prior observations of lower SciID or Eff in URM students as compared to non-URM students (Adedokun et al., 2013; Beltran et al., 2020; Cole & Beck, 2022; Hazari et al., 2013; MacPhee et al., 2013). In terms of Val, students from nondominant backgrounds may lack systemic exposure to scientific norms and practices, thereby experiencing a perceived misalignment between their personal values and those of the scientific community (Estrada et al., 2018). Thus, students with nondominant identities may report lower endorsement of Val compared to their peers (Shortlidge et al., 2024). For example, Cole and Beck (2022) observed that first-generation students had lower levels of Val that continued to decrease over time, highlighting the need for targeted interventions to support the development of Val in this nondominant population.

These ongoing demographic disparities in SciID, Eff, and Val have led to decreased persistence and engagement in STEM fields, especially in nondominant populations (Chemers et al., 2011; Estrada et al., 2011; Estrada et al., 2018). These findings confirm the need for creating equitable and inclusive learning environments that actively work to mitigate the negative impacts of systemic barriers on students' science connection, particularly for students from nondominant backgrounds. By cultivating equitable growth in SciID, Eff, and Val through CBL, we can propel *all* students—from both dominant and nondominant backgrounds—to develop deeper connections to science.

Our next nested hypothesis recognized that students in our course were coming from a wide variety of academic paths differentiated by their disciplinary identities, specific stage in college, and distinct semesters of enrollment. Given the heterogeneity of these academic paths, we sought to investigate how these varied identities, stages, and semester contexts may influence the development of science connection. Thus, we tested:

> **Academic paths hypothesis; H_2.** Our CBL course will effectively promote science connection for all students, regardless of their disciplinary identity, class status, or the specific semester in which they take the course (Figure 1). Specifically, we tested three distinct statistical sub-hypotheses within H_2: H_{2A}-Intended/declared major status (STEM/Non-STEM; Biology/Not Biology), H_{2B}- Class status (Lowerclassmen/Upperclassmen), and H_{2C}-Term (semester iteration) will have no effect on changes in science connection after experiencing a CBL course.

This nested hypothesis addresses three key dimensions of students' academic paths: 1) disciplinary identity, defined by their intended or declared major; 2) stage in college journey, defined by their class status as a lowerclassmen (Years 1 and 2) or upperclassmen (Years 2 and beyond); and 3) specific semester in which they took the course, differentiated by instructional impact of the COVID-19 pandemic. Exploring these dimensions provides critical insight into the effectiveness of our CBL course in promoting science connection development for the diverse academic paths represented in our student population.

The TIMSI framework proposes that students' integration into the scientific community is shaped by their experiences within that community (Estrada et al., 2011). Thus, students' level of exposure to and engagement with scientific disciplines may impact their SciID, Eff, and Val. Previous research has investigated potential differences between STEM and Non-STEM majors (i.e., majors and nonmajors), as well as between majors within STEM fields. Hazari et al. (2013) emphasized the importance of examining specific disciplinary identities (e.g., biology identity, chemistry identity, and physics identity) to better understand differences in participation and engagement between fields. More broadly, various attitudinal differences have been observed between STEM and Non-STEM majors. Non-STEM majors are less likely than biology majors to see science as personally relevant (Cotner et al., 2017). However, STEM majors tend to place less importance on social agency and working for social change compared to Non-STEM majors (Garibay, 2015). Interestingly, despite these attitudinal differences, nonmajors and majors may possess similar levels of science process skills at the beginning of their undergraduate studies (Hebert & Cotner, 2019). This implies that both groups have the capacity to participate in authentic scientific inquiry, and educators should prioritize instructional approaches that address attitudinal disparities between nonmajors and majors to advance scientific literacy (Hebert & Cotner, 2019; Knight & Smith, 2010). This aligns with calls for engaging Non-STEM majors in meaningful scientific inquiry to develop a scientifically literate citizenry capable of tackling the socioscientific issues of the future (Ballen et al., 2017a; Gormally & Heil, 2022). By assessing how our CBL course impacts science connection development across different disciplinary identities, we aim to enrich understanding of effective methods for promoting scientific literacy and engagement among undergraduate students from all disciplines.

Students' stage in their college journey, as defined by their class status (i.e., year in college), may also influence their science connection. As students progress through college, their self-perceptions with respect to science are likely to become more stable (Hazari et al., 2013). Thus, upperclassmen may have more rigid conceptions of their SciID, Eff, and Val that are resistant to change compared to lowerclassmen who may express more malleable conceptions in these constructs. Estrada et al. (2011) hypothesized that graduate students, being more socialized into the scientific community, would exhibit greater SciID, Eff, and Val compared to undergraduates. Extending this notion to undergraduates, upperclassmen—who have more knowledge, training, and exposure to their fields—might express stronger SciID, Eff, and Val compared to lowerclassmen. Knight and Smith (2010) investigated the effect of class standing on attitudinal changes in a genetics course, finding that upperclassmen exhibited a significant positive shift in attitudes towards science compared to underclassmen. These findings suggest that class status can impact students' responsiveness to instructional interventions geared toward developing attitudinal constructs like SciID, Eff, and Val. Additionally, in the specific timeline of our study, the semester in which students took the course may have influenced their science connection due to variations in instructional modality instigated by the COVID-19 pandemic. By assessing the potential impact of these academic path variables in our study, we strive to further support the tailoring of interventions that meet the needs of students at different stages of their academic journey.

Methods

Study Design

We employed a pre/post survey design to measure changes in TIMSI constructs (SciID, Eff, and Val) among students participating in the *Bioinspired Design* course, with a focus on examining potential differences across demographic groups. This approach allowed for both within-subjects comparisons of students' pre- and post-course responses and between-subjects analyses based on demographic variables. Data were collected from course iterations over five semesters (Spring 2019-Spring 2023), providing a sufficient sample size for investigating demographic trends. We analyzed matched pre/post pairs from students who completed both surveys, with sample sizes ranging from $N = 315$ to $N = 529$ depending on the specific statistical test. To encourage participation, students received one point of extra credit for voluntarily completing the pre- and post-surveys online via Qualtrics. The survey protocol included informed consent and confidentiality assurances, and all experimental procedures were approved by the Institutional Review Board (Protocol ID: 2017-12-10602).

Survey Design and Demographic Data Collection

This chapter utilizes the same survey response data from Chapter 1 based on the adapted TIMSI survey items. See Chapter 1-Table 2 for the complete survey instrument (19 Likert-type items with SciID, Eff, and Val subscales). As discussed in Chapter 1, the survey demonstrated good to excellent reliability at the overall and individual construct levels based on both Cronbach's alpha and McDonald's omega values, indicating high internal consistency of the instrument (See Chapter 1-Table 3 and Supplement 2).

In addition to pre/post survey results from the SciID, Eff, and Val measures, this chapter also presents key demographic data linked to our Equity hypotheses. Demographic survey items collected students' potentially nondominant identities (Gender, URM status, FirstGen status) and academic paths (Intended/Declared Major Status, Class Status, Term). Each of these demographic variables were split into dichotomous groups to test for statistically significant differences in pre/post survey results. STEM/Non-STEM was categorized based on responses to the selected choice survey item "*Major field of study or interest.*" STEM majors were students who selected 1) *Science*, 2) *Technology*, 3) *Engineering*, 4) *Mathematics*, or 5) *Computer Science*; Non-STEM majors were students who selected 1) *Design*, 2) *Uncertain/not sure*, and 3) *Other*. As a breadth course open to all students, we categorized "*Other*" as Non-STEM to capture the full range of majors beyond the designated categories. This ensured a comprehensive sampling of the course's diverse student body and enabled STEM versus Non-STEM group comparisons in our analysis. Response data for all demographic variables are presented within the Results section alongside the associated statistical analysis.

We recognize the limitations of binary classification for each of these demographic groups, particularly for gender and URM status. We want to make clear that the gender demographic item allowed students to select from male, female, non-binary, and other. Students also selected from disaggregated racial identities (e.g., disaggregated subgroups within the "Asian" demographic) that were later aggregated into URM/Non-URM based on the National Science Foundation (NSF) (2023) definition of underrepresented minorities—"Races or ethnicities whose representation in STEM employment and S&E [science & engineering] education is smaller than their representation in the U.S. population. This includes Blacks or African Americans, Hispanics or Latinos, and American Indians or Alaska Natives." There are known limitations with this definition, particularly in the context of undergraduate STEM education (Bhatti, 2021). We chose to use this definition to align with the demographic standards of the TIMSI framework literature. We encourage future research to consider alternatives to traditional notions of URM/Non-URM.

Statistical Analysis

For all statistical analyses, we used IBM SPSS Statistics for Macintosh (Version 28.0, IBM Corp., 2021). We calculated descriptive statistics for individual survey items and overall constructs (See Chapter 1-Results), and assessed normality using histograms, skewness, and kurtosis values. Given our large sample size, we proceeded with ANOVA and ANCOVA parametric tests, which are generally robust to normality violations (Blanca et al., 2017; Blanca et al., 2023). We generated visualizations of our ANCOVA results using R (version 4.4.1; R Core Team, 2024) with the *ggplot2*, *dplyr*, and *gridExtra* packages (Wickham, 2016) to create high-contrast dot plots with error bars for each construct. We conducted Box's test (Box, 1949) for homogeneity of covariance and Levene's test (Levene, 1960) for homogeneity of variance across groups to evaluate assumptions for group comparisons, with results detailed in Supplement 3. We mitigated potential for Type I errors in multiple comparisons by applying Bonferroni corrections to all relevant statistical tests.

To examine potential differences across the seven dichotomous demographic groups, we conducted a set of RM ANOVAs labeled "Demographic RM ANOVAs," or those with *time* as the within-subjects factor and *demographic group variables* as additional between-subjects factors. These RM ANOVAs tested H_1 (nondominant identities hypotheses) and H_2 (academic paths hypotheses). Similar to the RM ANOVA analysis in Chapter 1, multivariate and univariate tests examined potential differences in pre-to-post changes across the demographic groups at the combined and individual construct levels, respectively.

In contrast to Chapter 1, we recognized that the hypotheses in this chapter centered on investigating the absence of statistically significant differences (i.e., null hypotheses) based on demographic groups. Thus, we sought to strengthen our analytical approach by also conducting analyses of covariance (ANCOVAs) to investigate potential variations in course impact between demographic groups while controlling for pre-survey scores. This additional approach enhanced our ability to detect effects that may not have been apparent in the RM ANOVA alone (Dimitrov & Rumrill Jr, 2003; Maxwell et al., 2017). Importantly, due to potential limitations in statistical power, the absence of statistically significant differences does not definitively equate to lack of demographic differences. However, the combined use of RM ANOVA and ANCOVA provides a thorough assessment of our null hypotheses. By examining both changes over time and adjusted post-survey scores, our dual methodology strengthened the validity of findings that aligned with our hypotheses of no demographic differences in science connection development.

Results

In the results tables below, dashed lines within the tables separate the demographic variables for H_1 (nondominant identities hypotheses; $H_{1A\text{-}1C}$) and H_2 (academic paths hypotheses; $H_{2A\text{-}2C}$). The descriptive statistics (means and standard deviations) for each of the pre/post survey constructs can be found in Chapter 1-Table 4. The demographic response data for the RM ANOVAs are shown in Table 1 below.

Table 1
Demographic Response Data for RM ANOVAs

Demographic	Variable	N
Gender	Male	192
	Female	292
URM Status	URM	48
	Non-URM	438
FirstGen Status	Yes	45
	No	265
Biology Major	Yes	137
	No	356
STEM/Non-STEM	STEM	329
	Non-STEM	164
Class Status	Lowerclassmen	293
	Upperclassmen	201
Term	2019	127
	2020	127
	2021	97
	2022	72
	2023	71

No Significant Differences in Science Connection Development Across Demographic Group RM ANOVAs

Multivariate Tests Show Equitable Growth Across All Demographic Variables. Prior to analysis, we ran assumption tests including Box's test and Levene's test (Supplement 3-Tables S2-S4). Based on our assumption test results and the robustness of multivariate and univariate tests to departures from assumptions (Field, 2024; Supplement 3), we proceeded with our planned analyses. We first tested the nondominant identities hypotheses ($H_{1A\text{-}1C}$) and the academic paths hypotheses ($H_{2A\text{-}2C}$) by conducting multivariate RM ANOVA tests with each demographic variable as a between-subjects factor. This resulted in seven tests (one for each demographic group variable) as shown in Table 2. A Bonferroni corrected critical value of $p < .05/7$ was applied to all tests. We observed no statistically significant differences between any of the seven dichotomous demographic groups, supporting both H_1 and H_2.

Table 2
Demographic RM ANOVAs - Multivariate

Variable		Value	F	Hypothesis df	Error df	Sig.*	Partial Eta Squared
PrePost * Gender	Pillai's Trace	.003	.547	3	480	.650	.003
	Wilks' Lambda	.997	.547	3	480	.650	.003
PrePost * URM	Pillai's Trace	.002	.343	3	482	.794	.002
	Wilks' Lambda	.998	.343	3	482	.794	.002
PrePost * FirstGen	Pillai's Trace	.002	.184	3	306	.908	.002
	Wilks' Lambda	.998	.184	3	306	.908	.002
PrePost * Biology Major	Pillai's Trace	.005	.878	3	489	.452	.005
	Wilks' Lambda	.995	.878	3	489	.452	.005
PrePost * STEM/Non-STEM	Pillai's Trace	.007	1.211	3	489	.305	.007
	Wilks' Lambda	.993	1.211	3	489	.305	.007
PrePost * Class Status	Pillai's Trace	.013	2.073	3	490	.103	.013
	Wilks' Lambda	.987	2.073	3	490	.103	.013
PrePost * Term	Pillai's Trace	.028	1.132	12	1467	.330	.009
	Wilks' Lambda	.973	1.130	12	1288	.331	.009

*Bonferroni correction applied ($p < .05/7$)

Univariate Analyses Further Support Equitable Development in Individual Social Influence Constructs. We followed up these multivariate tests with univariate tests that also analyzed potential demographic differences at a more granular level of each individual social influence construct. This resulted in 21 tests (3 constructs x 7 demographic group variables) as shown in Table 3. A Bonferroni corrected critical value of $p < .05/21$ was applied to all tests. In alignment with the multivariate test, we observed no statistically significant differences across all 21 univariate tests, further supporting equitable developmental outcomes in science connection regardless of demographic classification.

Table 3
Demographic RM ANOVAs - Univariate

Variable	Construct	Type III Sum of Squares	df	Mean Square	F	Sig.*	Partial Eta Squared
PrePost * Gender	SciID	.008	1	.008	.016	.900	.000
	Eff	.328	1	.328	1.325	.250	.003
	Val	.195	1	.195	.649	.421	.001
PrePost * URM	SciID	.139	1	.139	.266	.606	.001
	Eff	.134	1	.134	.537	.464	.001
	Val	.245	1	.245	.808	.369	.002
PrePost * FirstGen	SciID	.009	1	.009	.017	.897	.000
	Eff	.052	1	.052	.224	.636	.001
	Val	.046	1	.046	.151	.698	.000
PrePost * Biology Major	SciID	.980	1	.980	1.876	.171	.004
	Eff	.021	1	.021	.084	.772	.000
	Val**	.467	1	.467	1.579	.209	.003
PrePost * STEM/Non-STEM	SciID	.140	1	.140	.267	.605	.001
	Eff	.603	1	.603	2.411	.121	.005
	Val**	.060	1	.060	.199	.656	.000
PrePost * Class Status	SciID	.543	1	.543	1.040	.308	.002
	Eff	.539	1	.539	2.159	.142	.004
	Val	.399	1	.399	1.330	.249	.003
PrePost * Term	SciID	2.402	4	.600	1.151	.332	.009
	Eff	1.978	4	.495	1.992	.095	.016
	Val	.648	4	.162	.538	.708	.004

*Bonferroni correction applied ($p < .05/21$)
**Val failed Levene's test in Bio and STEM

ANCOVA Results Support Equitable Outcomes with Exceptions in SciID

ANCOVA analyses examined whether any differences in post-survey scores across demographic groups were significant after controlling for pre-survey scores. These tests provided an additional and alternative assessment of the Equity hypotheses that adjusted for baseline variations. Prior to conducting each ANCOVA, Levene's test was performed to assess the assumption of homogeneity of variances across groups (Supplement 3-Tables S5-S7). If this assumption was not met, robust standard errors were used instead. Robust standard errors provide unbiased standard errors of the parameter estimates when the homoscedasticity assumption is not met (White, 1980), allowing for valid interpretation of the significance tests. A Bonferroni corrected critical value of $p < .05/7$ was applied to all tests. The demographic response data for the ANCOVAs are shown in Tables 4 and 5. ANCOVA results for each social influence construct (SciID, Eff, and Val) are presented in separate subsections below.

Table 4
Demographic Response Data for ANCOVA Tests

Variable	SciID	Eff	Val
Gender			
Male	208	202	199
Female	305	303	296
URM Status			
URM	50	50	48
Non-URM	464	457	450
FirstGen			
Yes	47	45	46
No	275	274	269
Biology Major			
Yes	148	149	141
No	376	367	365
STEM/Non-STEM			
STEM	310	346	341
Non-STEM	215	170	165
Class Status			
Lowerclassmen	293	306	300
Upperclassmen	201	211	207

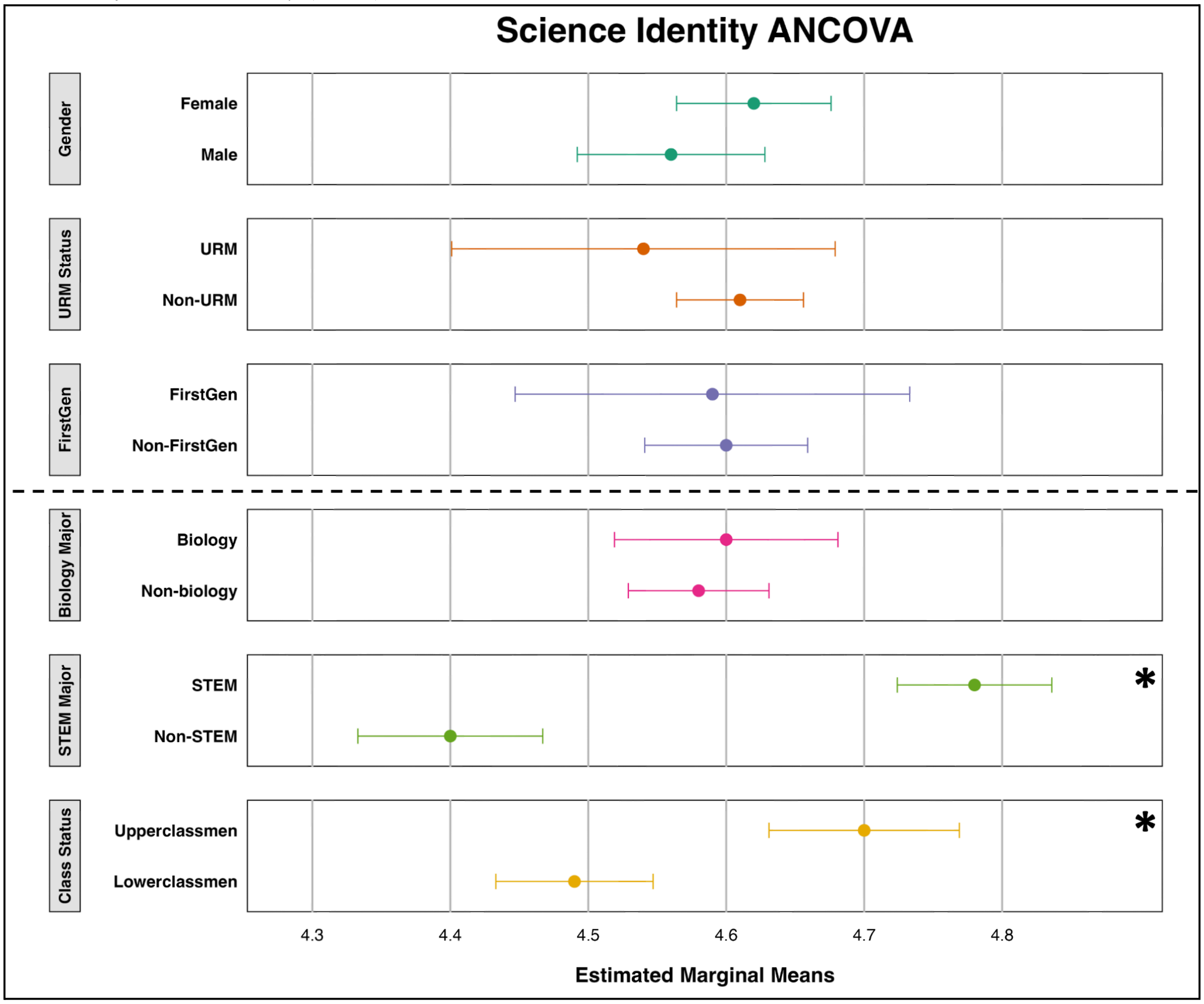
Table 5
Response Data (N) for Term ANCOVA

Term	SciID	Eff	Val
2019	132	131	130
2020	133	132	129
2021	104	101	101
2022	81	79	76
2023	75	74	71

SciID ANCOVA: Statistically Significant Variations Based on STEM/Non-STEM and Class Status. Levene's test showed homogeneity of variance was met for all covariates (Supplement 3-Table S5). SciID ANCOVA results are shown in Figure 2 and Table 6. For purposes of brevity, the effect of the Term demographic variable is shown only in the ANCOVA results tables. There were no significant differences based on Term across SciID, Eff, and Val. Testing H_1 (nondominant identities), there were no significant differences based on Gender, URM status, and FirstGen status in the SciID ANCOVA. Testing H_2 (academic paths), there were no significant differences based on Biology Major and Term. There was a statistically significant difference between STEM (M = 4.78, SE = 0.056) and Non-STEM (M = 4.40, SE = 0.067) with small-medium effect size; $F(1, 522) = 23.331$, $p < .001$, $\eta_p^2 = 0.043$. There was also a statistically significant difference based on Class Status between lowerclassmen (M = 4.49, SE = 0.057) and upperclassmen (M = 4.70, SE = 0.069) with small effect size; $F(1, 523) = 8.547$, $p = .004$, $\eta_p^2 = .016$.

Figure 2

Dot Plots of Science Identity (SciID) ANCOVA Results



Note. The plots display estimated marginal means for groups based on nondominant identity demographics (Gender, URM Status, FirstGen Status) and academic path demographics (Biology Major, STEM/Non-STEM, Class Status) separated by a dashed line. Asterisks (*) denote statistically significant differences between the compared groups after applying a Bonferroni correction ($p < .05/7$). Error bars represent 95% confidence intervals. The Likert scale ranges from 1 (strongly disagree) to 7 (strongly agree). Comparisons show that STEM/Non-STEM and Lowerclassmen/Upperclassmen groups had significant differences, while other demographic groups did not show statistically significant differences.

Table 6

SciID ANCOVA Results

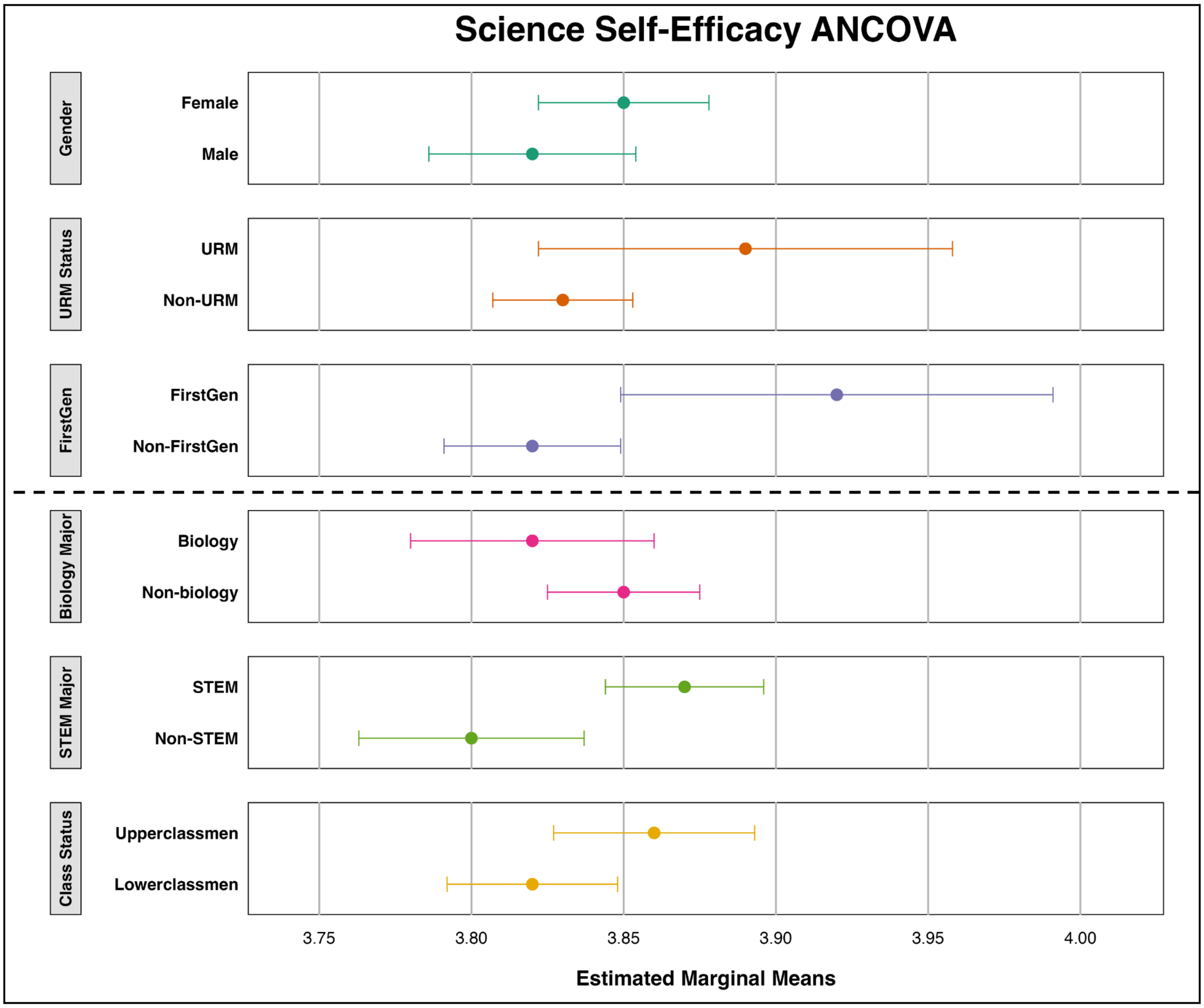
Variable	Type III Sum of Squares	F	Sig.*	Partial Eta Squared	Effect Size
Gender	.569	.607	.436	.001	N/A
URM	.177	.189	.664	.000	N/A
FirstGen	.002	.002	.962	.000	N/A
Biology Major	.046	.049	.825	.000	N/A
STEM/Non-STEM	**20.961**	**23.331**	**<.001**	**.043**	**Small-med**
Class Status	**7.878**	**8.547**	**.004**	**.016**	**Small**
Term	4.166	1.115	.349	.009	N/A

*Bold indicates significant difference; Bonferroni correction applied ($p < .05/7$)

Eff ANCOVA: Equitable Outcomes Across All Groups. Levene's test showed homogeneity of variance was met for all covariates except Class Status (Supplement 3-Table S6). To account for this violation, the Class Status ANCOVA was conducted using robust standard errors. Testing H_1 and H_2, no statistically significant differences were detected across all demographic variables (Figure 3; Table 7).

Figure 3

Dot Plots of Science Self-Efficacy (Eff) ANCOVA Results



Note. The plots display estimated marginal means for groups based on nondominant identity demographics (Gender, URM Status, FirstGen Status) and academic path demographics (Biology Major, STEM/Non-STEM, Class Status) separated by a dashed line. Bonferroni correction applied ($p < .05/7$) to all analyses. Error bars represent 95% confidence intervals. Likert scale range 1 (not at all confident) to 5 (absolutely confident). Comparisons between demographic groups showed no statistically significant differences.

Table 7
Eff ANCOVA Results

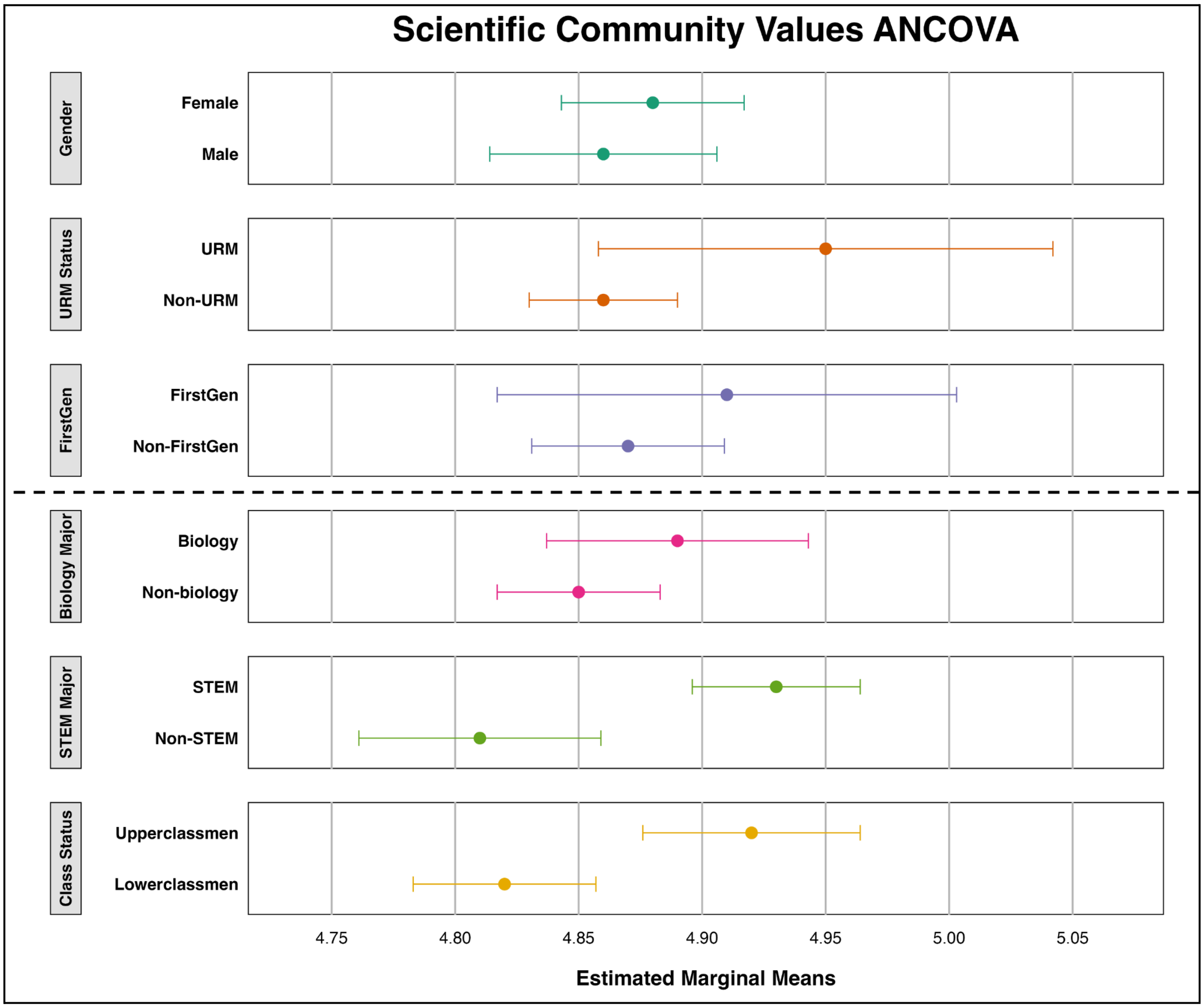
Variable	Type III Sum of Squares	F	Sig.*	Partial Eta Squared	Effect Size
Gender	.167	.458	.499	.001	N/A
URM	.391	1.076	.300	.002	N/A
FirstGen	.714	1.967	.162	.006	N/A
Biology Major	.158	.437	.509	.001	N/A
STEM/Non-STEM	.951	2.633	.105	.005	N/A
Class Status**	.350	.965	.306	.002	N/A
Term	1.743	1.204	.308	.009	N/A

*Bonferroni correction applied ($p < .05/7$)
**Sig. value adjusted based on robust standard error due to violation of Levene's test

Val ANCOVA: No Statistically Significant Demographic Variations. Levene's test showed homogeneity of variance was met for all covariates except STEM/Non-STEM (Supplement 3-Table S7). As a result, the STEM/Non-STEM ANCOVA was conducted using robust standard errors to account for this violation. Testing H_1, there were no significant differences based on Gender, URM status, or FirstGen Status. Testing H_2, there were no significant differences based on Biology Major, Class Status, and Term (Figure 4; Table 8). Applying the robust standard errors method to the STEM/Non-STEM ANCOVA resulted in an adjusted significance value of .030, which was greater than the Bonferroni corrected critical value of $p < .05/7$. Thus, there was no statistically significant difference between STEM and Non-STEM on the Val ANCOVA.

Figure 4

Dot Plots of Scientific Community Values (Val) ANCOVA Results



Note. The plots display estimated marginal means for groups based on nondominant identity demographics (Gender, URM Status, FirstGen Status) and academic path demographics (Biology Major, STEM/Non-STEM, Class Status) separated by a dashed line. Bonferroni correction applied ($p < .05/7$) to all analyses. Error bars represent 95% confidence intervals. Likert scale range 1 (not like me at all) to 6 (very much like me). Comparisons between demographic groups showed no statistically significant differences.

Table 8

Val ANCOVA Results

Variable	Type III Sum of Squares	F	Sig.*	Partial Eta Squared	Effect Size
Gender	.050	.092	.761	.000	N/A
URM	.507	.924	.337	.002	N/A
FirstGen	.080	.150	.699	.000	N/A
Biology Major	.173	.322	.570	.001	N/A
STEM/Non-STEM**	3.050	5.633	.030	.009	N/A
Class Status	1.715	3.152	.076	.006	N/A
Term	1.926	.880	.475	.007	N/A

*Bonferroni correction applied ($p < .05/7$)

**Sig. value adjusted based on robust standard error due to violation of Levene's test

A summary of all ANCOVA results across each social influence construct—SciID, Eff, Val—is shown in Table 9 (excluding Term, see Chapter 1). Overall, 19 out of 21 ANCOVA tests showed no statistically significant differences based on demographic groups. Viewed collectively, both the results from the RM ANOVAs (no statistically significant differences in all 7 multivariate tests and all 21 univariate tests) and these ANCOVA results align with our Equity hypothesis.

Table 9
Summary of ANCOVA Results for SciID, Eff, and Val

Construct	Variable	Group 1 (M, SE)	Group 2 (M, SE)	F	Sig.*	η_p^2
SciID	Gender	Male (4.56, 0.068)	Female (4.62, 0.056)	0.607	.436	.001
	URM	URM (4.54, 0.139)	Non-URM (4.61, 0.046)	0.189	.664	.000
	FirstGen	FirstGen (4.59, 0.143)	Non-FirstGen (4.60, 0.059)	0.002	.962	.000
	Biology Major	Bio (4.60, 0.081)	Non-Bio (4.58, 0.051)	0.049	.825	.000
	STEM/Non-STEM	**STEM (4.78, 0.056)**	**Non-STEM (4.40, 0.067)**	**23.331**	**<.001**	**.043**
	Class Status	**Lower (4.49, 0.057)**	**Upper (4.70, 0.069)**	**8.547**	**.004**	**.016**
Eff	Gender	Male (3.82, 0.034)	Female (3.85, 0.028)	0.458	.499	.001
	URM	URM (3.89, 0.068)	Non-URM (3.83, 0.023)	1.076	.300	.002
	FirstGen	FirstGen (3.92, 0.071)	Non-FirstGen (3.82, 0.029)	1.967	.162	.006
	Biology Major	Bio (3.82, 0.040)	Non-Bio (3.85, 0.025)	0.437	.509	.001
	STEM/Non-STEM	STEM (3.87, 0.026)	Non-STEM (3.80, 0.037)	2.633	.105	.005
	Class Status**	Lower (3.82, 0.028)	Upper (3.86, 0.033)	0.965	.306	.002
Val	Gender	Male (4.86, 0.046)	Female (4.88, 0.037)	0.092	.761	.000
	URM	URM (4.95, 0.092)	Non-URM (4.86, 0.030)	0.924	.337	.002
	FirstGen	FirstGen (4.91, 0.093)	Non-FirstGen (4.87, 0.039)	0.150	.699	.000
	Biology Major	Bio (4.89, 0.053)	Non-Bio (4.85, 0.033)	0.322	.570	.001
	STEM/Non-STEM**	STEM (4.93, 0.034)	Non-STEM (4.81, 0.049)	5.633	.030	.009
	Class Status	Lower (4.82, 0.037)	Upper (4.92, 0.044)	3.152	.076	.006

*Bold indicates significant difference; Bonferroni correction applied ($p < .05/7$)
**Sig. value adjusted based on robust standard error due to violation of Levene's test

Discussion

This chapter expanded on our initial growth findings from Chapter 1 by exploring how the development of science connection (SciID, Eff, and Val) potentially varies across different demographic groups within our *Bioinspired Design* course. We also sought to understand if a course based on a CBL perspective could mitigate previously observed disparities in these measures and promote equitable outcomes across diverse student populations. During this process, we examined how specific demographic variables interacted with the development of SciID, Eff, and Val in the context of our unique CBL environment. By investigating these critical areas, we aimed to provide a comprehensive assessment of equity within our course framework. Our analysis focused on both nondominant identities (gender, URM status, FirstGen status) and academic paths (major, class status, enrollment term), allowing us to explore potential differences across a wide spectrum of student backgrounds and experiences.

The results of our multiple analyses in this chapter revealed equitable outcomes across these diverse student groups, suggesting that our CBL approach may offer a promising model for inclusive STEM education. In the following discussion, we explore the specific components of our course that may have contributed to these equitable results, focusing on how the CBL structure and intentional design elements created an environment where all students could meaningfully connect with science, regardless of their background or intended career path. We structure the discussion around the two aims posed earlier, connecting the equitable results with specific course components that may have been especially influential in driving the outcomes. By understanding these mechanisms, we can better inform the design of future STEM education initiatives that prioritize equity and inclusion.

Equitable Development of Science Connection Across Demographic Groups Through a CBL Approach

We first examined the effect of demographic groups on the development of science connection (SciID, Eff, and Val) in our CBL course (Aim 1). Our analysis showed that our *Bioinspired Design* course promoted equitable outcomes in science connection for students across a variety of key demographic groups. This is a notable finding, as *lack* of differences in SciID, Eff, and Val development between historically dominant and nondominant groups are not typically observed and undergraduate STEM courses may even exacerbate differences between demographic groups (Cole & Beck, 2022; Estrada et al., 2019; Shortlidge et al., 2024). In contrast, our *Bioinspired Design* course showed no significant group differences in SciID, Eff, and Val across all the Demographic RM ANOVAs (Tables 2 and 3), suggesting effective and equitable development of science connection.

Our results could be connected to our *Bioinspired Design* course's emphasis on interdisciplinary collaboration and its structured support to promote inclusive teamwork (Full et al., 2021). Students engaged in a series of team-based design projects culminating in a final project where demographically diverse teams extracted a biological principle, created a bioinspired design, and presented their work in a public showcase. During this process, we initially formed interdisciplinary teams of students balanced across major/class year/design experience. Students then engaged in team building activities and collaboration training designed to highlight strategies for inclusive teamwork. Following team formation and training, students completed a scaffolded set of design projects ranging from guided to open-ended. This workflow allowed all students to build skills and confidence over time, while the open-ended final project gave teams autonomy to leverage their unique mix of disciplinary expertise. Throughout this process, we required reflection on team dynamics to continuously improve collaboration, encouraging teams to leverage their diverse strengths. Lastly, the public showcase of final projects was a culminating experience where all students could see themselves and their teammates as legitimate members of the scientific learning community.

These intentional team formation, training, and reflection activities created an inclusive environment where student contributions were valued, potentially mitigating differential outcomes in SciID, Eff, and Val development throughout the semester. For example, the course structure may have promoted SciID as students worked in diverse teams, mirroring the collaborative, interdisciplinary nature of the scientific community, and helping all students see themselves as valid contributors regardless of background. For Eff development, various teaming activities provided all students with strategies to effectively contribute, potentially supporting individual self-efficacy. Instructional strategies like these to promote equitable Eff growth are crucial considering the enhancing effects of Eff development on SciID development based on gender and URM status (Cole & Beck, 2022). For Val alignment, interdisciplinary teamwork embodied the scientific value of considering diverse perspectives to enrich discovery and innovation. This approach may have contributed to equitable Val development, an especially noteworthy outcome considering recently observed differences in Val (decreases) even after intervention based on URM and FirstGen status (Cole & Beck, 2022; Shortlidge et al., 2024). By making scientific work a collaborative, interdisciplinary endeavor, the course validated students' potential to meaningfully contribute, regardless of nondominant identity status or academic path status.

Interactions Between Demographic Variables and Science Connection Development

While our analysis showed a wide array of equitable outcomes, we did observe some notable interactions between specific demographic variables and the development of science connection constructs. These interactions related to Aim 2: "Evaluate the effectiveness of our CBL approach in promoting equitable science connection outcomes across seven key demographic variables." We focus our discussion of this aim on the ANCOVA results (Figures 2-4; Tables 6-8). The majority of the ANCOVA results matched the Demographic RM ANOVA findings supporting equitable outcomes. Five out of seven demographic variables showed no significant differences in post-survey scores after controlling for pre-survey scores. However, two findings emerged that warrant further discussion—statistically significant differences in SciID between STEM/Non-STEM majors and lowerclassmen/upperclassmen (Figure 2; Table 6). The detected differences were specific to only SciID and did not extend to Eff or Val measures, which showed equitable development across all groups. Importantly, this result demonstrates that our methods were indeed capable of detecting differences where they existed, thereby increasing confidence in the equitable outcomes observed across other variables and constructs.

The significant difference in SciID between STEM and Non-STEM majors ($F(1, 522) = 23.331$, $p < .001$, $\eta_p^2 = .043$) (Figure 2) suggests that while both groups experienced growth, STEM majors may have developed a comparatively stronger SciID by the end of the course. This could be due to STEM students' prior exposure to scientific practices and communities, which may have allowed them to more readily integrate the course experiences into their existing SciID. Non-STEM students, starting from a potentially lower baseline, may have shown growth (as evidenced by the RM ANOVA analysis) but not to the same extent as their STEM counterparts. The significant difference in SciID based on class status ($F(1, 523) = 8.547$, $p = .004$, $\eta_p^2 = .016$) (Figure 2) may reflect the cumulative effect of college experiences on students' SciID formation. Upperclassmen, with more exposure to scientific coursework and perhaps even undergraduate research experiences, may have been better positioned to integrate the CBL course experiences into their developing SciID. Lowerclassmen, while still benefiting from the course, may require more time and experiences to fully develop their SciID. These results align with previous research suggesting that SciID development is a complex process influenced by multiple factors including prior experiences and academic stage (Estrada et al., 2011; Hamilton, 2004; Hazari et al., 2013). These results also underscore the importance of considering students' diverse starting points when designing interventions to foster science connection. Future iterations of the course could explore targeted support for Non-STEM majors and lowerclassmen to reduce these gaps in SciID development, while maintaining the equitable growth observed in Eff and Val across all groups.

Science Connection as an Ongoing Process Needed for All Students

Based on the results of this and the previous chapter, we further support the notion that developing a strong connection to science is critical for promoting student engagement, persistence, and success in STEM (Estrada et al., 2018; Estrada et al., 2019; Hernandez et al., 2020). The results of this chapter specifically also indicate that cultivating science connection is a viable outcome in all students, including Non-STEM students (i.e., nonmajors). These students, just like their STEM counterparts, participate in science contexts within and outside of the classroom. We know that nearly all students participate in proximal science contexts such as STEM course communities (e.g., science breadth courses), but these same students will also have to engage in more distal science contexts after graduating, such as participation in science-related legislation, public health decision-making, and consumer decision-making. These contexts represent moments in which high levels of SciID, Eff, and Val can lead to informed participation in the democratic process, pro-science behaviors, and critical evaluation of scientific claims (Estrada et al., 2017; Ballen et al., 2017; Gormally & Heil, 2022). By fostering science connection in all students, both STEM and non-STEM graduates can contribute to a more scientifically literate society and make informed decisions across diverse aspects of their lives.

With our key results in mind, we propose a view of science connection as a continuous process that undergoes ongoing development through participation in different types of scientific communities, especially within courses that utilize CBL. Our results demonstrate that a single-semester breadth course can effectively promote equitable science connection for a diverse range of students, including Non-STEM majors, Non-biology majors, and students at all class levels. This adapts the idea that the study of integration through the TIMSI framework—and by extension, science connection—may require extensive time or is specific to certain demographic groups. Therefore, we propose both integration and science connection be studied not just as a long-term outcome, but as a continually developing process. Students' levels of science connection may increase or decrease due to their evolving personal conceptions of what it means to be a scientist (SciID), their ability to do science (Eff), and their alignment with the values of the scientific community (Val) as they progress through their educational experience. A course itself can serve as a meaningful scientific community, and connection with this proximal community is a valuable outcome that may spark further engagement with science and proscience attitudes. By intentionally designing courses to promote science connection, we can enhance retention for STEM students while also providing STEM-enriched experiences for Non-STEM students.

Limitations and Future Research

While our analyses showed only two instances of statistically significant differences between demographic groups (across roughly 50 tests and two different methodologies), limitations in statistical power may have obscured other disparities. We also express caution against conflating nondominant and dominant identities as being the "same" based on a lack of statistical differences, given the negative impacts of this assumption (Estrada et al., 2018). Future qualitative research should explore how specific course components uniquely impact and leverage diverse cultural strengths (Yosso, 2005). Qualitative investigations of student experiences could provide more granular insights that pinpoint differential impacts of key course elements for certain demographic groups.

Additionally, this study showcased short-term science connection outcomes within the timespan of a single semester. Longitudinal follow-up studies could provide insight into the persistence of these gains and whether the course influences future engagement with science and proscience attitudes for both STEM and Non-STEM students. As students progress through their educational journeys, each additional science or CBL course experience serves as a new opportunity to reinforce and deepen their connection to science. Examining distal outcomes, such as pursuing additional STEM courses or engaging in science-related civic behaviors, could help show the lasting impact of CBL experiences.

The findings from Chapters 1 and 2 highlight a particularly intriguing aspect of our CBLcourse—the significant and equitable development of Eff. Chapter 1 revealed the most substantial gains in Eff among the TIMSI constructs, while Chapter 2 demonstrated that these gains were equitably distributed across diverse student demographic populations. This consistent and inclusive growth in Eff suggests that there may be unique elements within our CBL approach that specifically foster self-efficacy development. To further examine this phenomenon, we recognize the need to further explore the concept of self-efficacy as it relates to our course context. Thus, Chapter 3 introduces a novel construct: Innovation Skills Self-Efficacy. In this next chapter, we show that Innovation Skills Self-Efficacy is a critical construct to develop in all students given the global imperative for innovation in addressing ever-evolving challenges.

Chapter 2 Conclusion

In the combined studies of Chapter 1 and 2, we investigated the impact of our *Bioinspired Design* course on cultivating science connection amongst undergraduate students within a CBL course environment. Using the TIMSI framework, we measured changes in SciID, Eff, and Val across a diverse student population. Our results demonstrated that the course equitably promoted science connection based on a pre/post survey analysis using RM ANOVAs, ANCOVAs, and paired *t*-tests. Considering these results, we also described how our course expands what it means to connect with science through CBL. We showed that self-reported science connection can be achieved rapidly and equitably in the context of a single course for students from diverse disciplinary and demographic backgrounds.

This inclusive view of science connection has important implications for broadening participation in science. By demonstrating the potential for science connection across a wide range of students in a short timeframe, breadth courses (like our *Bioinspired Design* course) offer inclusive, scalable, and adaptable models of change to support reform in undergraduate STEM education. As our world increasingly demands scientific literacy and interdisciplinary collaboration, providing more opportunities for meaningful science connection will be crucial. This then opens further possibilities for designing inclusive breadth courses that promote science connection and invites future research on the short-term and long-term impacts of such interventions. Ultimately, these efforts can then create more inclusive pathways for all students to meaningfully connect with science, building a more STEM-enriched society in this ongoing process.

Chapter 3

Assessing Self-Efficacy Growth in Innovation Skills Using a Developmental Perspective

Abstract

Future economic projections forecast the need for workers to pivot between professions with significantly different skill sets. Thus, educators need to prepare all students for imminent workforce redirections through pedagogy that promotes the development of transferable 21st century skills. In the STEM education landscape, traditional assessments have lacked alignment to skills-based learning outcomes, consequently leading to sparse measurement of students' self-efficacy in skill development. Building on the significant and equitable growth in Science Self-Efficacy observed in Chapters 1 and 2, this chapter introduces a novel construct, *Innovation Skills Self-Efficacy*, to further explore self-efficacy development in our Challenge-Based Learning (CBL) course context. We expand on the Tripartite Integration Model of Social Influence (TIMSI) framework used previously by developing a more targeted measure of self-efficacy aligned with the innovative thinking promoted in our *Bioinspired Design* course.

In this chapter, we present the application of a comprehensive assessment framework that emphasizes a developmental perspective and provides meaningful interpretations of student perceptions relative to intended growth outcomes. We utilized this framework to measure the affective domain of student development, specifically student self-efficacy growth in *Innovation Skills* as they participated in the *Bioinspired Design* course. We hypothesized that students would develop greater self-efficacy in *Innovation Skills* as a result of this course and we tested this through a survey-based methodology analyzed by item response theory. Our results showed approximately one standard deviation of growth between the pre and post samples with an especially large effect size in the context of educational interventions. These findings not only corroborate the growth in Science Self-Efficacy observed in previous chapters but also demonstrate the potential of CBL approaches to foster more specific, innovation-oriented self-efficacy. Overall, we show how the use of a comprehensive assessment framework can empower educators in measuring complex 21st century skills self-efficacy, particularly in CBL environments. This framework can be utilized across subject and course contexts to develop psychometrically robust assessments of skills-based constructs essential for advancing undergraduate STEM education.

Introduction

Chapters 1 and 2 of this dissertation demonstrated the effectiveness of our Challenge-Based Learning (CBL) *Bioinspired Design* (Full et al., 2021) course in fostering science connection through an adaptation of the Tripartite Integration Model of Social Influence (TIMSI) framework. Notably, we observed the highest levels of equitable growth in the Science Self-Efficacy (Eff) social influence construct. These findings suggested that our CBL approach may be especially consequential for promoting students' confidence in their scientific abilities. To further investigate this outcome, we recognized the need to develop a more targeted measure of Eff specifically aligned with the innovative thinking promoted in our course's unique interdisciplinary learning context. In this chapter, we introduce *Innovation Skills Self-Efficacy* as a novel construct that builds upon the Eff measure. This new construct shares significant overlap with Eff, as both focus on students' confidence in their scientific abilities. However, Innovation Skills Self-Efficacy provides a more granular view by specifically targeting students' belief in their capacity to innovate within scientific contexts.

This progression from the general Eff measure to the more specific Innovation Skills Self-Efficacy allows us to more precisely capture the unique outcomes of our CBL course in preparing students for future STEM challenges. It enables us to build upon the TIMSI framework while providing a more nuanced understanding of self-efficacy development in the context of innovation-focused, interdisciplinary STEM education. Through this approach, we aim to more effectively assess and understand the specific ways in which our CBL course is preparing students to confidently tackle complex, real-world challenges that require innovative thinking. By developing and validating a construct focused on self-efficacy in Innovation Skills, we highlight a critical aspect of student learning that is particularly relevant to the goals of our CBL course and the broader needs of 21st century STEM education.

Background

Preparing Students for the Known Jobs of Today and the Unknown Jobs of Tomorrow: The Need for Innovation Skills

We live in a society where educators inevitably face the task of preparing students for future careers that do not yet exist. In a 2022 working paper by the National Bureau of Economic Research, Autor et al. concluded that 74% of employment amongst professionals in 2018 was found in job titles that did not exist in 1940. Thus, educators at all levels of teaching are faced with a pressing question—*how do we prepare the students of today for the unknown jobs of tomorrow?* As science educators particularly, we must prepare *all* students, STEM and non-STEM, for future workforce pivoting through pedagogy that promotes the development of transferable STEM skills (National Academies of Sciences, Engineering, and Medicine [NASEM], 2021). In addressing this challenge, CBL approaches, as explored in Chapters 1 and 2, offer a promising pedagogical framework. By engaging learners in real-world, interdisciplinary challenges, CBL cultivates adaptability crucial for navigating the ever-evolving job landscape of the future. In the previous chapters, we saw how CBL promoted greater science connection among students, and in turn, fostered a learning environment where students developed more confidence in an array of science-related skills (Eff).

In this chapter, we specifically focus on *innovation* as a key skill needed for the future. The ability to innovate has consistently been ranked as the most important skill for the future workforce (World Economic Forum, 2020). Considering the current and future need for innovation, we base this chapter on the crucial role of *Innovation Skills* in Undergraduate STEM Education (USE)[2]. Various national reports on USE have called for fostering skills related to innovation. In the National Academies *Imagining the Future of Undergraduate STEM Education* report (2022), a series of forward-looking questions are asked about the world in 2040 including "How will they [scientists and engineers] learn the skills, practices, and concepts they need to contribute to innovation?" (NASEM, 2022, p. 1). Other reports describe innovation as "key to economic success" (NASEM, 2018, p. 50), the ability to innovate as a "major differentiator" (National Endowment for the Arts, 2017, p. 12), and call for "giving students broader repertoires for critical thinking and creative innovation" to solve the complex problems of today and tomorrow with creative solutions that are humane, technically robust, and elegant (NASEM, 2018, p. 54). These reports cohesively establish the importance of Innovation Skills for the future workforce, the scientific enterprise, and the long-term goals of USE.

How do we define *innovation*? Innovation is a multifaceted concept that has been differentially defined across various disciplines and contexts. A comprehensive analysis by Singh and Aggarwal (2022) synthesized over 200 definitions to propose that innovation is "the operationalization of creative potential with a commercial and/or social motive by implementing new adaptive solutions that create value, harness new technology or invention, contribute to competitive advantage and economic growth" (p. 177). This definition highlights the social and technological aspects of innovation, which are particularly relevant to the educational research context presented in this chapter. The Nature Index (2022) offers an additional definition that also resonates with our pedagogical focus of translating scientific research into real-world applications, describing innovation as "the practice of turning cutting edge basic research into inventions with real world application." These definitions collectively emphasize the importance of viewing innovation as a process that can lead to research-based novelty and societally impactful practical applications. Given the influential role of innovation in addressing complex global challenges and driving economic growth, we now shift to a consideration of what specific elements of Innovation Skills need to be fostered in USE.

The Need to Foster Constructs of Learning in the Affective Domain through Self-Efficacy Development

In educational assessment, a construct is defined as a theoretical attribute of interest, typically elicited and measured through instruments such as tests or surveys. Innovation Skills represent a construct within both the *cognitive* and *affective* domains of learning (Bloom et al., 1956). In the cognitive domain, Innovation Skills are the capabilities needed to innovate, such as creativity, critical thinking, and problem-solving. In contrast, our research aims to establish Innovation Skills as a construct within the affective domain. Krathwohl et al. (1964) initially defined this domain to encompass objectives that "emphasize a feeling tone, an emotion, or a degree of acceptance or rejection." Birbeck and Andre (2009) further explained that the affective domain is the gateway to learning, influencing students' motivation and willingness to engage in

[2] STEM represents science, technology, engineering, and mathematics. However, we also recognize other commonly used acronyms to aggregate disciplines such as STEAM (arts), STEMM (medicine), and S&E (science and engineering). Additionally, our definition of USE matches the NASEM definition (2018).

their coursework. Additionally, understanding student affect can even help adjust teaching to improve student learning and reduce STEM attrition (McConnell & van der Hoeven Kraft, 2011; NASEM, 2015). Taken together, this literature supports the need to develop and assess affective constructs made up of the emotional and attitudinal components that underpin intended behaviors.

The specific affective component of interest for our research was the development of *self-efficacy* in Innovation Skills. Self-efficacy, as defined by Bandura (1997), refers to "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments." The importance of developing self-efficacy is widely recognized throughout education research, as it plays a crucial role in students' persistence and performance. According to social cognitive theory, students with higher self-efficacy are more likely to persevere through challenges and apply their skills in novel situations, making it a critical factor in educational outcomes (Bandura, 1997; Zimmerman, 2000). When applied to our specific context of STEM education, self-efficacy represents a student's confidence in their ability to succeed in a field of science (Koballa & Glynn, 2013) alongside their belief in their ability to successfully perform scientific tasks and achieve scientific goals (Estrada et al., 2011). Building on this understanding, Chemers et al. (2011) noted that confidence in one's ability to perform specific behaviors or accomplish specific tasks is predictive of performance beyond what can be predicted by objective measures of ability alone. This suggests that self-efficacy plays a crucial role in student success in combination with their actual skill level. Furthermore, Estrada et al. (2011) concluded that self-efficacy consistently predicts students' interest, goals, and persistence in pursuing careers in STEM disciplines. This underscores the importance of fostering self-efficacy in STEM education, as it not only affects immediate academic performance but also influences long-term career choices and persistence in STEM fields. Further research by Ballen et al. (2017) found that increases in science self-efficacy mediated positive effects on student performance in introductory biology courses. Importantly, this effect was observed across many different student demographic groups, highlighting the universal importance of self-efficacy in USE.

By focusing on self-efficacy in Innovation Skills, we address a critical affective construct that underpins students' capacity to innovate. Strong self-efficacy in Innovation Skills is foundationally needed for students to engage effectively in actual innovation processes. As discussed by Birbeck and Andre (2009), affective attributes enable the transference of cognitive skills between contexts, such as between university and the workplace. This suggests that developing self-efficacy enhances students' confidence in their Innovation Skills while also improving their ability to apply those skills in real-world settings requiring the cognitive domain.

Despite the importance of assessing the affective domain, measuring aspects of affect (e.g., students' attitudes, beliefs, and expectations) is less prevalent in science education research compared to measuring student learning and cognition (Koballa & Glynn, 2013; Maric et al., 2023). A key hallmark of measuring student affect is the use of self-reported measures related to self-efficacy and confidence, which are known to influence motivation, persistence, and achievement in STEM (NASEM, 2012; Trujillo & Tanner, 2014). However, there remains a need for more comprehensive assessment approaches that address both cognitive and affective domains, as called for in recent reports on science education (NASEM, 2021). Our focus on self-efficacy in Innovation Skills bridges the gap between affective development and cognitive performance in innovation assessment. We recognize that while cognitive skills are necessary for innovation, they are insufficient without the accompanying belief in one's ability to apply those skills effectively. By fostering and measuring self-efficacy in Innovation Skills, we aim to empower students with the confidence and motivation necessary to fully utilize their cognitive capabilities in tasks requiring innovation.

Defining the Construct: Measuring Self-Efficacy in Innovation Skills, a Critical 21st Century Skill Needed for the Future

Here, we introduce a novel *Innovation Skills* self-efficacy construct derived from assessment in the affective domain. We initially developed this construct by linking the CBL activities students complete in the course with broader skills frameworks used throughout K-16 that emphasize *21st Century Skills* (21CS). There have been widespread efforts to foster "deep learning" through the implementation of curricula that develops 21CS such as innovation, creativity and creative problem-solving (Bellanca & Brandt, 2010; Griffin & Care, 2015; NASEM, 2012). These 21CS skills are characterized using "transferable knowledge that can be applied to solve new problems or respond effectively to new situations," which resonated well with the CBL activities embedded throughout the course (Pellegrino, 2017, p. 228). For a comprehensive review and comparison of major 21CS frameworks, see Chu et al. (2017). Though there are differences between each of these 21CS frameworks, most frameworks are broadly consistent with each other (Voogt & Roblin, 2012) and agree on the need to develop Innovation Skills. However, the inherently multifaceted nature of these skills makes their measurement complex (Geisinger, 2017) and most scholars agree that current assessment instruments do not adequately measure 21CS due to insufficient reporting of reliability evidence and validity arguments (Siddiq et al., 2017).

The development of Innovation Skills as a specific class of 21CS has also been studied. In the area of Innovation Skills assessment, current instruments recognize the importance of measuring innovation competencies, especially in the context of business and economic advancement (Luke, 2013; Chirazi et al., 2019). The development of previous self-report surveys supports the need to measure the affective domain related to Innovation Skills, particularly students' self-efficacy beliefs (Carberry et al., 2018; Gerber et al., 2012b; Nelson et al., 2009). As a result, innovation self-efficacy, defined as an individual's belief in his or her ability to accomplish tasks necessary for innovating (Gerber et al., 2012a; Gerber et al., 2012b), has emerged as a construct of focus in the context of engineering education. In our study, we present a more inclusive context that enhances previous assessments of Innovation Skills self-efficacy by including students from diverse backgrounds and disciplines to better reflect the interdisciplinary nature of innovation in real-world contexts. While Innovation Skills have been highlighted as increasingly essential in business, economics, and design fields (Durand et al., 2015; Keinänen et al., 2018; Luke, 2013), the development of assessments to measure self-efficacy in Innovation Skills in CBL contexts that encompass a wide range of disciplines and student backgrounds remains sparse. Additionally, there is significant potential to advance prior survey instruments by situating them within a comprehensive assessment framework that utilizes more robust psychometric techniques. Previous efforts have predominantly used factor analysis methods (Gerber et al., 2012b; Keinänen et al., 2018; Carberry et al., 2018) which represent an important first step that can be enhanced with modern test theory perspectives like item response modeling.

Defining the Context: Measuring Self-Efficacy in Innovation Skills within a Bioinspired Design Course

Considering the national, scientific, and educational importance of innovation, how can we empower undergraduate students to think and act innovatively *now*, in the classroom, so that they can become the innovators we need in the future? To address this question, we developed an undergraduate course centered on *bioinspired design* (also referred to as *biomimicry* or *bionics*). Tasking students to engage in a self-efficacy construct based on Innovation Skills requires a CBL course context that fosters the application of innovative thinking. Our 180-student *Bioinspired Design* course—open to all majors, all years, with no prerequisites—represents an especially appropriate context because of the inherent nature of the discipline (Full et al., 2021). Bioinspired design is a process that can be viewed as a problem solving approach, an innovation paradigm for creative thinking, and a methodology for knowledge transfer between disciplines (Rovalo et al., 2020; Wanieck et al., 2020). The interdisciplinary nature of bioinspired design, which combines biology and technology to solve practical problems, makes it a valuable context for fostering innovative thinking and problem-solving skills in students (Jacobs et al., 2022). A recent 2023 “Dear Colleague Letter” by the National Science Foundation even regarded bioinspired design as a powerful means of addressing the need for the United States to do more to ensure that discoveries are translated into innovations (as outlined in the National Science Board’s [2020] Vision 2030 report).

From a learning perspective, students’ confidence in their abilities can influence their persistence to engage in the bioinspired design process as they attempt to solve problems within CBL. Thus, high levels of self-efficacy in Innovation Skills may be crucial to effectively translate biological principles into novel solutions. Considering this context, there is a lack of a comprehensive, integrated system for assessing, interpreting, and monitoring student performance at the intersections of CBL, 21CS acquisition, and Innovation Skills self-efficacy development. This study directly addresses several needs for assessment in each of those spaces to ultimately help answer Songer and Ruiz-Primo’s (2012) question: “How can we develop assessments in science that tap adaptability, complex communication and social skills, non-routine problem solving, self-management, and system thinking?”

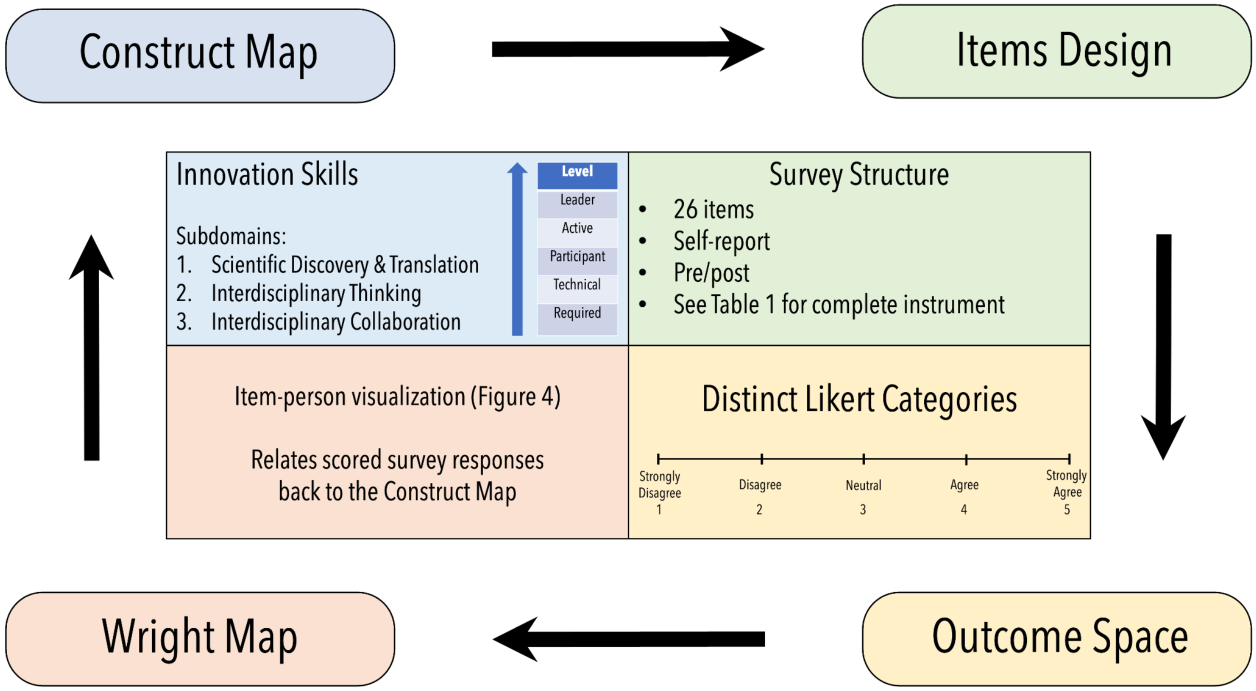
This question is especially relevant to the advancement of assessment that measures learning outcomes in the bioinspired design education space. For example, Wanieck et al. (2020) connected 18 learning objectives of bioinspired design with Bloom’s taxonomy, but these connections were not followed by assessments to measure reliability and validity, making it difficult to support any claims regarding the successful completion of the learning objectives. Other key aspects of bioinspired design education include discussions by Gvili et al. (2016) and Fried et al. (2020) regarding the importance of teaching cross-cutting concepts like structure-function relationships. Additionally, Nagel and Pidaparti (2016) summarized various engineering competencies of the 21st century engineer and how these relate to bioinspired design education. In each of these educational overviews, the inclusion of self-efficacy assessment evidence explicitly connected to the intended learning outcomes would have improved the overall understanding of student growth within bioinspired design education.

Assessment Framework: The Berkeley Evaluation & Assessment Research System (BAS) and the Four Building Blocks

Considering the need to align assessment with evidence-based educational practices, one method is to develop new instruments that measure purported changes in students as they progress through a course, curriculum, or program (NASEM, 2018). How do we develop these instruments to address the call for more holistic, practices-aligned affective assessment in USE? Here, we present an initial step toward more comprehensive assessment based on practices aligned with robust psychometric standards (American Educational Research Association [AERA] et al., 2014). We build on previous efforts to implement assessment frameworks in USE, such as the three-dimensional learning assessment protocol by Laverty et al. (2016), which was developed based on an approach to assessment as an evidentiary argument, or Evidence-Centered Design (Mislevy et al., 2003). Through this study, we advance this approach by utilizing a comprehensive assessment framework based not only on (1) evidentiary argument, but also (2) a developmental perspective, (3) a match between instruction and assessment, and (4) management by instructors to promote iteration (Wilson & Scalise, 2006). All four of these principles result in the Berkeley Evaluation & Assessment Research Assessment System (BAS), an integrated approach to developing assessments that provide meaningful interpretations of student work relative to the cognitive and developmental goals of a curriculum (Wilson & Sloane, 2000). Assessment centered on developmental progressions provides critical insights into differentiated learner growth over time, enabling responsive instruction tailored to varied skill levels. This comprehensive approach moves beyond static assessments to gain robust understanding of student advancement along multiple dimensions, empowering educators to promote equitable outcomes. For example, this assessment framework enhances the pre/post assessment approaches used in Chapters 1 and 2 by tracking student growth through tiered levels of development. While the previous chapters assessed overall measures of growth in science connection, the developmental perspective within this assessment framework better accommodates the multiple levels of development inherent to Innovation Skills Self-Efficacy (see Building Block 1: Construct Map; Figure 1).

In the present research, we specifically show how a developmental perspective to assessment, realized through the Four Building Blocks of assessment (Wilson, 2023), allows educators to gain novel insights on student growth—insights that would otherwise not be possible to see without the application of this comprehensive assessment framework. We explain how each of the Four Building Blocks—1) Construct Map, 2) Items Design, 3) Outcome Space, and 4) Wright Map—connect to our assessment of students in the subsequent sections. Figure 1 summarizes this approach.

Figure 1
Utilizing the Four Building Blocks in the BEAR Assessment System to Measure the Innovation Skills Construct



Note. A) The Construct Map for the Innovation Skills construct contains three subdomains—Scientific Discovery & Translation Process, Interdisciplinary Thinking, and Interdisciplinary Collaboration. Each of these subdomains is hypothesized to develop across five qualitatively distinct, ascending ability levels (Required, Technical, Participant, Active, Leader); B) The Items Design procedures resulted in the development of a 26 item pre/post Likert-type survey to measure the Innovation Skills construct. Each of the 26 items is mapped onto a level within a Construct Map as shown in Table 1; C) The Outcome Space translates survey responses into scores that map directly back to levels of the Construct Map. In this study, the Outcome Space includes the distinct Likert categories ranging from "Strongly Disagree" to "Strongly Agree." The items mapped onto the highest levels ("Leader") are the most difficult to agree with on the Likert scale whereas the items mapped onto the lowest levels ("Required") are the easiest to agree with on the Likert scale; D) The Wright Map involves relating the scored survey responses back to the Construct Map by translating the scores into "locations" on the construct's continuum (Wilson, 2023). This process is visualized as item-person Wright maps. These Wright maps represent a visualization of the construct based on student data and responses. See Figure 3 for more details.

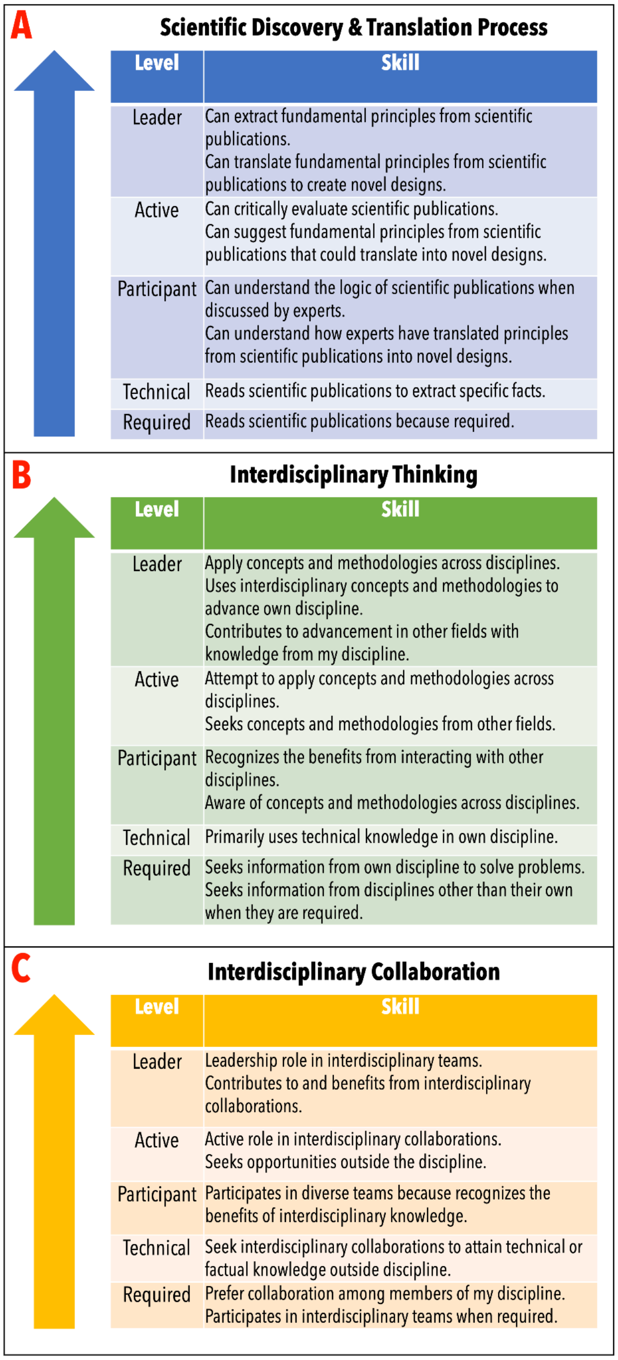
Building Block 1: Construct Map. The first building block, the Construct Map, is a construct definition tool that relies on a developmental perspective to assess student achievement and growth (Wilson, 2023). In a Construct Map, this developmental perspective manifests through qualitatively different levels of performance on the latent construct. A latent construct represents a theoretical attribute, and, in our assessment, we considered Innovation Skills as an overall latent construct, or variable of measurement, consisting of three subdomains—Scientific Discovery & Translation Process (SDTP), Interdisciplinary Thinking (IT), and Interdisciplinary Collaboration (IC). These subdomains were adapted from a previously validated Construct Map for interdisciplinary research skills developed in the context of a learning laboratory course on organismal biomechanics (Full et al., 2015).

Each of the subdomains (SDTP, IT, and IC) were aligned to a developmental perspective ranging from "Required" (lowest level) to "Leader" (highest level), resulting in three Construct Maps hypothesized to encompass the development of student self-efficacy in Innovation Skills

(Figure 2). We considered how self-efficacy might manifest differently at each level. For instance, at the “Required” level, students might have confidence in following course requirements, while at the “Leader” level, they confidently endorse taking on leadership roles in tasks requiring innovative thinking. We encourage readers to first observe the contrast between the lowest and highest level of each subdomain and then notice the stepwise development in the various intermediate levels. As part of the variable clarification process necessary for this first building block (Wilson, 2023), we explain the purpose of each subdomain below.

Figure 2

Construct Maps for A) Scientific Discovery & Translation Process (SDTP), B) Interdisciplinary Thinking (IT), and C) Interdisciplinary Collaboration (IC). (Adapted from S17 in Full et al., 2021)



A **Scientific Discovery & Translation Process**

Level	Skill
Leader	Can extract fundamental principles from scientific publications. Can translate fundamental principles from scientific publications to create novel designs.
Active	Can critically evaluate scientific publications. Can suggest fundamental principles from scientific publications that could translate into novel designs.
Participant	Can understand the logic of scientific publications when discussed by experts. Can understand how experts have translated principles from scientific publications into novel designs.
Technical	Reads scientific publications to extract specific facts.
Required	Reads scientific publications because required.

B **Interdisciplinary Thinking**

Level	Skill
Leader	Apply concepts and methodologies across disciplines. Uses interdisciplinary concepts and methodologies to advance own discipline. Contributes to advancement in other fields with knowledge from my discipline.
Active	Attempt to apply concepts and methodologies across disciplines. Seeks concepts and methodologies from other fields.
Participant	Recognizes the benefits from interacting with other disciplines. Aware of concepts and methodologies across disciplines.
Technical	Primarily uses technical knowledge in own discipline.
Required	Seeks information from own discipline to solve problems. Seeks information from disciplines other than their own when they are required.

C **Interdisciplinary Collaboration**

Level	Skill
Leader	Leadership role in interdisciplinary teams. Contributes to and benefits from interdisciplinary collaborations.
Active	Active role in interdisciplinary collaborations. Seeks opportunities outside the discipline.
Participant	Participates in diverse teams because recognizes the benefits of interdisciplinary knowledge.
Technical	Seek interdisciplinary collaborations to attain technical or factual knowledge outside discipline.
Required	Prefer collaboration among members of my discipline. Participates in interdisciplinary teams when required.

Note. Upward arrow represents an ascending continuum of ability levels. Levels in the Construct Map range from low (Required) to high (Leader) and correspond to perceived self-efficacy in the subdomain.

Scientific Discovery & Translation Process (SDTP). The bioinspired innovations of the future will require students to develop skills related to both scientific discovery and translation. Regarding scientific discovery, we recognize that "the work of science is complex: it is a process, a product, and an institution. As a result, engaging in science—whether using knowledge or creating it—necessitates some level of familiarity with the enterprise and practice of science" (NASEM, 2016, p. 11). Our course aims to illuminate the black box of the scientific process that leads to original discoveries and furthermore, the process that leads to the successful translation of those discoveries (Full et al., 2021). To this end, by situating student learning through the researcher's frame (Lave & Wenger, 1991), we task students to use a solution-driven approach to bioinspired design (Helms et al., 2009). In this approach, the design process begins with finding inspiration based on an original scientific discovery from primary literature. Students then use our *Discovery Decomposition* tool (Figure S1) to "decompose" their selected publication into its core components and extract the fundamental principles behind the scientific discovery. After breaking down the publication and extracting the fundamental principles, students then engage in the translation process. Students use our *Analogy Check* tool (Figure S2) to create an analogy to solve a human problem by actively comparing the fundamental principles from the publication with a proposed design problem. Students compare various elements between the design solution from their publication and the design problem they want to solve such as function, structure, operating environment, size, mechanisms, specification, performance criteria, and constraints. By utilizing both the Discovery Decomposition and Analogy Check tools multiple times throughout the semester, we hypothesize that students develop self-efficacy in the skills related to the SDTP subdomain (Figure 2A) of the Innovation Skills construct.

Interdisciplinary Thinking (IT) and Interdisciplinary Collaboration (IC). In addition to SDTP, Innovation Skills require both IT (Figure 2B) and IC (Figure 2C). The American Association for the Advancement of Science *Vision and Change* (2011) report classified "the ability to tap into the interdisciplinary nature of science" and "the ability to communicate and collaborate with other disciplines" (p. 15) as core competencies within any undergraduate biology course. Why does interdisciplinarity matter, particularly in the context of innovation? We view IT and IC as a means of achieving *scientific convergence*, an integrative approach spanning across multiple disciplines that "stimulates innovation from basic science discovery to translational application" to tackle scientific and societal challenges (NRC, 2014, p. 1). Additionally, skills related to IT and IC promote *integrative learning*, or the "ability to connect, apply, and/or synthesize information coherently from disparate contexts and perspectives, and make use of these new insights in multiple contexts" (Barber, 2012, p. 593). This type of learning aligns with the aforementioned skills-based frameworks (e.g., 21CS). Most importantly, regarding the Innovation Skills construct, integrative learning is necessary because "innovation today requires integrative thinking and collaboration, while technology development decreases the need to perform repetitive tasks and leaves more time for innovation and interdisciplinary collaboration among colleagues and occupations" (Carnevale et al., 2011; NASEM, 2018, p. 51). Overall, fostering self-efficacy in IT and IC skills leads to the integration of knowledge from multiple disciplines (i.e., across over 40 different majors), thereby promoting the innovative thinking needed to develop significant scientific breakthroughs (i.e., societally impactful bioinspired designs) (NASEM, 2018).

The Construct Maps for IT and IC follow the same ascending continuum as the SDTP Construct Map. These constructs differentiate the interdisciplinary skills needed for "thinking" and those needed for "collaboration." IT skills refer to the knowledge, concepts, and methodologies a student uses when engaging in interdisciplinary work while IC skills refer to the varying levels of student participation in interdisciplinary teams. Both are necessary, alongside SDTP, to participate in the inherently interdisciplinary process of bioinspired design.

Overall, this developmental perspective embedded within the Construct Map represents a hypothesis of growth in students' Innovation Skills self-efficacy. As students move through the levels of our Construct Map, we expect to see corresponding increases in their self-efficacy beliefs. Importantly, we must rigorously test this developmental hypothesis to verify our assumptions related to the levels within the Innovation Skills construct. This testing can be done by analyzing item responses from students mapped to locations within our hypothesized construct. Thus, the next step in the assessment cycle is to develop items that offer us a way to observe the variables underlying the Innovation Skills construct. This leads us to the second building block, Items Design (Figure 1).

Building Block 2: Items Design. The second building block, the Items Design, focuses on the match between instruction and the types of assessment, often determined by the match between assessment tasks and levels within the Construct Maps. The purpose of this building block is to develop a causal link (the transduction) between the construct underlying the Construct Map (Innovation Skills) and the responses to items within an instrument. This causal link requires a conscious and purposeful design of items because items represent a way to reveal a particular construct within a respondent (i.e., student in the course). As stated by Wilson (2005), "the measurer needs to build a structure that links the construct closely to the items—one that brings the inferences as close as possible to the observations" (p. 8). Establishing this linking structure embodies the "design" feature of this building block, which includes designing (1) a procedure that allows observations to be made under a set of standard conditions that span the intended range of the item contexts, and (2) a procedure for classifying those observations into a set of standard categories (Bhatti et al., 2023; Wilson, 2023).

In developing our items, we also drew upon Bandura's (2006) guidelines for constructing self-efficacy scales. We ensured content validity by crafting items that accurately reflect the specific Innovation Skills targeted in our construct. Following the principle of domain specificity, our items were tailored to Innovation Skills within the context of interdisciplinary STEM education. Furthermore, we incorporated gradations of challenge in our items, aligning with our Construct Map's levels from "Required" to "Leader," to capture the full range of task demands within each subdomain of Innovation Skills. The results of these design procedures are discussed in the Methods (See Survey Instrument and Administration). We further explain the item mapping procedure in the next building block, the Outcome Space.

Building Block 3: Outcome Space. The third building block, the Outcome Space, refers to a procedure for classifying or mapping responses to survey items. Recognizing the cyclical nature of the assessment cycle using Four Building Blocks, the Outcome Space represents the first step in the inference process because it begins to tie together the hypothesis (the Construct Map), observation (the responses), and the measurement (the scores or ordered categories) (Bhatti et al., 2023; Wilson, 2023). In other words, through this building block, we can translate survey responses into quantitative data or scores that map directly back to levels of the Construct Map. Our Innovation Skills survey is a Likert scale instrument in which the Outcome Space includes the distinct Likert categories ranging from "Strongly Disagree" to "Strongly Agree." The items mapped onto the highest levels ("Leader") are the most difficult to agree with on the Likert scale whereas the items mapped onto the lowest levels ("Required") are the easiest to agree with on the Likert scale. Seven items map to the "Leader" level, six items map to "Active," five items map to "Participant," three items map to "Technical," and five items map to "Required." Together, the mapped Likert items form the basis of the Outcome Space for our

Innovation Skills instrument. This leads us to the fourth and final building block of our assessment framework, the Wright Map (Figure 1).

Building Block 4: Wright Map. The final building block, the Wright Map (Measurement Model) involves relating the scored survey responses back to the Construct Map by translating the scores into "locations" on the construct's continuum (Wilson, 2023). This process is typically visualized as item-person Wright maps[3]. These Wright maps represent a visualization of the construct based on student data and responses. We further elaborate on the details of this building block and its connection to our survey response data in the Results and Discussion.

Research Question and Hypothesis

Building on the findings of Chapters 1 and 2 (significant growth in Eff across diverse student groups), our research question for this chapter centered on developing and assessing a specific form of self-efficacy relevant to the CBL environment of our *Bioinspired Design* course. Based on the previously described assessment framework and the motivations outlined in our definitions of the construct and context, our guiding research question was: *How does the Bioinspired Design course affect students' self-efficacy in Innovation Skills?* We hypothesized that, through engagement with learning activities in the *Bioinspired Design* course, students' self-reported measures of self-efficacy in Innovation Skills would *increase* across each subdomain of the construct. We assessed this hypothesis through both a general Likert scale analysis and an anchored pre/post analysis using Rasch modeling. We predicted support for our hypothesis from both analyses, with the general Likert scale analysis resulting in an initial means of data-based support followed by more psychometrically robust support from the pre/post Rasch analysis.

Methods

Study Design: Utilizing BAS and the Four Building Blocks to Conduct a Pre/Post Survey

In this study, we implemented the BAS framework to design and deliver a pre/post survey instrument that asked students to self-report their beliefs related to their self-efficacy in Innovation Skills. This pre/post design allowed us to analyze potential development of students' self-efficacy in Innovation Skills over time. This design was similar to what was done in Chapters 1 and 2 but extends the methodology by incorporating the Four Building Blocks. We used the Four Building Blocks as an interconnected and iterative cycle of assessment to ground our analysis of the Innovation Skills construct (See Figure 1).

Survey Instrument and Administration

This survey instrument originated in the context of a learning laboratory course on organismal biomechanics where we developed and validated an interdisciplinary research skills construct using the BAS framework (Full et al., 2015). The original survey consisted of 32 Likert-type items that measured students' self-efficacy in interdisciplinary research skills.

[3] For more information, see "Some Notes on the Term: 'Wright Map.'" by Mark Wilson at https://www.rasch.org/rmt/rmt253b.htm

Utilizing the iterative nature of our assessment framework, we adapted this survey into 26 Likert-type items specifically targeting self-efficacy in Innovation Skills. This resulted in a new survey instrument with 26 items mapped onto a specific subdomain and level, as shown in Table 1. We administered this pre/post Innovation Skills survey to students enrolled in *Bioinspired Design* before and after the course. Students voluntarily completed the survey for one point of extra credit. This study was IRB reviewed and approved (Protocol ID: 2017-12-10602).

Table 1
Innovation Skills Instrument with Subdomains and Levels

Item	Subdomain	Level
1. I read scientific publications to extract specific facts.	SDTP	Technical
4. I read scientific publications when they are required.	SDTP	Required
6. I can understand the logic of scientific publications when discussed by experts.	SDTP	Participant
8. I can translate fundamental principles from scientific publications to create novel designs.	SDTP	Leader
13. I can understand how experts have translated principles from scientific publications into novel designs.	SDTP	Participant
19. I can extract fundamental principles from scientific publications.	SDTP	Leader
22. I can suggest fundamental principles from scientific publications that could translate into novel designs.	SDTP	Active
25. I can critically evaluate scientific publications.	SDTP	Active
2. I use interdisciplinary concepts and methodologies to advance my own discipline.	IT	Leader
9. I am aware of concepts and methodologies across disciplines.	IT	Participant
11. I seek concepts and methodologies from other fields.	IT	Active
12. I recognize the benefits from interacting with other disciplines.	IT	Participant
15. I seek information from other disciplines when required.	IT	Required
16. I can apply concepts and methodologies across disciplines.	IT	Leader
17. I seek information from my own discipline to solve problems.	IT	Required
20. I attempt to apply concepts and methodologies across disciplines.	IT	Active
21. I primarily use technical knowledge in my own discipline.	IT	Technical
24. I contribute to advancements in other fields with knowledge from my discipline.	IT	Leader
3. I take a leadership role in interdisciplinary teams.	IC	Leader
5. I take an active role in interdisciplinary collaborations.	IC	Active
7. I seek interdisciplinary collaborations to attain technical or factual knowledge outside my discipline.	IC	Technical
10. I prefer collaboration among members of my discipline.	IC	Required

14. I contribute to and benefit from interdisciplinary collaborations.	IC	Leader
18. I seek opportunities outside my discipline.	IC	Active
23. I participate in interdisciplinary teams when required.	IC	Required
26. I participate in diverse teams because I recognize the benefits of interdisciplinary knowledge.	IC	Participant

Items grouped based on three subdomains of Scientific Discovery & Translation Process (SDTP), Interdisciplinary Thinking (IT), and Interdisciplinary Collaboration (IC). Each item level is also noted (Required to Leader). (Adapted from S16 in Full et al., 2021)

Demographics and Sample Characteristics

The demographics of our survey population was a broadly representative sample of undergraduate students at the university because of the inclusive nature of the course (i.e., open to all majors, all years, no prerequisites, satisfies a breadth requirement). Table 2 contains a demographic breakdown of our survey population and shows notable trends such as a diverse distribution of class standing, ethnicity, and major. In terms of class standing, survey respondents came from a well spread population of freshmen through seniors (Year 1: ~33%, Year 2: ~22%, Year 3: ~25%, Year 4+: ~20%). Regarding ethnicity, the majority of the respondents were non-White (~76%). Additionally, a large portion of the student population at this university would be categorized into what is typically labeled as “Asian,” but in our efforts to recognize the expansive within-group diversity of this category (Bhatti, 2021; Nguyen et al., 2017; Vue et al., 2023), the Asian ethnicity was disaggregated into various sub-populations (Supplement 4). With regard to major, the survey respondents came from over forty different majors throughout the spectrum of STEM (~68%) and non-STEM (~32%). Regarding gender, the majority of survey respondents identified as female (~59%), breaking the traditional trend of STEM courses—particularly those with engineering and design components—being taken mostly by students who identify as male. Lastly, the majority of survey respondents indicated that their current/intended major was not biology (~74%).

Table 2
Demographics and Sample Characteristics of Survey Population

Demographic	Category	Percentage (%)
Class standing	Year 1	32.67
	Year 2	22
	Year 3	25
	Year 4+	20.33
Ethnicity*	White	23.87
	Non-white	76.13
STEM major	STEM	68.07
	Non-STEM	31.93
Gender	Female	58.73
	Male	40.17
	Non-binary/Other	1.10
Biology major	Yes	26.33
	No	73.67

*See Supplement 4 for disaggregated distributions

Data Collection and Analysis: Likert and Rasch Approaches

We collected course survey data each Spring semester from 2016 to 2020 resulting in 514 pre- and 432 post-survey responses. Our pre/post survey instrument remained consistent throughout that time. Item responses were coded 1-5, with 1 representing "Strongly Disagree" and 5 representing "Strongly Agree." We compiled two separate datasets that represented the pre (N = 514) and post-survey (N = 432) responses. Each dataset contained pre/post responses to all items. We obtained matched student data across some years (2016-2018), but attrition in the later years (2019-2020) led to some unmatched pre/post survey responses as shown in Table 3. We opted to include all pre/post responses to ensure a large sample size for analysis. See S4 for further details on the difference in pre and post sample sizes. We considered all the pre responses as one large collection of data to compare with the post responses, a separate collection of data. Therefore, any changes from pre to post served as a proxy for changes in students' self-efficacy in Innovation Skills resulting from completing the *Bioinspired Design* course. In other words, we considered the *Bioinspired Design* course to represent an "intervention" through which we could compare students' self-efficacy in Innovation Skills before and after the intervention (i.e., before and after the course).

Table 3
Pre- and Post-Survey Responses

Year	Pre	Post
2016	38	38
2017	85	85
2018	32	32
2019	173	142
2020	186	135
Total	**514**	**432**

We conducted two types of analyses on the pre- and post-survey responses—a general Likert scale analysis and a follow-up Rasch analysis using ACER ConQuest (Adams et al., 2020). For the general Likert scale analysis, we calculated the mean Likert rating for each of the 26 items in the pre and post-survey datasets. We then calculated a delta value to measure the difference between the mean ratings in the pre- versus post-survey responses. We also calculated the Standard Deviation (SD) of each of the pre/post delta values across the 26 items. We conducted item level paired t-tests for a subset of the sample that included matched pairs (Supplement 5-Table S8). After conducting this general Likert scale analysis, we conducted a follow-up Rasch analysis to address the limitations of analyzing raw scores on self-reported Likert instruments (Chimi & Russell, 2009). We explain the added value of this follow-up Rasch analysis in the section below.

Advancing Assessment: Utilizing Item Response Theory and Rasch Analysis to Improve Likert Survey Research

The use of Likert-type survey instruments often results in the collection of qualitative ordinal data (e.g., "Strongly Disagree" to "Strongly Agree") that are subsequently converted into quantitative interval data (e.g., 1-5). This conversion assumes the scored data are interval level and allows for parametric statistical analyses. In turn, the interval data can then be used to calculate common Likert survey statistics such as the mean score and SD of individual items. Though there is widespread use of these procedures in survey research, literature within the field of measurement has cautioned the inherent assumptions tied to these analyses (Boone, 2016; Wilson et al., 2022). Harwell and Gatti (2001) classified such ordinal-to-interval conversion procedures as an appeal to Classical Test Theory (CTT), a measurement method with widely recognized limitations as compared to more robust psychometric methods like Item Response Theory (IRT; Wang, 1999). For a summative comparison of CTT and IRT, see Embretson (1996). Through IRT, ordinal data (e.g., Likert response data) can be "rescaled" to an interval scale (Harwell & Gatti, 2001). In other words, raw nonlinear response data can be converted to a linear scale, which can then be analyzed using parametric statistics (Boone, 2016).

This rescaling is critical in the analysis of Likert response data because the presumed "distance" between each response category on the Likert scale may differ within a single item and between several items (Boone, 2016). For example, in a single Likert-type item, the distance any one individual respondent considers between "Strongly Disagree" and "Disagree" may not be the same as the distance they consider between "Agree" and "Strongly Agree" even though both sets of distances are theoretically equal (i.e., one step distance). This complication is further compounded when considering the distance judgements made by a multitude of respondents across multiple items throughout a survey (Wang et al., 2006). Thus, it becomes necessary to utilize a measurement model like IRT that can accommodate these complications. In our analysis, we use the Rasch model (a specific subclass of IRT models) as a scaling model that can transform survey responses to an estimated score on a latent variable. Importantly, estimates based on the Rasch models are interval-level and do not have to assume any type of distribution (Wang, 1999). As summarized by Boone (2016), "Rasch analysis allows researchers to use a respondent's raw test or scale scores and express the respondent's performance on a linear scale that accounts for the unequal difficulties across all test items. Rasch techniques involve corrections for several psychometric issues (e.g., rating scales are ordinal, not all survey items mark the same part of the variable) so that accurate person measures can be computed" (p. 3).

To conduct our Rasch analysis, we fit the Innovation Skills survey data to various Rasch models using a computer program (ACER ConQuest v5.29). This analysis included a comparison of both unidimensional and multidimensional Rasch models as well as a comparison of the Rating Scale Model (RSM) and the Partial Credit Model (PCM). For purposes of brevity, the Rasch analysis results henceforth assume a unidimensional PCM. We plan to elaborate on this methodological decision in a future publication that compares the use of unidimensional and multidimensional Rasch analyses, as well as factor analysis, when analyzing the Innovation Skills survey data. Additionally, the main findings from our Rasch analysis focus on an anchored pre/post comparison. Item difficulty estimates of the same items from separate administrations were anchored to establish a common metric. Specifically, for the pre-survey, the item parameters were calibrated by anchoring the items to the estimates obtained from post-survey calibration. As a result, both pre- and post-survey were put on the common scale. This approach to analyzing the pre- and post-survey provides a strong factorial measurement invariance (Millsap, 2012).

Results

Survey Analysis Shows Strong Reliability Evidence

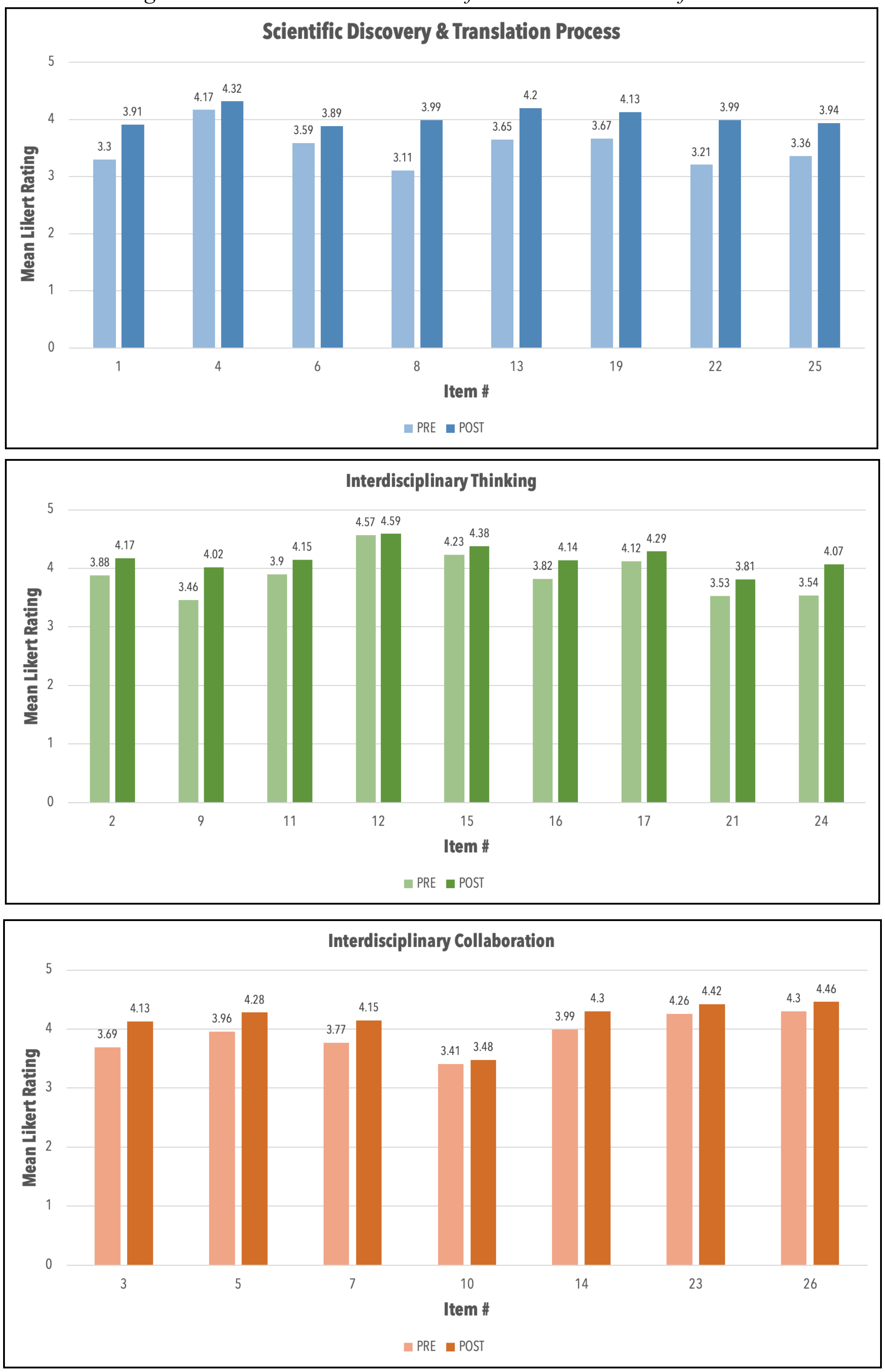
Reliability is a measure to show whether an instrument has demonstrated sufficient consistency to be useful. Here we present reliability evidence from an internal consistency perspective, or in other words, the consistency of item responses across the set of items in the instrument (Wilson, 2023). We report measures of reliability for the whole instrument rather than each subdomain based on the aforementioned assumption of a unidimensional PCM. The reliability of our Innovation Skills instrument is supported by a high Weighted Likelihood Estimates (WLE) person separation reliability (.93), a high expected a posteriori/plausible value (EAP/PV) reliability (.90), and a high coefficient (Cronbach's) alpha (.92).

General Likert Scale Analysis: Initial Support for Growth Hypothesis

The general Likert scale analysis consisted of a pre to post comparison of the mean Likert ratings for each of the 26 items on the Innovation Skills instrument. These results signify a *first step* of inquiry leading to a *first line* of empirical evidence that tests our hypothesis of student growth due to participation in the course. As shown in Figure 3, the mean Likert ratings for all 26 items were greater in the post-survey as compared to the pre-survey.

Figure 3

Pre/post Mean Likert Ratings Show Growth Across All Items for Each Subdomain of Innovation Skills Construct



Note. X axis shows survey item number for pre (light colored bar) and post (dark colored bar). Y axis shows mean Likert rating from 0-5.

In other words, there was an increase in the overall "agreeability" for each item. This initial analysis showed the mean Likert rating for each of the 26 items on the survey increased from pre to post, resulting in a positive delta value for each item. After completing the *Bioinspired Design* course, students showed growth in their self-efficacy for all Innovation Skills in every year of our analysis, including the overall combined analysis across the four-year dataset. The results from this general Likert scale analysis are summarized in Table 4. Even though many assessment studies utilize self-reported Likert data, we previously described the psychometric limitations and potentially problematic statistical procedures associated with this analysis (See Advancing Assessment section). We next present the results to our follow-up Rasch analysis.

Table 4
General Likert Scale Analysis Results Arranged in Descending Order of Delta (Δ)

Item	Subdomain	Level	Pre	Post	Δ	SD
8. I can translate fundamental principles from scientific publications to create novel designs.	SDTP	Leader	3.11	3.99	0.88	1.26
22. I can suggest fundamental principles from scientific publications that could translate into novel designs.	SDTP	Active	3.21	3.99	0.79	1.23
1. I read scientific publications to extract specific facts.	SDTP	Technical	3.3	3.91	0.61	1.39
25. I can critically evaluate scientific publications.	SDTP	Active	3.36	3.94	0.58	1.3
9. I am aware of concepts and methodologies across disciplines.	IT	Participant	3.46	4.02	0.56	1.19
13. I can understand how experts have translated principles from scientific publications into novel designs.	SDTP	Participant	3.65	4.2	0.54	1.13
24. I contribute to advancements in other fields with knowledge from my discipline.	IT	Leader	3.54	4.07	0.53	1.22
19. I can extract fundamental principles from scientific publications.	SDTP	Leader	3.67	4.13	0.46	1.13
3. I take a leadership role in interdisciplinary teams.	IC	Leader	3.69	4.13	0.44	1.22
7. I seek interdisciplinary collaborations to attain technical or factual knowledge outside my discipline.	IC	Technical	3.77	4.15	0.38	1.19
5. I take an active role in interdisciplinary collaborations.	IC	Active	3.96	4.28	0.32	1.11
16. I can apply concepts and methodologies across disciplines.	IT	Leader	3.82	4.14	0.31	1.05
14. I contribute to and benefit from interdisciplinary collaborations.	IC	Leader	3.99	4.3	0.31	1.08
6. I can understand the logic of scientific publications when discussed by experts.	SDTP	Participant	3.59	3.89	0.3	1.23
2. I use interdisciplinary concepts and methodologies to advance my own discipline.	IT	Leader	3.88	4.17	0.29	1.14

20. I attempt to apply concepts and methodologies across disciplines.	IT	Active	3.86	4.15	0.29	1.05
18. I seek opportunities outside my discipline.	IC	Active	3.87	4.16	0.29	1.16
21. I primarily use technical knowledge in my own discipline.	IT	Technical	3.53	3.81	0.27	1.39
11. I seek concepts and methodologies from other fields.	IT	Active	3.9	4.15	0.25	1.09
17. I seek information from my own discipline to solve problems.	IT	Required	4.12	4.29	0.18	1
26. I participate in diverse teams because I recognize the benefits of interdisciplinary knowledge.	IC	Participant	4.3	4.46	0.17	0.99
23. I participate in interdisciplinary teams when required.	IC	Required	4.26	4.42	0.16	0.97
4. I read scientific publications when they are required.	SDTP	Required	4.17	4.32	0.15	1.01
15. I seek information from other disciplines when required.	IT	Required	4.23	4.38	0.15	0.94
10. I prefer collaboration among members of my discipline.	IC	Required	3.41	3.48	0.07	1.42
12. I recognize the benefits from interacting with other disciplines.	IT	Participant	4.57	4.59	0.03	0.93

Delta was calculated based on the difference in mean Likert scale rating between the post-survey and pre-survey. (Adapted from S17 in Full et al., 2021).

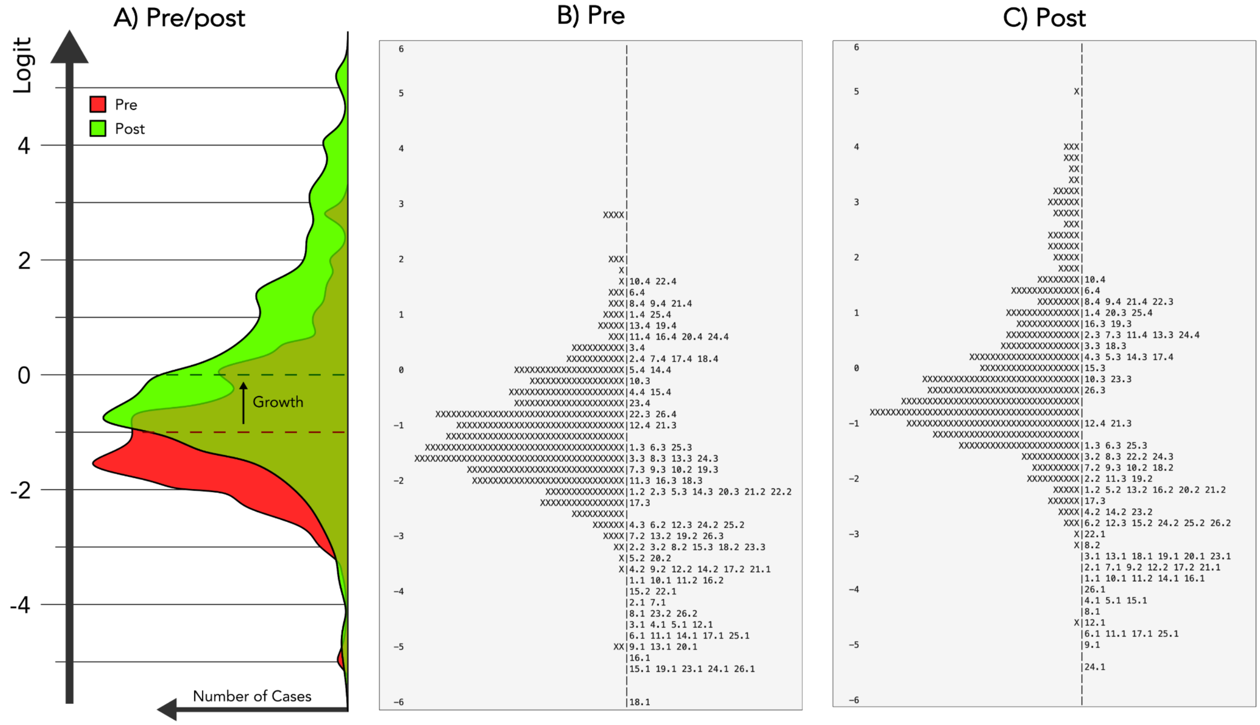
Rasch Analysis: Cross-Validating and Deepening Support for Growth Hypothesis

The following Rasch analysis results provide validity evidence for the Innovation Skills instrument by demonstrating that the data fit the Rasch model, an IRT approach that assesses psychometric properties at both item and person levels. These results support the validity of the instrument, indicating that the instrument measures the intended construct and that the items function as expected. The validity evidence gained from this approach complements traditional factor analysis methods of validation.

For example, in our Rasch analysis, we compared the results from the anchored pre-analysis with the results from the calibrated post-analysis, producing the pre- and post-Wright maps in Figure 4.

Figure 4

Wright Maps Based on Innovation Skills Pre/Post Survey Results



Note. Different sets of item steps are shown in these two Wright Maps. This is due to differing rates of response to the items in the two samples and does not affect the validity of the common scale. A) Growth from pre to post (+0.97 logits) visualized through overlaid Wright maps; B) Pre-survey Wright map, each 'X' represents 1.2 cases (persons), the labels for thresholds show the levels of item and category, respectively; C) Post-survey Wright map, each 'X' represents 1.0 cases (person), the labels for thresholds show the levels of item and category, respectively

The Wright maps show item responses and persons on a logit scale in which the more positive responses map to higher levels of the construct. Item parameter thresholds show the average difficulty (in logits) of the item thresholds relative to the other item thresholds. The distribution of the item parameter thresholds on a logit scale is displayed as the item number and threshold level (e.g., 10.4) on the right-hand side of the Wright map. The larger the item parameter threshold (or the higher on the Wright Map), the more difficult the item is to agree with, and therefore, less likely for a student to strongly agree with the item statement. The left-hand side of the Wright map displays the distribution of student abilities as X's. As a student's location increases, that student is more likely to agree with the item statements, indicating greater self-efficacy in Innovation Skills.

The Wright maps in Figure 4 indicate that there is strong quantifiable growth from pre to post in overall student ability, consistent with the results from the general Likert scale analysis. Additionally, because of the anchored nature of this Rasch analysis, a comparison of the constant in the anchored pre and in the post results in a difference that represents the "anchored" difference between the pre and post on the logit scale. This anchoring links the entire ability distribution between two time points—the post (0.00) and the anchored pre (-0.97). The difference between these two values indicates that the ability distribution of the anchored post was about 1 logit above the pre.

Overall, our Rasch analysis comparison of pre and post results showed an approximately 1 logit gain (0.97) in student ability as a result of completing the *Bioinspired Design* course. This increase is roughly equivalent to a one "step" increase in the Likert scale across all items (i.e., "Agree" to "Strongly agree"). Put differently, this logit gain translates to approximately 0.89 standard deviations of growth between pre and post. The logit gain can also be converted to a pooled standard deviation between the pre and the post, resulting in a Cohen's *d* effect size value of 0.75. This effect size is nearly a "large" effect (>0.8) based on Cohen's (1988) standards and an especially large effect in the context of educational interventions (>0.2) (Kraft, 2020).

Discussion

Connecting Innovation Skills Self-Efficacy to CBL and Science Connection

This study on Innovation Skills Self-Efficacy builds upon and extends the findings from the previous two chapters, offering a more comprehensive understanding of student development in our CBL *Bioinspired Design* course. In Chapter 1, we observed significant growth in science connection measures, with highest gains in Eff. The current study's findings corroborate and refine these previous results. The observed growth of approximately one standard deviation in Innovation Skills Self-Efficacy corresponds with the Eff gains while also providing a more refined understanding of the specific skills in which students are developing confidence. This growth suggests that our CBL approach enhances general scientific confidence alongside self-efficacy in the specific competencies needed for innovative thinking and problem-solving in STEM fields. The development of Innovation Skills Self-Efficacy also complements the growth in Science Identity (SciID) observed in Chapter 1. As students gain confidence in their ability to innovate within scientific contexts, they may increasingly see themselves as capable practitioners of science. This connection between Innovation Skills and SciID underscores the potential of our CBL approach to foster a more holistic development of students as future STEM-enriched professionals.

Chapter 2 showed that the growth in science connection measures, including Eff, was equitably distributed across diverse demographic groups. While this chapter's study did not explicitly examine demographic factors, the substantial overall growth in Innovation Skills Self-Efficacy suggests that our CBL approach may have the potential to foster these skills across a broad student population. This aligns with our course's inclusive design, which is open to students from all majors and backgrounds. Collectively, the findings across all three chapters demonstrate how a CBL environment can foster comprehensive student development. From broader science connection measures to specialized innovation-related self-efficacy, our *Bioinspired Design* course contributes to multiple constructs of student growth in the affective domain of learning. These outcomes align with the needs of 21st century STEM education, which emphasize both content knowledge and confidence in the skills necessary to apply that knowledge in challenge-based scenarios. Below, we further discuss the key outcomes from this chapter focused on Innovation Skills Self-Efficacy.

Outcomes of Measuring Innovation Skills Self-Efficacy through a Comprehensive Assessment Framework

Considering the intersecting areas for advancement in 21CS, bioinspired design, and Innovation Skills assessment, we utilized a comprehensive assessment framework (BAS and the Four Building Blocks) to develop and validate a survey instrument that more effectively shows student self-efficacy growth along developmental progressions. This framework enabled the alignment of instruction, assessment, and measurement tools to a clearly defined construct (self-efficacy of Innovation Skills) and a corresponding learning progression. This alignment yielded critical insights into stratified learner progress that traditional self-efficacy assessments may not capture. We structured this approach by implementing domain-specific self-efficacy measures (Bandura, 1997) that addressed the need for more robust assessment of affective constructs in STEM education (Koballa & Glynn, 2013; Maric et al., 2023).

We implemented our assessment framework in a large enrollment *Bioinspired Design* course open to all students. This is unlike many other discovery-based STEM courses that often have disciplinary barriers to entry (i.e., introductory prerequisites such as general biology), especially for students from non-STEM backgrounds. This course explicitly tasked all students to be innovators as they worked in interdisciplinary teams engaging in the bioinspired design process. Collectively, the course activities required students to use Innovation Skills. We set out to measure students' self-efficacy development in these skills based on a pre/post survey design analyzed through the Four Building Blocks of assessment. Our focus on self-efficacy in Innovation Skills builds upon previous work on innovation self-efficacy (Carberry et al., 2018) by extending it to a broader, interdisciplinary learning context. Crucial to our study was the development of a Construct Map used to test our hypothesis of student growth, further discussed below.

Utilizing the Construct Map to Foster a Developmental Perspective of Growth in Innovation Skills

In this study, we hypothesized that Innovation Skills form a continuum in which hierarchical levels of innovation connect with certain skills related to innovation. This resulted in a hypothesized Construct Map, which meaningfully tracks changes by situating measurement within developmental progressions. This approach tied closely with Bandura's (1997) social cognitive theory, particularly his concept of self-efficacy development through mastery experiences. In testing our hypothesis, both the pre- and the post-Wright maps showed expected banding patterns consistent with leveled thresholds. This revealed that the ordered levels of responses ("Strongly Disagree" to "Strongly Agree") were consistent with the hypothesized Construct Map levels ("Required" to "Leader"). This progressive growth is supported by Bandura's (1997) notion that self-efficacy development involves not just the acquisition of skills, but also the belief in one's ability to use these skills effectively across various situations. This is particularly relevant in the context of Innovation Skills, where students must develop confidence in their ability to apply their knowledge and skills to novel, complex problems.

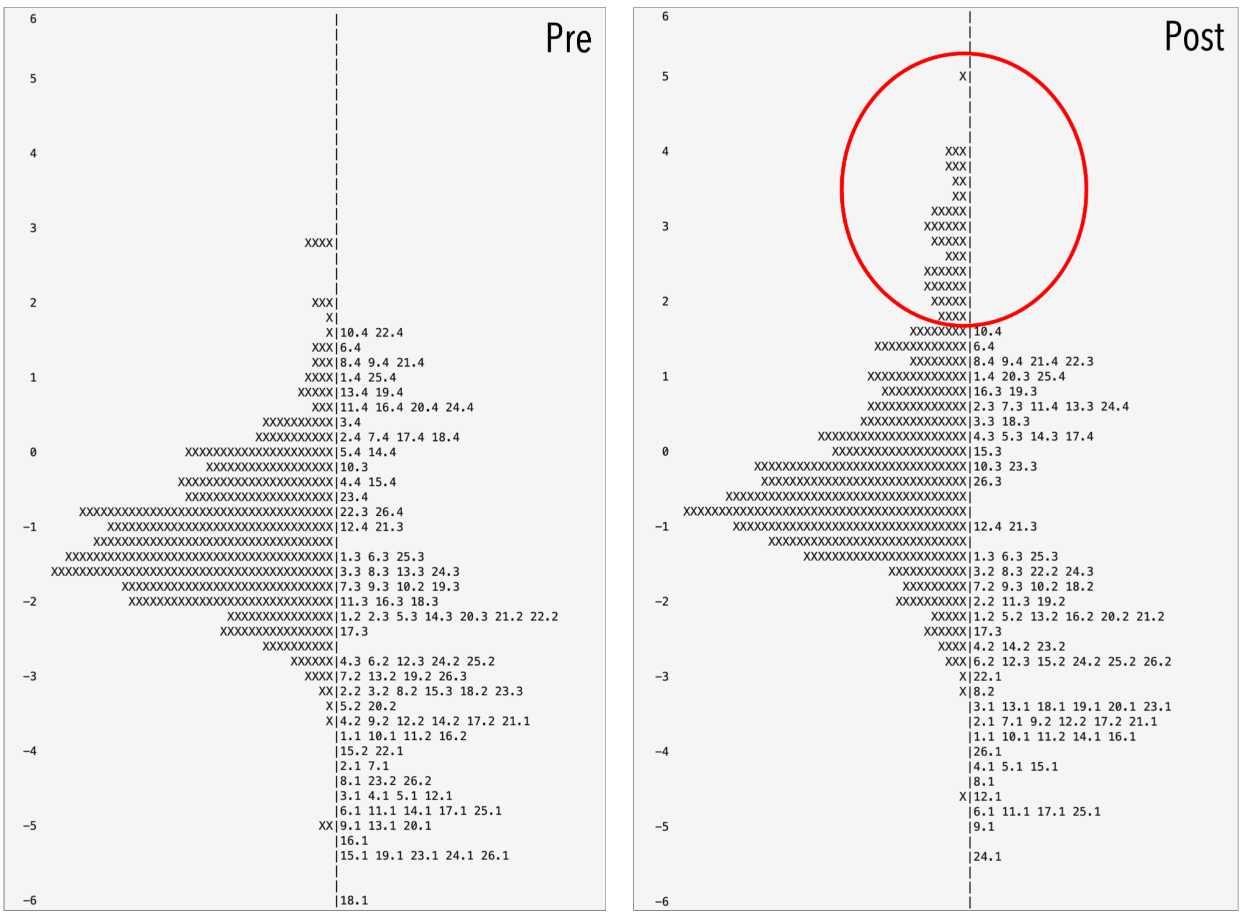
Additionally, since self-efficacy in Innovation Skills were hypothesized to develop throughout the course, students likely began the course with more limited preexisting self-efficacy. Self-efficacy is a malleable construct that can be influenced by environmental factors and learning experiences (Bandura, 1997). The observed growth from pre to post survey suggests that our course activities may have promoted mastery experiences, a key source of self-efficacy development (Bandura, 1997). The survey itself represents a tool to reveal and shape this developing self-efficacy in Innovation Skills. By mapping self-reported perceptions of growth on this novel construct, we have achieved an important first step in the eventual goal of developing an assessment that directly measures Innovation Skills.

Iterating on the Items Design to Expand Measurement of Self-Efficacy Development

Considering the cyclical nature of the Four Building Blocks, an important next step would be to iterate on the Items Design building block. This can be done by refining the instrument at the item level (i.e., developing new items that more effectively measure the Innovation Skills latent construct). For example, as compared to the pre-survey Wright map, the distribution of student abilities in the post-survey Wright map indicates that more items are needed to assess the higher end of the ability scale. In other words, the pre shows the distribution of student abilities appropriately mapped to more difficult items, but the post shows a large set of student abilities on the higher end of the logit scale well above any item thresholds (Figure 5).

Figure 5

Pre- and Post-Survey Wright maps on Logit Scale Vertical Axis (-6 to 6)



Note. Central dividing line separates persons (left side of central line) from item thresholds (right side of central line). Circled area in the post-survey Wright map indicates student abilities on the higher end of logit scale above item thresholds. This comparison shows growth in ability from pre to post, but also suggests more difficult-to-agree-with items are needed in the post-survey to accurately assess the higher end of the student ability distribution.

Bandura (2006) noted that effective self-efficacy scales should include items representing different levels of task demands, which corresponds to our Construct Map progression from “Required” to “Leader” levels. Our findings suggest an opportunity to further differentiate the upper end of the scale, allowing for more comprehensive measurement of high self-efficacy levels. More difficult-to-agree-with items are needed to accurately assess the higher end of the student ability distribution in the post. We interpret this to be a result of the *Bioinspired Design* course contributing to significant growth in students' self-efficacy in Innovation Skills. Future iterations on the Items Design, guided by this finding, will refine our instrument to better unveil the full spectrum of growth, particularly at the higher levels of ability.

Elevating Affective Assessment to Better Showcase Growth in Innovation Skills Self-Efficacy

Student voices are pivotal in understanding the ways our instruction impacts their development in the affective domain of learning. Gaining such insights requires strong measurement practices. As illustrated through the development of an assessment for Innovation Skills self-efficacy in our interdisciplinary *Bioinspired Design* course, the BAS framework centered student perceptions within an overarching developmental continuum spanning novice to expert abilities. This approach yielded key insights into stratified learner progress that static assessments fail to showcase. For instance, our Wright maps revealed a shift in student abilities towards the higher end of the logit scale post-course, indicating substantial growth in self-efficacy that might be missed by traditional pre-post comparisons. Students' self-efficacy growth was particularly pronounced in areas related to translating scientific principles into novel designs, a key aspect of innovation in STEM fields.

Overall, it is critical that students engage with courses that develop their Innovation Skills. At the same time, it is just as important to develop valid and reliable assessments of Innovation Skills to measure self-efficacy development in these skills. There is a need for a careful measurement approach that better differentiates the strengths and weaknesses of future STEM practitioners beyond just the STEM content they know or do not know. We need to better understand students' skills and their development of self-efficacy in those skills to improve USE. Our results, showing growth in self-efficacy across all items of our instrument, underscore the potential of targeted coursework to foster these crucial abilities. Additionally, our application of the BAS framework to assess gains in perceived innovation abilities represents an initial step toward more reliable and valid assessment. Future work must also expand beyond self-reports to include interviews, observations, and detailed mappings of student work products to evaluate alignment with survey responses.

Limitations and Future Research

In earlier years of our data collection, administrative challenges prevented us from matching pre and post-survey data for each individual respondent. Thus, we were unable to conduct paired analyses comparing pre/post survey responses for statistically significant differences (e.g., paired *t*-tests). In Table S8 of Supplement 5, we show a specific subset of paired analyses that corroborate with the conclusions from the overall Rasch analysis. Future studies should prioritize the collection of matched-pair data to analyze individual student development. Additionally, self-reported data may be subject to various biases (e.g., social desirability or response bias), potentially influencing validity. Despite this, numerical measures of non-cognitive variables, including those based on self-report, still provide scientifically useful data that accurately reflect respondents' inner emotional states (Kaiser & Oswald, 2022). Nonetheless, we plan to investigate how biases may be impacting item interpretations through cognitive interviews in future work. Further, while self-reported data provides valuable insights into students' own perceptions, our study did not include objective measures of students' Innovation Skills, such as direct assessments of their performance or products. In a future study, we plan to assess already collected course products alongside self-reported measures for a more comprehensive understanding of students' Innovation Skills development. Lastly, our study was conducted within the specific context of one course at a single university. Future research could explore the generalizability of the Innovation Skills construct in alternative educational contexts.

Chapter 3 Conclusion

We wish to conclude by emphasizing the importance of self-efficacy in Innovation Skills with the following thought experiment. Consider two job candidates for a STEM-centric position within a technology company. Both candidates have identical technical skills and identical background experiences. However, one of the candidates exemplifies far superior confidence in her ability to apply "soft" skills. She believes strongly in her capacity to communicate effectively within disciplinarily and demographically diverse teams, she is confident in her ability to integrate ideas from her non-technical peers in ways that enhance technical products, and she feels capable of solving interdisciplinary problems through effective team collaborations. The candidate's strong belief in her innovative capabilities is likely to differentiate her from other technically savvy colleagues. In today's knowledge-driven economy, this self-efficacy in Innovation Skills may drive her to attempt more innovative approaches, persist in the face of challenges, and ultimately contribute more effectively to her field. This robust sense of self-efficacy in innovation-related tasks also reflects a strong science connection, enhancing her potential for success in STEM-related fields.

We hope that all undergraduate courses integrate opportunities to develop self-efficacy in Innovation Skills because they represent a critical affective component underlying the transferable skills needed for both the known jobs of today and the unknown jobs of tomorrow. By fostering this self-efficacy through CBL approaches and measuring its development using comprehensive assessment frameworks, we can better prepare students to confidently tackle the complex, interdisciplinary challenges they will face in their future careers. This focus on self-efficacy in Innovation Skills represents a crucial step in bridging the gap between technical knowledge and the belief in one's ability to apply that knowledge innovatively in real-world contexts that enhance society.

Dissertation Conclusion

As Songer & Ruiz-Primo (2012) remarked, "What is measured, why and how it is measured, whose achievement is measured, what type of inferences are made, and the kinds of evidence supporting such inferences—all remain enduring issues in assessment development in science education (p. 688)." This dissertation addressed these enduring questions through three interconnected studies, each building upon the last to provide a substantive assessment of student development in a CBL environment. We began by establishing the overall effectiveness of CBL in cultivating science connection, then demonstrated the equitable nature of this development across diverse student populations, and finally introduced a novel construct and robust assessment framework to measure a specific skill set fostered by this learning environment. The progression through these three studies illustrated a comprehensive approach to assessment in USE, particularly within CBL environments.

The first chapter of this dissertation focused on measuring science connection, defined through an adaptation of the TIMSI framework and its three social influence constructs: SciID, Eff, and Val. We assessed these constructs in a diverse population of undergraduate students from over 40 majors participating in a single-semester *Bioinspired Design* course. Our aim was to evaluate the impact of a CBL intervention on fostering science connection in this interdisciplinary group of students. Using pre/post surveys analyzed through repeated measures ANOVAs and paired *t*-tests, we found significant increases in SciID and Eff, while Val remained stable. These findings demonstrated the effectiveness of CBL in promoting science connection within the span of a single semester.

Building on these initial results, the second chapter investigated the equity aspect of our findings. We examined the same TIMSI constructs but with a focus on potential differences across demographic groups. Our analysis encompassed several demographic variables, namely gender, underrepresented minority status, first-generation status, intended/declared major, class status, and term of enrollment. This investigation aimed to determine whether our CBL environment could promote equitable outcomes in science connection. We expanded our statistical analyses to include RM ANOVAs with demographic variables as between-subjects factors and ANCOVAs to control for pre-survey scores. The results showed equitable development in science connection across the majority of demographic groups, with differences in SciID development based on STEM/Non-STEM and class status.

The third chapter of this dissertation introduced a novel construct: self-efficacy in Innovation Skills. This construct was developed specifically for the CBL environment of our *Bioinspired Design* course, addressing the need to assess critical 21st century skills necessary for tackling complex global challenges. We measured this construct in the same diverse undergraduate student population, employing a comprehensive assessment framework known as the BAS. This framework utilized the Four Building Blocks approach and Rasch analysis to provide a more thorough evaluation of our research question. Our findings revealed significant growth in Innovation Skills self-efficacy, with students showing approximately one standard deviation of improvement between pre and post measurements.

Throughout these studies, several overarching themes emerged. The CBL framework consistently proved effective in promoting science connection and Innovation Skills across a diverse student population, demonstrating its potential as a scalable and adaptable model for USE reform. The use of TIMSI provided valuable insights into how students develop a connection to science, especially as it relates to growth in SciID and Eff. The equitable outcomes observed across demographic groups positively expands notions of who can succeed in STEM and highlights the potential of inclusive pedagogy. By focusing on constructs within the affective domain, particularly self-efficacy, this research highlights the importance of developing students' beliefs in their abilities alongside their cognitive skills. The development and assessment of Innovation Skills self-efficacy responds to a crucial need in preparing students for future workforce demands. The progression from the statistical analyses in earlier chapters to the BAS demonstrates the value of comprehensive assessment methods in capturing 21st century skills. By integrating instruction with evaluation tools like BAS, this research empowers educators to develop, refine, and validate assessments that capture complex competencies essential for student success. The Four Building Blocks approach, in particular, provides a powerful tool for creating assessments aligned with developmental progressions, offering deeper insights into student growth.

In conclusion, this dissertation contributes to USE by demonstrating how CBL can foster equitable development of science connection and self-efficacy in Innovation Skills. It also provides a model for deeper assessment that addresses the enduring issues mentioned above by Songer & Ruiz-Primo (2012). As we strive to cultivate the skills needed in our future workforce, integrated assessment frameworks will be more necessary than ever to promote STEM-enriched learning. By understanding not just what students know, but also how they perceive their abilities and connect with science, we can better prepare all students to take on the challenges of our rapidly changing world. This research lays a foundation for future studies that can further explore the intersection of innovative pedagogical approaches, equitable outcomes, and robust assessment methodologies in USE.

References

Adams, R. J., Wu, M. L., Cloney, D., Berezner, A., & Wilson, M. (2020). ACER ConQuest: Generalised Item Response Modelling Software (Version 5.29) [Computer software]. Australian Council for Educational Research. https://www.acer.org/au/conquest

Adedokun, O. A., Bessenbacher, A. B., Parker, L. C., Kirkham, L. L., & Burgess, W. D. (2013). Research skills and STEM undergraduate research students' aspirations for research careers: Mediating effects of research self-efficacy. Journal of Research in Science Teaching, 50(8), 940-951.

Ainscough, L., Foulis, E., Colthorpe, K., Zimbardi, K., Robertson-Dean, M., Chunduri, P., & Lluka, L. (2016). Changes in biology self-efficacy during a first-year university course. CBE—Life Sciences Education, 15(2), ar19.

AlShebli, B. K., Rahwan, T., & Woon, W. L. (2018). The preeminence of ethnic diversity in scientific collaboration. Nature communications, 9(1), 5163.

American Association for the Advancement of Science. (2011). Vision and change: A call to action, final report. Washington, DC. Retrieved July 3, 2018.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). Standards for educational and psychological testing. American Educational Research Association.

Apple, Inc. (2008). Challenge Based Learning: An Approach for Our Time. Retrieved from https://www.apple.com/ca/education/docs/Apple-ChallengedBasedLearning.pdf

Aschbacher, P.R., Li, E. and Roth, E.J. (2010), Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine. J. Res. Sci. Teach., 47: 564-582. https://doi.org/10.1002/tea.20353

Autor, D., Chin, C., Salomons, A. M., & Seegmiller, B. (2022). New Frontiers: The Origins and Content of New Work, 1940–2018. National Bureau of Economic Research.

Bakshi, A., Patrick, L. E., & Wischusen, E. W. (2016). A Framework for Implementing Course-Based Undergraduate Research Experiences (CUREs) in Freshman Biology Labs. The American Biology Teacher, 78(6), 448–455.

Ballen, C. J., Blum, J. E., Brownell, S., Hebert, S., Hewlett, J., Klein, J. R., McDonald, E. A., Monti, D. L., Nold, S. C., Slemmons, K. E., Soneral, P. A. G., & Cotner, S. (2017a). A Call to Develop Course-Based Undergraduate Research Experiences (CUREs) for Nonmajors Courses. CBE—Life Sciences Education, 16(2), mr2.

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017b). Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. CBE—Life Sciences Education, 16(4), ar56.

Bandura, A. (1997). Self-efficacy: The exercise of control. (pp. ix, 604). W H Freeman/Times Books/ Henry Holt & Co.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), Self-efficacy beliefs of adolescents (Vol. 5, pp. 307-337). Greenwich, CT: Information Age Publishing.

Barber, J. P. (2012). Integration of Learning: A Grounded Theory Analysis of College Students' Learning. American Educational Research Journal, 49(3), 590–617. https://doi.org/10.3102/0002831212437854

Batchelor, R. L., H. Ali, K. G. Gardner-Vandy, A. U. Gold, J. A. MacKinnon, and P. M. Asher (2021), Reimagining STEM workforce development as a braided river, Eos, 102, https://doi.org/10.1029/2021EO157277. Published on 19 April 2021.
Bellanca, J. A., & Brandt, R. S. (Eds.). (2010). 21st century skills: Rethinking how students learn. Solution Tree Press.
Beltran, R. S., Marnocha, E., Race, A., Croll, D. A., Dayton, G. H., & Zavaleta, E. S. (2020). Field courses narrow demographic achievement gaps in ecology and evolutionary biology. Ecology and Evolution, 10(12), 5184-5196.
Bhatti, H. A., (2021). Toward "Inclusifying" the Underrepresented Minority in STEM Education Research. Journal of Microbiology & Biology Education, 22(3), e00202-21. https://doi.org/10.1128/jmbe.00202-21
Bhatti, H.A., Mehta, S., McNeil, R., Yao, SY., Wilson, M. (2023). A Scientific Approach to Assessment: Rasch Measurement and the Four Building Blocks. In: Liu, X., Boone, W.J. (eds) Advances in Applications of Rasch Measurement in Science Education. Contemporary Trends and Issues in Science Education, vol 57. Springer, Cham. https://doi.org/10.1007/978-3-031-28776-3_7
Birbeck, D., & Andre, K. (2009). The affective domain: beyond simply knowing. ATN Assessment Conference, 40-47.
Blanca, M. J., Alarcón, R., & Arnau, J. (2017). Non-normal data: Is ANOVA still a valid option? Psicothema, 29.4, 552–557.
Blanca, M. J., Arnau, J., García-Castro, F. J., Alarcón, R., & Bono, R. (2023). Non-normal Data in Repeated Measures ANOVA: Impact on Type I Error and Power. Psicothema, 35.1, 21–29.
Bloodhart, B., Balgopal, M. M., Casper, A. M. A., Sample McMeeking, L. B., & Fischer, E. V. (2020). Outperforming yet undervalued: Undergraduate women in STEM. PLOS ONE, 15(6), e0234685.
Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay Company.
Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? CBE—Life Sciences Education, 15(4), rm4. https://doi.org/10.1187/cbe.16-04-0148
Börner, K., Contractor, N., Falk-Krzesinski, H. J., Fiore, S. M., Hall, K. L., Keyton, J., ... & Uzzi, B. (2010). A multi-level systems perspective for the science of team science. Science translational medicine, 2(49), 49cm24-49cm24.
Box, G. E. P. (1949). A general distribution theory for a class of likelihood criteria. Biometrika, 36(3/4), 317-346.
Camfield, E. K., Schiller, N. R., & Land, K. M. (2021). Nipped in the Bud: COVID-19 Reveals the Malleability of STEM Student Self-Efficacy. CBE—Life Sciences Education, 20(2), ar25. https://doi.org/10.1187/cbe.20-09-0206
Carberry, A. R., Gerber, E. M., & Martin, C. K. (2018). Measuring the innovation self-efficacy of engineers*. International Journal of Engineering Education, 34(2), 590-598.
Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. Journal of Research in Science Teaching, 44(8), 1187–1218. https://doi.org/10.1002/tea.20237
Carnevale, A. P., Smith, N., & Melton, M. (2011). STEM: Science Technology Engineering Mathematics. In Georgetown University Center on Education and the Workforce.

Georgetown University Center on Education and the Workforce. https://eric.ed.gov/?id=ED525297
Chemers, M. M., Zurbriggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. Journal of Social Issues, 67(3), 469-491.
Chimi, C. J., & Russell, D. L. (2009, November). The Likert scale: A proposal for improvement using quasi-continuous variables. In Information Systems Education Conference, Washington, DC (pp. 1-10).
Chirazi, J., Wanieck, K., Fayemi, P.-E., Zollfrank, C., & Jacobs, S. (2019). What Do We Learn from Good Practices of Biologically Inspired Design in Innovation? Applied Sciences, 9(4), 650. https://doi.org/10.3390/app9040650
Chu, S. K. W., Reynolds, R. B., Tavares, N. J., Notari, M., & Lee, C. W. Y. (2017). 21st Century Skills Development Through Inquiry-Based Learning. Springer Singapore. https://doi.org/10.1007/978-981-10-2481-8
Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587
Cole, M. F., & Beck, C. W. (2022). Developmental trajectories of student self-perception over a yearlong introductory biology sequence. CBE—Life Sciences Education, 21(3), ar59.
Cotner, S., Thompson, S., & Wright, R. (2017). Do biology majors really differ from non–STEM majors?. CBE—Life Sciences Education, 16(3), ar48.
Dear Colleague Letter: Bioinspired Design Collaborations to Accelerate the Discovery-Translation Process (BioDesign) (nsf23055) | NSF - National Science Foundation. (n.d.). Retrieved July 6, 2023, from https://www.nsf.gov/pubs/2023/nsf23055/nsf23055.jsp?WT.mc_ev=click&WT.mc_id=&utm_medium=email&utm_source=govdelivery
Dimitrov, D. M., & Rumrill Jr, P. D. (2003). Pretest-posttest designs and measurement of change. Work, 20(2), 159-165.
Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British journal of psychology, 105(3), 399-412.
Durand, F., Helms, M. E., Tsenn, J., McTigue, E., McAdams, D. A., & Linsey, J. S. (2015). Teaching Students to Innovate: Evaluating Methods for Bioinspired Design and Their Impact on Design Self Efficacy. Volume 7: 27th International Conference on Design Theory and Methodology, V007T06A003. https://doi.org/10.1115/DETC2015-47716
Embretson, S. E. (1996). The new rules of measurement. Psychological Assessment, 8(4), 341.
Estrada, M., Burnett, M., Campbell, A. G., Campbell, P. B., Denetclaw, W. F., Gutiérrez, C. G., ... & Zavala, M. (2016). Improving underrepresented minority student persistence in STEM. CBE—Life Sciences Education, 15(3), es5.
Estrada, M., Hernandez, P. R., & Schultz, P. W. (2018). A Longitudinal Study of How Quality Mentorship and Research Experience Integrate Underrepresented Minorities into STEM Careers. CBE—Life Sciences Education, 17(1), ar9.
Estrada, M., Matsui, J., Eppig, A., & Flores, L. (2019, July 30). A Longitudinal Study of the Biology Scholars Program: Maintaining Student Integration and Intention to Persist in Science Career Pathways: Published in Understanding Interventions. Understanding Interventions.

Estrada, M., Schultz, P. W., Silva-Send, N., & Boudrias, M. A. (2017). The Role of Social Influences on Pro-Environment Behaviors in the San Diego Region. Journal of Urban Health, 94(2), 170–179.

Estrada, M., Woodcock, A., Hernandez, P. R., & Schultz, P. W. (2011). Toward a model of social influence that explains minority student integration into the scientific community. Journal of Educational Psychology, 103(1), 206–222.

Estrada, M., Young, G. R., Flores, L., Yu, B., & Matsui, J. (2021). Content and Quality of Science Training Programs Matter: Longitudinal Study of the Biology Scholars Program. CBE—Life Sciences Education, 20(3), ar44.

Field, A. (2024). Discovering statistics using IBM SPSS statistics (6th ed.). SAGE Publications.

Forakis, J., & March, J. L. (2023). COVID-19 and the Incoming Chemistry Student: The Effect of the Pandemic on Student Self-Efficacy and Identity. Journal of Chemical Education, 100(1), 371–375. https://doi.org/10.1021/acs.jchemed.2c00312

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., ... & Barabási, A. L. (2018). Science of science. Science, 359(6379), eaao0185.

Fried, E., Martin, A., Esler, A., Tran, A., & Corwin, L. (2020). Design-based learning for a sustainable future: Student outcomes resulting from a biomimicry curriculum in an evolution course. Evolution: Education and Outreach, 13(1), 22. https://doi.org/10.1186/s12052-020-00136-6

Full, R. J., Bhatti, H. A., Jennings, P., Ruopp, R., Jafar, T., Matsui, J., Flores, L. A., & Estrada, M. (2021). Eyes Toward Tomorrow Program Enhancing Collaboration, Connections, and Community Using Bioinspired Design. Integrative and Comparative Biology, icab187. https://doi.org/10.1093/icb/icab187

Full, R. J., Dudley, R., Koehl, M. A. R., Libby, T., & Schwab, C. (2015). Interdisciplinary Laboratory Course Facilitating Knowledge Integration, Mutualistic Teaming, and Original Discovery. Integrative and Comparative Biology, 55(5), 912–925. https://doi.org/10.1093/icb/icv095

Gallagher, S. E., & Savage, T. (2020). Challenge-based learning in higher education: An exploratory literature review. Teaching in Higher Education, 1-23. https://doi.org/10.1080/13562517.2020.1863354

Garibay, J. C. (2015). STEM students' social agency and views on working for social change: Are STEM disciplines developing socially and civically responsible students?. Journal of Research in Science Teaching, 52(5), 610-632.

Geisinger, K. F. (2016). 21st Century Skills: What Are They and How Do We Assess Them? Applied Measurement in Education, 29(4), 245–249. https://doi.org/10.1080/08957347.2016.1209207

Gerber, E. M., Marie Olson, J., & Komarek, R. L. (2012a). Extracurricular design-based learning: Preparing students for careers in innovation. International Journal of Engineering Education, 28(2), 317.

Gerber, E., Martin, C. K., Kramer, E., Braunstein, J., & Carberry, A. R. (2012b). Work in progress: Developing an Innovation Self-efficacy survey. 2012 Frontiers in Education Conference Proceedings, 1–3. https://doi.org/10.1109/FIE.2012.6462435

Gormally, C., & Heil, A. (2022). A Vision for University Biology Education for Non-science Majors. CBE—Life Sciences Education, 21(4), es5.

Griffin, P., & Care, E. (Eds.). (2015). Assessment and Teaching of 21st Century Skills. Springer Netherlands. https://doi.org/10.1007/978-94-017-9395-7

Gutiérrez, K. D., Morales, P. Z., & Martinez, D. C. (2009). Re-mediating Literacy: Culture, Difference, and Learning for Students From Nondominant Communities. Review of Research in Education, 33(1), 212-245. https://doi.org/10.3102/0091732X08328267

Gvili, I. E. F., Weissburg, M. J., Yen, J., Helms, M. E., & Tovey, C. A. (2016). Development of scoring rubric for evaluating integrated understanding in an undergraduate biologically-inspired design course. Synthesis, 11, 14.

Hamilton, V. L. (2004). Identification as a Challenge to Dual-Process Theories of Persuasion. In A. H. Eagly, R. M. Baron, & V. L. Hamilton (Eds.), The social psychology of group identity and social conflict: Theory, application, and practice (pp. 65–76). American Psychological Association. https://doi.org/10.1037/10683-004

Handelsman, J., Elgin, S., Estrada, M., Hays, S., Johnson, T., Miller, S., Mingo, V., Shaffer, C., & Williams, J. (2022). Achieving STEM diversity: Fix the classrooms. Science, 376(6597), 1057–1059.

Harwell, M. R., & Gatti, G. G. (2001). Rescaling Ordinal Data to Interval Data in Educational Research. Review of Educational Research, 71(1), 105–131. https://doi.org/10.3102/00346543071001105

Hazari, Z., Sadler, P. M., & Sonnert, G. (2013). The science identity of college students: Exploring the intersection of gender, race, and ethnicity. Journal of College Science Teaching, 42(5), 82-91.

Hebert, S., & Cotner, S. (2019). A comparison of nonmajors' & majors' incoming science process skills. The American Biology Teacher, 81(8), 554-560.

Helms, M., Vattam, S. S., & Goel, A. K. (2009). Biologically inspired design: Process and products. Design Studies, 30(5), 606–622. https://doi.org/10.1016/j.destud.2009.04.003

Hernandez, P. R., Agocha, V. B., Carney, L. M., Estrada, M., Lee, S. Y., Loomis, D., Williams, M., & Park, C. L. (2020). Testing models of reciprocal relations between social influence and integration in STEM across the college years. PLOS ONE, 15(9), e0238250.

Hernandez, P. R., Schultz, P., Estrada, M., Woodcock, A., & Chance, R. C. (2013). Sustaining optimal motivation: A longitudinal analysis of interventions to broaden participation of underrepresented students in STEM. Journal of educational psychology, 105(1), 89.

Hofstra B, Kulkarni VV, Munoz-Najar Galvez S, He B, Jurafsky D, McFarland DA. 2020. The Diversity–Innovation Paradox in Science. Proc Natl Acad Sci. 117:9284–91.

IBM Corp. Released 2021. IBM SPSS Statistics for Macintosh, Version 28.0. Armonk, NY: IBM Corp

Innovation. Nature Index. (2022, December 7). https://www.nature.com/collections/ghgedjjigf

Jacobs, S., Eggermont, M., Helms, M., & Wanieck, K. (2022). The Education Pipeline of Biomimetics and Its Challenges. Biomimetics, 7(3), 93. https://doi.org/10.3390/biomimetics7030093

Jayaram, K., & Full, R. J. (2016). Cockroaches traverse crevices, crawl rapidly in confined spaces, and inspire a soft, legged robot. Proceedings of the National Academy of Sciences, 113(8), E950–E957. https://doi.org/10.1073/pnas.1514591113

Johnson, L. F., Smith, R. S., Smythe, J. T., & Varon, R. K. (2009). Challenge-based learning: An approach for our time. The New Media Consortium.

Kaiser, C., & Oswald, A. J. (2022). The scientific value of numerical measures of human feelings. Proceedings of the National Academy of Sciences, 119(42), e2210412119. https://doi.org/10.1073/pnas.2210412119

Keinänen, M., Ursin, J., & Nissinen, K. (2018). How to measure students' innovation competences in higher education: Evaluation of an assessment tool in authentic learning environments. Studies in Educational Evaluation, 58, 30–36. https://doi.org/10.1016/j.stueduc.2018.05.007
Kelman, H. C. (1958). Compliance, identification, and internalization three processes of attitude change. Journal of Conflict Resolution, 2(1), 51-60. https://doi.org/10.1177/002200275800200106
Knight, J. K., & Smith, M. K. (2010). Different but equal? How nonmajors and majors approach and learn genetics. CBE—Life Sciences Education, 9(1), 34-44.
Koballa, T. R., & Glynn, S. M. (2013). Attitudinal and motivational constructs in science learning. In Handbook of research on science education (pp. 75-102). Routledge.
Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. Educational Researcher, 49(4), 241–253. https://doi.org/10.3102/0013189X20912798
Krathwohl, D.R., Bloom, B.S., and Masia, B.B. (1964). Taxonomy of educational objectives: Handbook II: Affective domain. New York: David McKay Co.
Lave, J., & Wenger, E. (1991). Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511815355
Laverty, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., Caballero, M. D., Fata-Hartley, C. L., Ebert-May, D., Jardeleza, S. E., & Cooper, M. M. (2016). Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol. PLOS ONE, 11(9), e0162333. https://doi.org/10.1371/journal.pone.0162333
Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, & H. B. Mann (Eds.), Contributions to probability and statistics: Essays in honor of Harold Hotelling (pp. 278-292). Stanford University Press.
Li, T., Oseguera, L., & Kirk, C. (2024). Examining the influence of first-generation status on STEM socialization among undergraduates in a STEM scholars program. Research in Higher Education, 65, 417-438.
Linn, M. C., Palmer, E., Baranger, A., Gerard, E., & Stone, E. (2015). Undergraduate research experiences: Impacts and opportunities. Science, 347(6222), 1261757.
Luke, R. (2013). Measuring Innovation Skills Acquired by College and Polytechnic Students through Applied Research. Technology Innovation Management Review.
MacPhee, D., Farro, S., & Canetto, S. S. (2013). Academic self-efficacy and performance of underrepresented STEM majors: Gender, ethnic, and social class patterns. Analyses of Social Issues and Public Policy, 13(1), 347-369.
Maher, J. M., Markey, J. C., & Ebert-May, D. (2013). The Other Half of the Story: Effect Size Analysis in Quantitative Research. CBE—Life Sciences Education, 12(3), 345–351. https://doi.org/10.1187/cbe.13-04-0082
Malmqvist, J., Rådberg, K. K., & Lundqvist, U. (2015). Comparative analysis of challenge-based learning experiences. Proceedings of the 11th International CDIO Conference, Chengdu, Sichuan, PR China.
Maric, D., Fore, G. A., Nyarko, S. C., & Varma-Nelson, P. (2023). Measurement in STEM education research: A systematic literature review of trends in the psychometric evidence of scales. International Journal of STEM Education, 10(1), 39. https://doi.org/10.1186/s40594-023-00430-x

Marshman, E. M., Kalender, Z. Y., Nokes-Malach, T., Schunn, C., & Singh, C. (2018). Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm?. Physical Review Physics Education Research, 14(2), 020123.

Maxwell, S.E., Delaney, H.D., & Kelley, K. (2017). Designing Experiments and Analyzing Data: A Model Comparison Perspective, Third Edition (3rd ed.). Routledge. https://doi.org/10.4324/9781315642956

McConnell, D. A., & van Der Hoeven Kraft, K. J. (2011). Affective Domain and Student Learning in the Geosciences. Journal of Geoscience Education, 59(3), 106–110. https://doi.org/10.5408/1.3604828

Means, B., & Neisler, J. (2020). Unmasking Inequality: STEM Course Experience During the COVID-19 Pandemic. Digital Promise. https://doi.org/10.51388/20.500.12265/102

Membrillo-Hernández, J., Ramírez-Cadena, M. J., Martínez-Acosta, M., Cruz-Gómez, E., Muñoz-Díaz, E., & Elizalde, H. (2019). Challenge based learning: The importance of world-leading companies as training partners. International Journal on Interactive Design and Manufacturing (IJIDeM), 13(3), 1103-1113. https://doi.org/10.1007/s12008-019-00569-4

Miller, C. T., Drewery, M., Waliczek, T. M., Contreras, R. N., & Kubota, C. (2023). Engaging Undergraduate Students in Research. HortTechnology, 33(1), 1–7.

Miller, D. I., & Wai, J. (2015). The bachelor's to Ph.D. STEM pipeline no longer leaks more women than men: A 30-year analysis. Frontiers in Psychology, 6. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2015.00037

Millsap, R. E. (2012). Statistical approaches to measurement invariance. Routledge.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A Brief Introduction To Evidence-Centered Design. ETS Research Report Series, 2003(1), i–29. https://doi.org/10.1002/j.2333-8504.2003.tb01908.x

Morán-Soto, G., Marsh, A., González Peña, O. I., Sheppard, M., Gómez-Quiñones, J. I., & Benson, L. C. (2022). Effect of the COVID-19 Pandemic on the Sense of Belonging in Higher Education for STEM Students in the United States and Mexico. Sustainability, 14(24), 16627. https://doi.org/10.3390/su142416627

Mukherjee, S., Romero, D. M., Jones, B., & Uzzi, B. (2017). The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot. Science advances, 3(4), e1601315.

Nagel, J. K. S., & Pidaparti, R. M. (2016). Significance, Prevalence and Implications for Bio-Inspired Design Courses in the Undergraduate Engineering Curriculum. Volume 3: 18th International Conference on Advanced Vehicle Technologies; 13th International Conference on Design Education; 9th Frontiers in Biomedical Devices, V003T04A009. https://doi.org/10.1115/DETC2016-59661

National Academies of Sciences, Engineering, and Medicine. (2012). Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. Washington, DC: The National Academies Press. https://doi.org/10.17226/13398.

National Academies of Sciences, Engineering, and Medicine. (2014). Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond. Washington, DC: The National Academies Press. https://doi.org/10.17226/18722.

National Academies of Sciences, Engineering, and Medicine. (2015). Reaching Students: What Research Says About Effective Instruction in Undergraduate Science and Engineering. Washington, DC: The National Academies Press. https://doi.org/10.17226/18687.
National Academies of Sciences, Engineering, and Medicine. (2016). Science Literacy: Concepts, Contexts, and Consequences. Washington, DC: The National Academies Press. https://doi.org/10.17226/23595.
National Academies of Sciences, Engineering, and Medicine. (2018). The Integration of the Humanities and Arts with Sciences, Engineering, and Medicine in Higher Education: Branches from the Same Tree (D. Skorton & A. Bear, Eds.; p. 24988). National Academies Press. https://doi.org/10.17226/24988
National Academies of Sciences, Engineering, and Medicine. (2018). Indicators for Monitoring Undergraduate STEM Education (M. B. Rosenberg, M. L. Hilton, & K. A. Dibner, Eds.; p. 24943). National Academies Press. https://doi.org/10.17226/24943
National Academies of Sciences, Engineering, and Medicine. (2021). Call to Action for Science Education: Building Opportunity for the Future (M. Honey, H. Schweingruber, K. Brenner, & P. Gonring, Eds.; p. 26152). National Academies Press. https://doi.org/10.17226/26152
National Academies of Sciences, Engineering, and Medicine. (2022). Imagining the Future of Undergraduate STEM Education: Proceedings of a Virtual Symposium (K. Brenner, A. Beatty, & J. Alper, Eds.; p. 26314). National Academies Press. https://doi.org/10.17226/26314
National Academies of Sciences, Engineering, and Medicine. 2021. Undergraduate and Graduate STEM Students Experiences During COVID-19: Proceedings of a Virtual Workshop Series. Washington, DC: The National Academies Press. https://doi.org/10.17226/26024.
National Academy of Engineering. 2015. Educate to Innovate: Factors That Influence Innovation: Based on Input from Innovators and Stakeholders, Washington, DC: National Academies Press
National Endowment for the Arts. (2017). Industrial Design: A Competitive Edge for U.S. Manufacturing Success in the Global Economy. Washington, DC: National Endowment for the Arts.
National Research Council, Division on Earth, Life Studies, Board on Life Sciences, & Committee on Key Challenge Areas for Convergence. (2014). Convergence: Facilitating transdisciplinary integration of life sciences, physical sciences, engineering, and beyond. National Academies Press.
National Science Board, "Vision 2030." May 2020 Report #: NSB-2020-15, page 7.
National Science Board. 2015. Revisiting the STEM Workforce, A Companion to Science and Engineering Indicators 2014, Arlington, VA: National Science Foundation (NSB-2015-10)
Nelson, B., Wilson, J., & Yen, J. (2009). A study of biologically-inspired design as a context for enhancing student innovation. 2009 39th IEEE Frontiers in Education Conference, 1–5. https://doi.org/10.1109/FIE.2009.5350871
Newell, M. J., & Ulrich, P. N. (2022). Gains in Scientific Identity, Scientific Self-Efficacy, and Career Intent Distinguish Upper-Level CUREs from Traditional Experiences in the Classroom. Journal of Microbiology & Biology Education, 23(3), e00051-22.

Nguyen, B. M. D., Alcantar, C. M., Curammeng, E. R., Hernandez, E., Kim, V., Paredes, A. D., Freeman, R., Nguyen, M. H., & Teranishi, R. T. (2017). The Racial Heterogeneity Project: Implications for Educational Research, Practice, and Policy. ACT, Inc.

Pellegrino, J. (2017), "Teaching, learning and assessing 21st century skills", in Guerriero, S. (ed.), Pedagogical Knowledge and the Changing Nature of the Teaching Profession, OECD Publishing, Paris, https://doi.org/10.1787/9789264270695-12-en.

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Ramírez-Lugo, J. S., Toledo-Hernández, C., Vélez-González, I., & Ruiz-Diaz, C. P. (2021). CREARE: A Course-Based Undergraduate Research Experience to Study the Responses of the Endangered Coral Acropora Cervicornis to a Changing Environment †. Journal of Microbiology & Biology Education, 22(1), ev22i1.2253.

Robinson, K. A., Perez, T., Carmel, J. H., & Linnenbrink-Garcia, L. (2019). Science identity development trajectories in a gateway college chemistry course: Predictors and relations to achievement and STEM pursuit. Contemporary Educational Psychology, 56, 180-192.

Robnett, R. D., & Thoman, S. E. (2017). STEM success expectancies and achievement among women in STEM majors. Journal of Applied Developmental Psychology, 52, 91-100.

Robnett, R. D., Chemers, M. M., & Zurbriggen, E. L. (2015). Longitudinal associations among undergraduates' research experience, self-efficacy, and identity. Journal of Research in Science Teaching, 52(6), 847–867.

Rodenbusch, S. E., Hernandez, P. R., Simmons, S. L., & Dolan, E. L. (2016). Early Engagement in Course-Based Research Increases Graduation Rates and Completion of Science, Engineering, and Mathematics Degrees. CBE—Life Sciences Education, 15(2), ar20.

Rovalo, E., McCardle, J., Smith, E., & Hooker, G. (2020). Growing the practice of biomimicry: Opportunities for mission-based organisations based on a global survey of practitioners. Technology Analysis & Strategic Management, 32(1), 71–87. https://doi.org/10.1080/09537325.2019.1634254

Sadler, T. D. (2009). Situated learning in science education: Socio‐scientific issues as contexts for practice. Studies in Science Education, 45(1), 1–42. https://doi.org/10.1080/03057260802681839

Savery, J. R. (2006). Overview of Problem-Based Learning: Definitions and Distinctions. Interdisciplinary Journal of Problem-Based Learning, 1(1), 9-20. doi:10.7771/1541-5015.1002

Seymour, E., & Hunter, A.-B. (Eds.). (2019). Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education. Springer International Publishing.

Seymour, E., Hunter, A. B., Laursen, S. L., & DeAntoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three‐year study. Science education, 88(4), 493-534.

Shortlidge, E. E., Gray, M. J., Estes, S., & Goodwin, E. C. (2024). The value of support: STEM intervention programs impact student persistence and belonging. CBE—Life Sciences Education, 23(2), ar23.

Shuster, M. I., Curtiss, J., Wright, T. F., Champion, C., Sharifi, M., & Bosland, J. (2019). Implementing and Evaluating a Course-Based Undergraduate Research Experience (CURE) at a Hispanic-Serving Institution. Interdisciplinary Journal of Problem-Based Learning, 13(2).

Siddiq, F., Gochyyev, P., & Wilson, M. (2017). Learning in Digital Networks – ICT literacy: A novel assessment of students' 21st century skills. Computers & Education, 109, 11–37. https://doi.org/10.1016/j.compedu.2017.01.014

Singh, S., & Aggarwal, Y. (2022). In search of a consensus definition of innovation: A qualitative synthesis of 208 definitions using grounded theory approach. Innovation: The European Journal of Social Science Research, 35(2), 177–195. https://doi.org/10.1080/13511610.2021.1925526

Songer, N. B., & Ruiz-Primo, M. A. (2012). Assessment and science education: Our essential new priority? Journal of Research in Science Teaching, 49(6), 683–690. https://doi.org/10.1002/tea.21033

Spell, R. M., Guinan, J. A., Miller, K. R., & Beck, C. W. (2014). Redefining Authentic Research Experiences in Introductory Biology Laboratories and Barriers to Their Implementation. CBE—Life Sciences Education, 13(1), 102–110.

Taber, K.S. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. Res Sci Educ 48, 1273–1296 (2018). https://doi.org/10.1007/s11165-016-9602-2

Trenholm, S. (1989). Persuasion and social influence. Englewood Cliffs, NJ: Prentice Hall.

Trujillo, G., & Tanner, K. D. (2014). Considering the Role of Affect in Learning: Monitoring Students' Self-Efficacy, Sense of Belonging, and Science Identity. CBE—Life Sciences Education, 13(1), 6–15. https://doi.org/10.1187/cbe.13-12-0241

Vincent-Ruz, P., Schunn, C.D. The nature of science identity and its role as the driver of student choices. IJ STEM Ed 5, 48 (2018).

Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. Journal of Curriculum Studies, 44(3), 299–321. https://doi.org/10.1080/00220272.2012.668938

Vue, Z., Vang, C., Vue, N., Kamalumpundi, V., Barongan, T., Shao, B., Huang, S., Vang, L., Vue, M., Vang, N., Shao, J., Coombes, C., Katti, P., Liu, K., Yoshimura, K., Biete, M., Dai, D.-F., Phillips, M. A., & Behringer, R. R. (2023). Asian Americans in STEM are not a monolith. Cell, 186(15), 3138–3142. https://doi.org/10.1016/j.cell.2023.06.017

Wang, W.-C. (1999). Direct estimation of correlations among latent traits within IRT Framework. Methods of Psychological Research, 4(2).

Wang, W., Wilson, M., & Shih, C. (2006). Modeling randomness in judging rating scales with a random‐effects rating scale model. Journal of Educational Measurement, 43(4), 335–353.

Wanieck, K., Ritzinger, D., Zollfrank, C., & Jacobs, S. (2020). Biomimetics: Teaching the tools of the trade. FEBS Open Bio, 10(11), 2250–2267. https://doi.org/10.1002/2211-5463.12963

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48: 817–838.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.

Wilson, M. (2005). Constructing Measures: An Item Response Modeling Approach. New York, Routledge.

Wilson, M. (2023). Constructing Measures: An Item Response Modeling Approach (2nd ed.). Routledge. https://doi.org/10.4324/9781003286929

Wilson, M. and Sloane, K. (2000). From principles to practice: An embedded assessment system. Applied Measurement in Education,13(2), 181–208.

Wilson, M., & Scalise, K. (2006). Assessment to improve learning in higher education: The BEAR Assessment System. Higher Education, 52(4), 635–663. https://doi.org/10.1007/s10734-004-7263-y

Wilson, M., Bathia, S., Morell, L., Gochyyev, P., Koo, B. W., & Smith, R. (2022). Seeking a better balance between efficiency and interpretability: Comparing the likert response format with the Guttman response format. Psychological Methods. https://doi.org/10.1037/met0000462

World Economic Forum (2020), “The future of jobs report 2020”, available at: http://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf (accessed 16 March 2023).

Yosso, T. J. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. Race Ethnicity and Education, 8(1), 69–91. https://doi.org/10.1080/1361332052000341006

Zimmerman, B. J. (2000). Self-Efficacy: An Essential Motive to Learn. Contemporary Educational Psychology, 25(1), 82–91. https://doi.org/10.1006/ceps.1999.1016

Appendices

Supplement 1

Further Explanation of Effect Size Interpretation based on Recommendations by Kraft (2020)

In line with recent discussions in educational research (Kraft, 2020), we aimed to interpret our effect sizes within the specific context of our short-term educational intervention. This approach involved considering factors such as the duration of the intervention, the nature of the outcome measures, and typical effect sizes observed in similar educational studies. By contextualizing our effect sizes in this way, we sought to provide a more nuanced and meaningful interpretation of the practical significance of our results, acknowledging that even relatively small effect sizes can be meaningful in educational settings, particularly for brief interventions.

We obtained and evaluated effect sizes (partial eta squared [η_p^2] values for ANOVA results and Cohen's *d* values for *t*-tests) to assess the practical significance of the observed changes. For partial eta squared, we maintained the generally accepted field guidelines of .01 (small effect), .06 (medium effect), and .14 (large effect) (Cohen, 1988). We retained these benchmarks as they are more encompassing than Cohen's *d*, providing flexibility for our context, while still offering a familiar point of reference. Importantly, we also interpret these benchmarks relative to our short-term intervention and diverse student sample, factors which typically yield smaller effects.

For Cohen's *d*, we adapted the generally accepted benchmarks to better fit our specific context. While the generally accepted values for small, medium, and large effects are 0.2, 0.5, and 0.8 respectively, we adjusted these to < 0.1 for small, 0.1 to < 0.3 for medium, and ≥ 0.3 for large effects. These values fall between Cohen's benchmarks and recently proposed benchmarks for educational interventions by Kraft (2020) (small < 0.05, medium = 0.05 to < 0.20, large = $\geq$ 0.20). We justify this adaptation based on several factors specific to our study below.

First, our intervention was relatively short-term (one semester) and involved a large, demographically diverse sample of students, which typically results in smaller effect sizes compared to longer-term or more targeted interventions. Second, we used broad measures of science connection (SciID, Eff, and Val) rather than narrow, specialized tests. These broader measures typically yield smaller effect sizes compared to more focused, intervention-specific assessments. Third, our intervention targeted undergraduate students across various disciplines. At this educational stage, students' attitudes, identities, and self-efficacy related to science are likely more established and potentially resistant to change compared to those of K-12 students. Consequently, even relatively small shifts in these constructs may represent meaningful developments in students' science connection.

Supplement 2

Further Explanation of Reliability based on Recommendations by Maric et al. (2023)

Based on a range of Cronbach's alpha (α) values from 0.894 to 0.927 (see Table 2 in main text), we labeled the overall internal consistency of our instrument as "good to excellent." This labeling was based on other commonly cited thresholds and their subsequent qualitative descriptors (Maric et al., 2023), although the choice of threshold and descriptor is somewhat arbitrary (Taber, 2018). Some sources suggest $\alpha > 0.7$ is "acceptable," while others argue for a higher standard of $\alpha > 0.9$ for "excellent" consistency. Given the evaluation purposes of this instrument and the measurement of TIMSI constructs in previous studies, we set a threshold of $\alpha > 0.8$ and compared each of our values to that threshold, thus our qualitative labeling of "good to excellent."

Considering this context, the high α values (all $\alpha > 0.89$) obtained for each subscale and the overall instrument indicate that the items are measuring the same underlying construct in a consistent way. If the instrument lacked sufficient internal consistency, (i.e., low alpha values; $\alpha < 0.7$), it would suggest the items are not reliably measuring the intended constructs of SciID, Eff, and Val (and the overall construct of science connection). This could occur if some items were poorly written, misinterpreted, or not actually relevant to the construct. Inconsistent responses across items would reduce confidence that the instrument can dependably assess students' SciID, Eff, Val, and, by extension, science connection.

To further examine reliability and align with suggestions by both Maric et al. (2023) and Taber (2018), McDonald's omega (ω) was calculated to supplement and verify Cronbach's alpha results. McDonald's omega makes fewer assumptions (e.g., equal item loadings) that are often violated in practice (Dunn et al., 2014). Omega values were comparably high: 0.912 and 0.923 for the pre and post-surveys respectively, supporting the internal consistency of the instrument. Overall, the reliability evidence collected through Cronbach's alpha (pre $\alpha = 0.916$, post $\alpha = 0.925$) and McDonald's omega (pre $\omega = 0.912$, post $\omega = 0.923$), together with the instrument's grounding in previously validated measures, provides confidence the survey consistently measures SciID, Eff, and Val among this population of students.

Table S1
Reliability of Survey Constructs (McDonald's Omega [ω])

Construct	N of Items	N (pre)	N (post)	ω (pre)	ω (post)
SciID	5	966	554	0.926	0.929
Eff	8	958	546	0.906	0.909
Val	6	956	540	0.894	0.905
Overall	19	944	529	0.912	0.923

Supplement 3

Results for RM ANOVA Assumptions Tests

In RM ANOVA, when adding between-subjects factors such as demographic group variables, Box's test results are obtained to test the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups. Box's test assumptions are not met when significance values are < .001. In our analysis, Box's test assumptions were not met for the Biology Major and STEM/Non-STEM variables as shown in Table S2.

Table S2
Box's Test Results for Demographic Variables

Variable	Box's M	F	df1	df2	Sig.	Assumptions
Gender	10.920	.513	21	616089	.967	Met
URM	20.730	.941	21	24236	.536	Met
FirstGen	19.311	.873	21	22163	.627	Met
Biology Major	54.321	2.542	21	259438	<.001	Not met
STEM/Non-STEM	59.125	2.773	21	411735	<.001	Not met
Class Status	35.596	1.672	21	679655	.027	Met
Term	120.055	1.389	84	309721	.011	Met

Given Box's test's high sensitivity to departures from normality and unequal group sizes (Field, 2024), we supplemented our analysis with Levene's test. Levene's test evaluates the equality of variances for each dependent variable individually and is recognized as a more robust test for this assumption. Levene's test assumptions are not met when significance values are < .05. Levene's test results for Biology Major and STEM/Non-STEM are shown in Tables S3 and S4, respectively. For the Biology Major group, Levene's test assumptions were not met for the Val construct in the pre-survey. For the STEM/Non-STEM group, Levene's test assumptions were not met for the Val construct in both pre and post surveys.

Table S3
RM ANOVA - Levene's Test - Biology Major

Construct	Levene Statistic	Sig.	Assumptions
SciID_pre	2.391	.123	Met
SciID_post	1.109	.293	Met
Eff_pre	3.528	.061	Met
Eff_post	3.035	.082	Met
Val_pre	8.075	.005	Not met
Val_post	.391	.532	Met

Table S4
RM ANOVA - Levene's Test - STEM/Non-STEM

Construct	Levene Statistic	Sig.	Assumptions
SciID_pre	.092	.762	Met
SciID_post	.063	.803	Met
Eff_pre	.683	.409	Met
Eff_post	.993	.319	Met
Val_pre	13.640	<.001	Not met
Val_post	12.860	<.001	Not met

These findings suggest that although there are differences in covariance matrices across groups (as indicated by Box's test), the assumption of equal variances is upheld for the vast majority of constructs and time points, excluding the Val construct in comparisons involving Biology Major and STEM/Non-STEM groups. Considering these assumption test results, we present both multivariate tests (Pillai's Trace, Wilks' Lambda) and univariate tests in the main text. Multivariate tests are typically more resilient to violations of homogeneity assumptions (Field, 2024). For univariate tests, results related to the Val construct may be less reliable, particularly for the Biology Major and STEM/Non-STEM comparisons.

Results for ANCOVA Assumptions Tests

Table S5

Levene's Test Results for SciID ANCOVA

Variable	F	df1	df2	Sig.	Assumptions
Gender	.044	1	511	.834	Met
URM	2.492	1	512	.115	Met
FirstGen	.002	1	320	.961	Met
Biology Major	.503	1	522	.479	Met
STEM/Non-STEM	.529	1	522	.468	Met
Class Status	2.984	1	523	.085	Met
Term	1.569	4	520	.181	Met

Table S6

Levene's Test Results for Eff ANCOVA

Variable	F	df1	df2	Sig.	Assumptions
Gender	.014	1	503	.905	Met
URM	1.217	1	505	.271	Met
FirstGen	1.406	1	317	.237	Met
Biology Major	.000	1	514	.991	Met
STEM/Non-STEM	1.817	1	514	.178	Met
Class Status	6.187	1	515	.013	Not met
Term	.275	4	512	.894	Met

Table S7

Levene's Test Results for Val ANCOVA

Variable	F	df1	df2	Sig.	Assumptions
Gender	.027	1	493	.869	Met
URM	.055	1	496	.814	Met
FirstGen	1.960	1	313	.163	Met
Biology Major	.388	1	504	.534	Met
STEM/Non-STEM	7.723	1	504	.006	Not met
Class Status	.000	1	505	.990	Met
Term	.513	4	502	.726	Met

Supplement 4

Detailed Demographics of Innovation Skills Survey Respondents

Class standing:

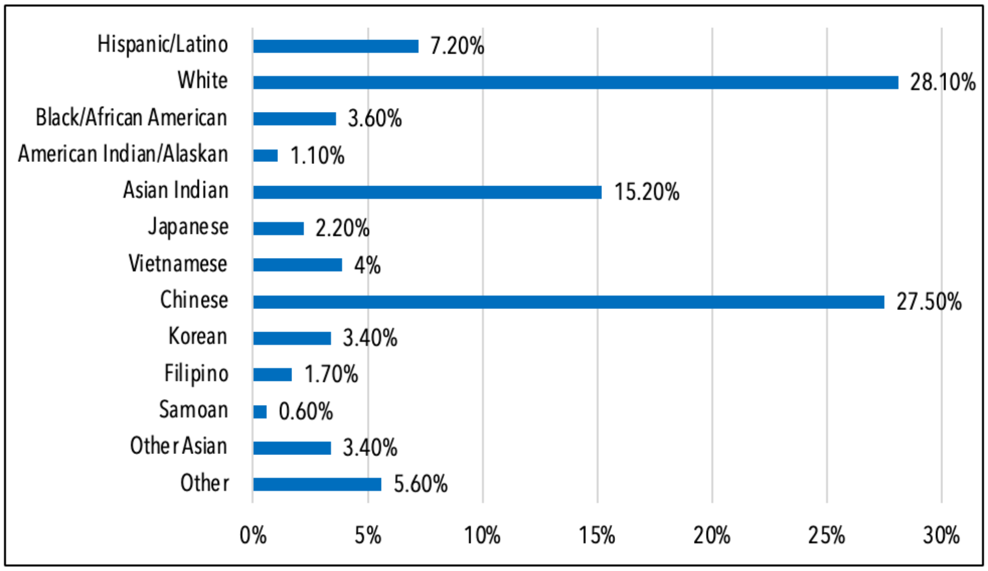
	2018	**2019**	**2020**
First year	26%	27%	45%
Second year	25%	19%	22%
Third year	26%	34%	15%
Fourth year	21%	20%	19%

Gender:

2018	
Female	55.5%
Male	43.9%
Other or declined to respond	0.6%
2019	
Female	57%
Male	41%
Other	1%
Non-binary	1%
2020	
Female	63.7%
Male	35.6%
Non-binary	0.7%

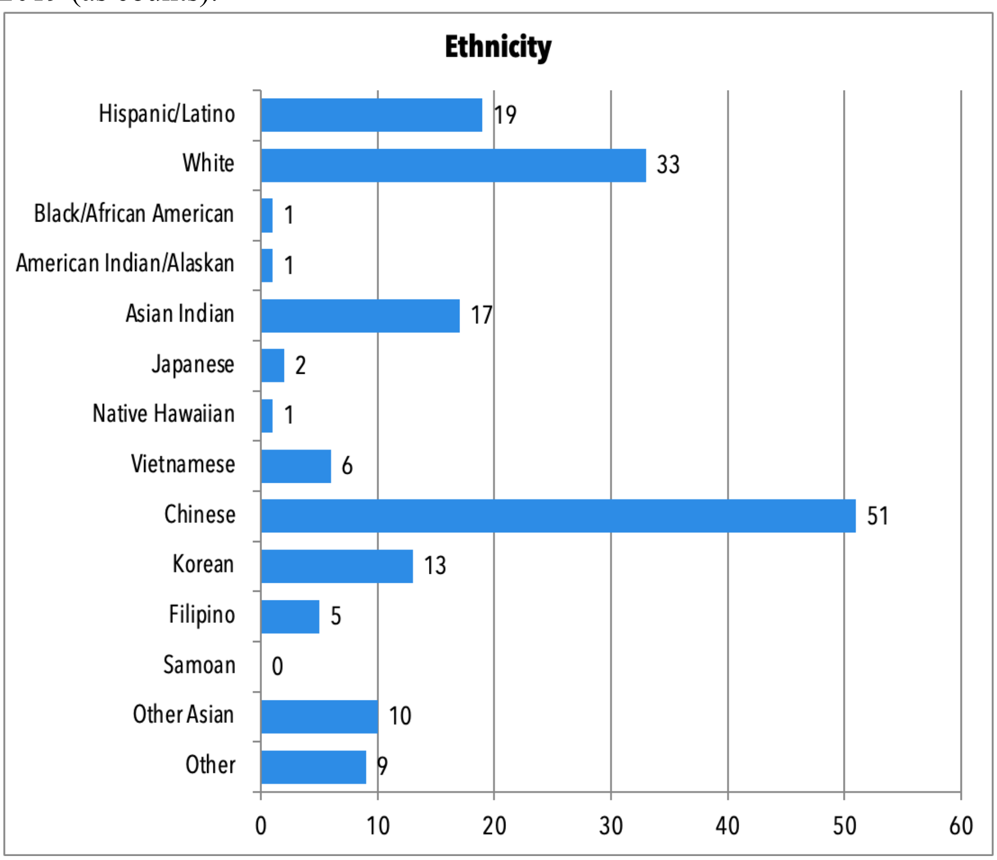
Ethnicity:

2018:



Participants were able to choose more than one option.

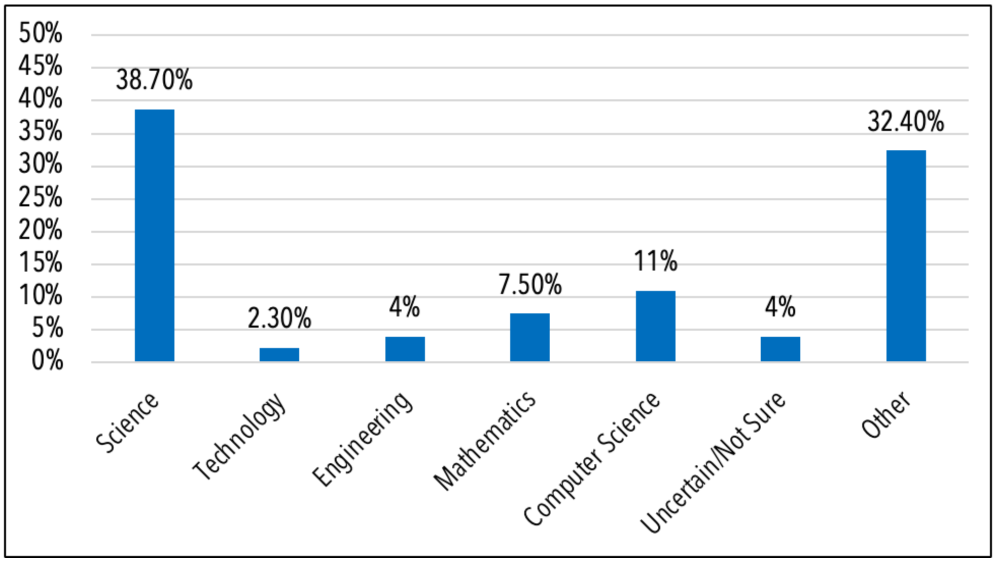
2019 (as counts):



Participants were able to choose more than one option. Responses in the "Other" category included: Nepali, Pakistani, Taiwanese, Basque, Cambodian, Middle Eastern, and Persian.

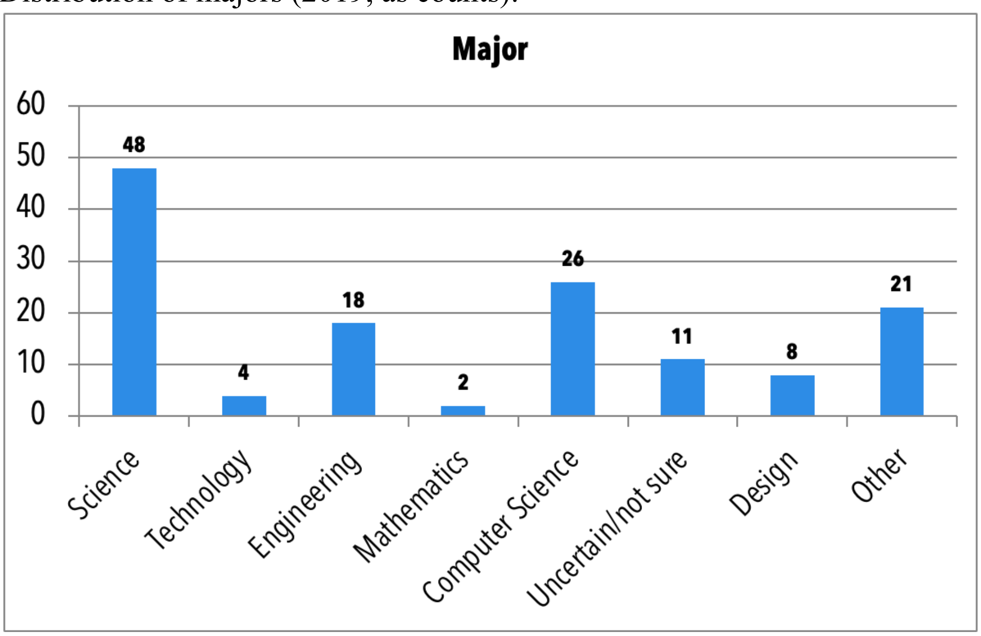
Majors:

Distribution of majors (2018):

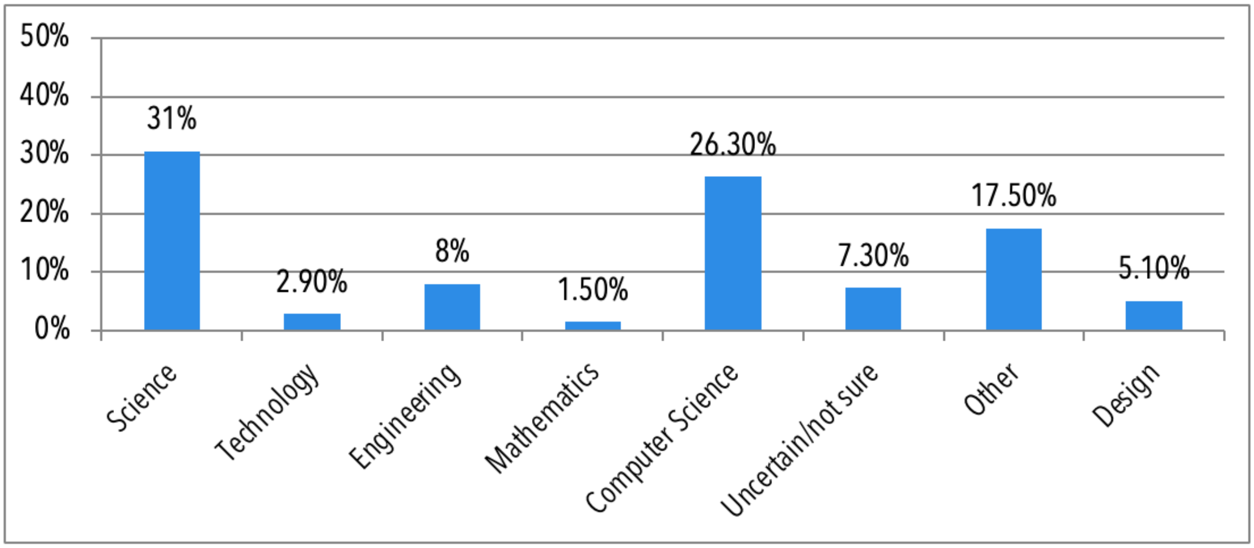


The "Other" category included: Social Sciences, Economics, Political Science, Architecture, Business, Art, City Planning, Cognitive Studies, Environmental Economics and Policy, Interdisciplinary Studies, International Relations, Media Studies, Political Economics, Sociology/Legal Studies, and Sustainability

Distribution of majors (2019, as counts):



The "Other" category included: Cognitive Science, French/Linguistics/Education, Social Science, Data science, Business, Economics, Public Health, Environmental Economics and Policy, Political Economy, Architecture, Comparative Ethnic Studies, and Political Science
Distribution of majors (2020):



The “Other” category included: English, Linguistics, Social Science, Global Management, Business, Economics, Data Science, Bioethics, Public Health, Political Science, Legal Studies, Architecture, Art Practice, and History

Is your current/intended major biology?					
2018		2019		2020	
Yes	31%	Yes	23%	Yes	25%
No	69%	No	77%	No	75%

Supplement 5

Paired Comparisons

Table S8
Item Level Paired t-Tests for Matched Pairs Subset of Survey Respondents

Item (Post-Pre)	Mean Difference	SD	Std. Error Mean	95% CI Lower	95% CI Upper	t	df	Sig. (2-tailed)
Item1_SDTP	.356	1.131	.059	.240	.473	6.016	364	**<.001**
Item2_IT	.110	.889	.047	.018	.201	2.355	364	**.019**
Item3_IC	.266	.908	.048	.173	.360	5.601	363	**<.001**
Item4_SDTP	.047	.767	.040	-.032	.126	1.160	364	.247
Item5_IC	.143	.871	.046	.053	.233	3.134	362	**.002**
Item6_SDTP	.204	.903	.047	.111	.297	4.303	362	**<.001**
Item7_IC	.163	.920	.048	.068	.258	3.372	361	**<.001**
Item8_SDTP	.679	.989	.052	.577	.781	13.084	363	**<.001**
Item9_IT	.413	.997	.052	.310	.516	7.895	362	**<.001**
Item10_IC	.028	1.129	.059	-.089	.144	.465	362	.642
Item11_IT	.157	.842	.044	.070	.245	3.557	361	**<.001**
Item12_IT	-.044	.719	.038	-.118	.030	-1.168	362	.244
Item13_SDTP	.507	.916	.048	.413	.601	10.575	364	**<.001**
Item14_IC	.144	.826	.043	.058	.229	3.307	361	**.001**
Item15_IT	.030	.752	.039	-.047	.108	.768	362	.443
Item16_IT	.262	.842	.044	.175	.349	5.928	361	**<.001**
Item17_IT	.083	.844	.044	-.004	.171	1.874	359	.062
Item18_IC	.150	.904	.048	.056	.243	3.145	360	**.002**
Item19_SDTP	.465	.906	.048	.372	.559	9.754	360	**<.001**
Item20_IT	.168	.858	.045	.079	.257	3.730	362	**<.001**
Item21_IT	.254	1.056	.056	.145	.363	4.579	361	**<.001**
Item22_SDTP	.634	.981	.051	.532	.735	12.308	362	**<.001**
Item23_IC	.080	.868	.046	-.010	.169	1.754	362	.080
Item24_IT	.389	.964	.051	.289	.489	7.655	359	**<.001**
Item25_SDTP	.511	.972	.051	.411	.611	10.030	363	**<.001**
Item26_IC	.052	.799	.042	-.030	.135	1.246	363	.214

*Bold indicates significant difference between post item mean and pre item mean ($p < 0.05$)

Supplemental Figures

Figure S1

Discovery Decomposition tool used to decompose Autumn et al. (2000) paper read as part of class assignment. (Retrieved from S2 in Full et al., 2021).

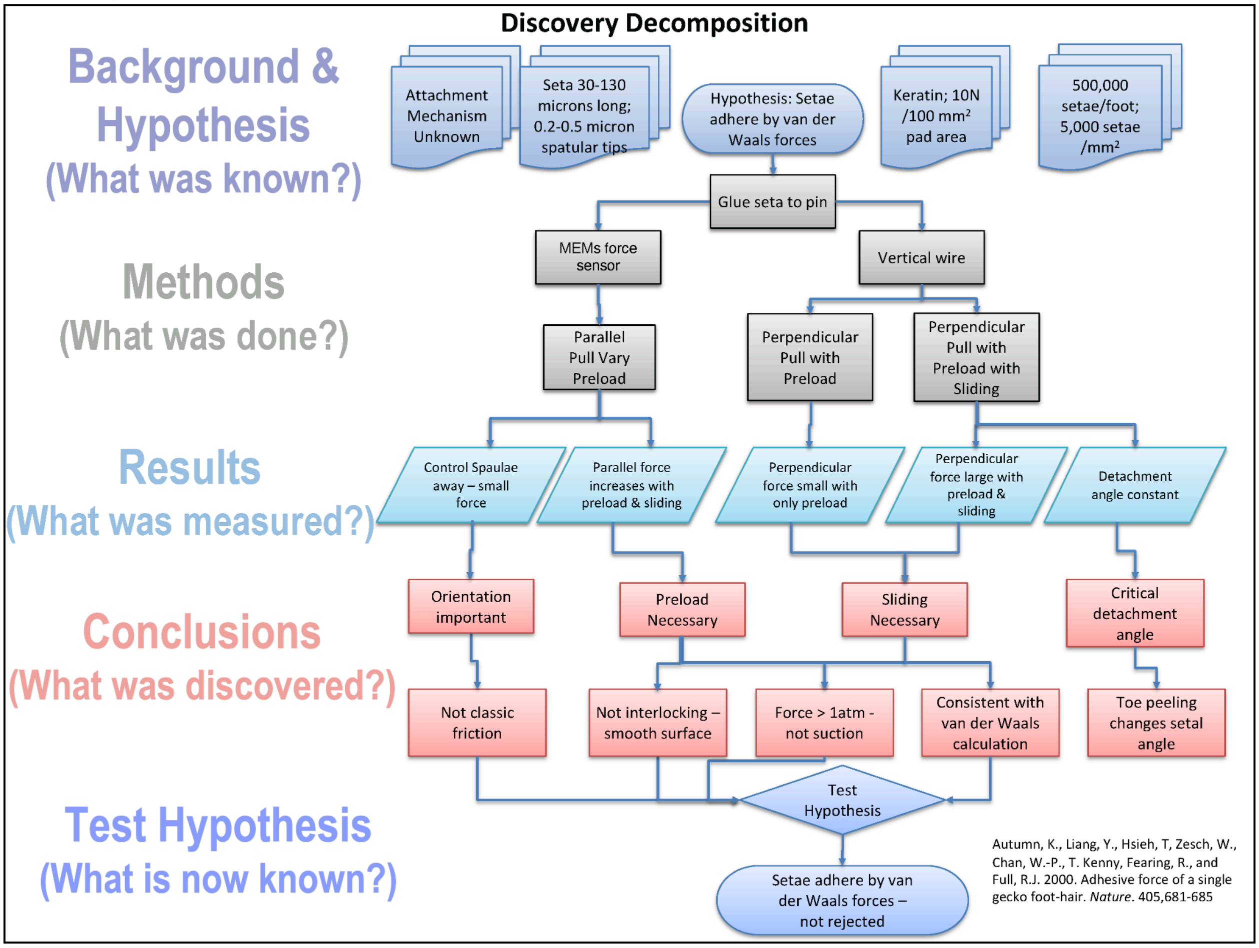


Figure S2

Analogy Check tool used to create an analogy based on a design solution from nature. (Analogy based on discoveries in Jayaram & Full, 2016; Retrieved from S2 in Full et al., 2021).

Design Solution	Analogy Check	Design Problem
Cockroach (*Periplaneta americana)*		**Search-and-Rescue Robot**
Behaviors (What does system or organism do?)		**Behaviors** (What do you want system to do?)
Crevice traversal	Similar	Crevice traversal
Rapid confined space locomotion	Similar	Rapid confined space locomotion
Structural Components (What is structure or organization of system?)		**Structural Components** (What can the structure be?)
Chitin	Different	Cardboard
Compliant membranes (arthrodial)	Uncertain	Polyester
Stiff plates and tubes	Uncertain	Cardboard
Operating Environment (Where?)		**Operating Environment** (Where?)
Crevice - Plexiglas	Uncertain	Crevice - rubble
Confined space (tunnel) - Plexiglas	Uncertain	Confined space (tunnel) - rubble
Size (What is size?)		**Size** (What size needed?)
Standing height (6mm)	Different	Standing height (75mm)
Compressed height (3mm)	Different	Compressed height (35mm)
Functional Mechanisms (How does system work?)		**Functional Mechanisms** (How do you want the system to work?)
Compresses by soft membranes	Uncertain	Compresses by polyester membranes
Propulsors (legs) assist	Uncertain	Propulsors (legs) assist
Body friction low	Uncertain	Body friction low with polyester
Viscoelastic, but tough	Uncertain	Viscoelastic, but tough
Characteristics/Specification (Which are distinguishing?)		**Characteristics/Specification** (What are your specifications?)
High compression	Similar	High compression
Robust to compression	Different	Far less robust to compression
Performance Criteria (How well does system work?)		**Performance Criteria** (How well must the system work?)
50% body compression	Similar	50% body compression
20 body lengths/s, confined space	Different	1 body lengths/s, confined space
800X body weight compression	Different	20X body weight compression
Constraints (What compromises system?)		**Constraints** (Can compromises be removed?)
Developmental - moulting	Uncertain	Fixed material
Multi-functional – many other behaviors	Uncertain	Single behavior
Evolution – chitin inherited from ancestors	Uncertain	Material can be varied