8-2023

A Mixed Methods Analysis of Influences Surrounding Undergraduate Science Recruitment: Identifying Challenges and Opportunities

Evelyn Abagayle Boyd
eboyd2@clemson.edu

A MIXED METHODS ANALYSIS OF INFLUENCES SURROUNDING UNDERGRADUATE SCIENCE RECRUITMENT: IDENTIFYING CHALLENGES AND OPPORTUNITIES

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Engineering and Science Education

By
Evelyn Abagayle Boyd
August 2023

Accepted by:
Dr. Kelly Best Lazar, Committee Chair
Dr. Brian N. Dominy
Dr. Marian S. Kennedy
Dr. Bridget G. Trogden

ABSTRACT

Often prevalent in the sciences, undergraduate research experiences (UREs) are highly beneficial engaged learning experiences in higher education. Though the benefits of UREs are well established, there is little information about which students participate in these experiences and the pathways they take to become engaged in them. This study consists of three works which surround the overarching research question: *How do undergraduate science students initialize research experiences such that departments and institutions can improve access to these high-impact practices?* Utilizing the intersection of the theories of Science Capital and Social Cognitive Career Theory, this study provides insight and recommendations into ways in which science departments and their respective institutions can improve the equity of access to UREs.

The first section of this sequential explanatory mixed methods study analyzes data from publicly available datasets and a multi-institutional survey. This section analyzed participation rates in undergraduate research across demographic groups and the effect of literature-identified influences on their participation. The second section applies topological data analysis to quantitative survey responses to identify influences common between groups of students that responded with varying numbers of opportunities impacting their undergraduate research participation. Based on the populations that answered the survey, an opportunity presented itself to study an understudied population in the literature of individuals with concealable identities and the intersectionality of the influences on their participation in undergraduate research. The third portion describes the experiences of ten women and the effect of their concealable identities on their interactions with undergraduate research.

This study provides a novel approach to considerations of entry into UREs and, by doing so, expands upon the multidimensional data methodologies available in discipline-based education research. The results of this study demonstrate common opportunities and barriers to participation across student communities. Examples of these influences include the benefit of faculty interaction and the importance for positive communication about research experiences and the pathways to entry available to students. Results also provide insight into the ways students' identities influence their experiences and highlight the importance of targeted approaches to meet specific student needs.

DEDICATION

"Eyes set on what I want ahead, how many miles, pbj piles, tears, and smiles did accrue, I'd be lying if I said I knew, but the vision became the view."

-Alexi Pappas

To all those who have shared miles, pbj piles, tears, and smiles with me along the way to and through grad school. Thanks for always feeding my mind, body, and soul; it certainly takes a village! Now on to the next adventure.

ACKNOWLEDGMENTS

I would like to thank my Ph.D. advisor, Dr. Kelly Best Lazar, for her unwavering support throughout the grad school process and for ensuring that I leave with two Clemson degrees, which is one more than I planned on when I came!

To the rest of my committee, Dr. Brian Dominy, Dr. Molly Kennedy, and Dr. Bridget Trogden thank you for your mentorship and advice. A special thank you to Bridget, who has helped me through my whole higher ed career. Go Bears!

Thanks to the other members of the People and Places Lab – Cole Bowman, Shannon Conner, Gavin Gleasman, and Megan Lapkoff. Also, to the many ESED students and grads, Dr. Catherine Kenyon, Dr. Baker Martin, Sarah Otterbeck, and Tyler Sullivan to name a few, for all of their research help and for making grad school way more fun.

I would also like to acknowledge the help I received from Dr. Allison Godwin and Dr. Stephen Moysey in setting up and utilizing TDA. I also want to thank The Geological Society of America Graduate Student Research Grants for financial support for this study, and the participants, particularly the ten interview participants, for their time and stories, which made this work possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

High-impact educational practices (HIPs; Kuh, 2008) are efforts in higher education that are proven to help increase student retention and engagement in students from varying backgrounds. Undergraduate research experiences (UREs) have been listed as one of these HIPs, as the positive outcomes for students have been well documented. But, despite the growing amount of evidence of the effects these experiences have on students, few studies have analyzed the recruitment practices that bring students into UREs (Haeger et al., 2021; Morales et al., 2017; Thompson & Jensen-Ryan, 2018). Additionally, there is little information available regarding the students that are participating in UREs, and there are a growing number of calls to improve the equity of access to undergraduate research experiences (Bangera & Brownell, 2014; Haeger et al., 2021; Hewlett, 2018; Krim et al., 2019; Lopatto, 2009; Vincent-Ruz et al., 2018).

Representation is an important piece of student success. Amongst the benefits of proper representation are increases in student sense of belonging, psychosocial wellbeing, and contributions to recruitment and retention efforts (Fogg-Rogers & Hobbs, 2019; Lewis et al., 2016; Robinson, 2018). Representation across demographic groups within the sciences in the United States (US) does not match national representation of the population in general. Biennially, the National Science Foundation (NSF), in collaboration with the National Center for Science and Engineering Statistics (NCSES), issues a report on the state of representation regarding race/ethnicity and disability status across many areas of science (and engineering). In 2019, The US science workforce was comprised of 15% individuals that identified as belonging to traditionally marginalized racial/ethnic groups (compared to 34% of the US population) and 9% of scientists identified as disabled (compared to 13% of the overall US population; Hamrick, 2022). Additionally, issues of equity and representation are not limited to race or disability status; non-binary gender and sexuality are among the demographic factors that should also be considered but are currently not considered (AERA, 2022).

The majority of higher education equity efforts involve improving rates at which students progress to graduation, but increasing persistence alone is not sufficient for actualization of career goals (Bangera & Brownell, 2014). The American Association of Colleges and Universities (AAC&U) suggests that “increasing access to and participation in HIPs” and “increasing student awareness and understanding of the value of HIPs for workforce preparation and engaged citizenship” are two of the best ways to promote equity in higher education (McNair & Veras, 2017). In the sciences, undergraduate research has been identified as so vital that professional societies such as the American Chemical Society (ACS, 2015) and American Physics Society (APS, 2008) have called for the incorporation of UREs across their disciplines. Three of the four participating institutions in this study require undergraduate research in the degree pathways for one or more of their science majors. This study analyzes entry pathways for science majors into a HIP, undergraduate research experiences and can serve as a first step toward the intentional improvements in equity efforts encouraged by AAC&U.

Literature Review

Previous Work Surrounding Opportunities, Barriers, and Recruitment Practices in UREs

The majority of studies surrounding undergraduate research focus on assessment or outcome-based approaches. However, there are several potential opportunities, barriers, and recruitment practices for URE participation that have been discussed in the literature. Haeger et al. (2021) carried out a study on opportunities and barriers to research participation in which they included undergraduate students, faculty members, and academic advisors as their participants. Many of their identified barriers fall into the categories of institutional barriers (e.g., finding a mentor, fitting it into one's curriculum), other commitments (e.g., having to use that time for an outside job, familial commitments), and affective concerns (e.g., lack of sense of belonging). Bangera and Brownell (2014) expressed many of these opportunities and barriers but also issues of student awareness regarding URE opportunities, how to pursue them, and the benefits of UREs. Discussions of the effects of COVID-19 on student participation in engaged learning events and general accessibility concerns for students have also been discussed as barriers to research participation (Bingham, 2021; Gin et al., 2022; Grineski et al., 2022; Pierszalowski et al., 2021).

Haeger et al. (2021) also found that the most common method for a student to get involved in research was through professors. Students described frustration because many faculty have differing expectations about how students should become involved in research (e.g., some faculty display research information as an advertisement to undergraduate students while others display it as general information). Additionally, faculty acknowledged that much of their choice of who they ask to join their labs is inherently biased as they are more likely to reach out to students who participate more in class or that they have built a relationship with over multiple semesters. The idea of faculty being biased when asking students to join their labs is further supported by the work of Bangera and Brownell (2014).

Frameworks focusing on student interest (e.g., Harackiewicz et al., 2008) are largely focused on major selection rather than participation in engaged learning activities. Few studies analyze the connection of student interest in science to URE participation, and among those, it is most often within discussions of persisting in research participation rather than initial entry (Canaria et al., 2012; Cooper, Gin, et al., 2019).

Other studies have analyzed opportunities and barriers to UREs at a programmatic level. Their identified positive and negative influences include institutional financial resources, faculty availability, limited student preparation, faculty support with curriculum development, and department/administrative support of UREs (Frantz et al., 2017; Hewlett, 2018; Kirkpatrick et al., 2019; Lopatto et al., 2014; Morales et al., 2017). Some of these identified influences conflict with established recommendations for when students should participate. For example, faculty concerns that students may not be academically prepared to participate in research directly opposes suggestions that to receive the maximum benefit of research experiences, students should begin participating in research as early as possible (Sandquist et al., 2019; Vincent-Ruz et al., 2018)). These can all serve as either opportunities or barriers (or both) to research participation depending on implementation.

Not every student will want or need a URE, but by examining these recruitment practices we can create a more equitable space for all students.

Improving Equity in UREs

A vast body of research has demonstrated the positive effects participation in UREs, including increased student interest in the discipline, enhanced career preparation, clarification of future goals, improved technical and professional skills, improved confidence and science self-efficacy, and increased sense of belonging in science settings (Harsh et al., 2011; Chemers et al., 2011; Carpi et al., 2017; Thompson & Jensen-Ryan, 2018). Positive outcomes of URE participation are well established for all student populations and have been found to be especially beneficial for students who are traditionally underrepresented in STEM, transfer students, first-generation college students (students who are the first in their family to go to college), and students from lower-income families (Carpi et al., 2017; Castillo & Estudillo, 2015; Chemers et al., 2011; Eagan et al., 2013; Haeger et al., 2021). However, these students are less likely to participate in UREs than their peers (NASEM, 2011). Some explanations for their lack of participation could be due in part to job, familial, or course requirements impacting their availability. These students may not know what resources are available to them at their institutions to help them balance engaged learning experiences with their other responsibilities, or, they may not know UREs exist at all or how to access them at their institution (E. Ramirez, 2011). Additionally, little data is available for students with concealable identities (identities that students may choose not to reveal, such as sexuality, disabilities, and gender diversity) because of the methods of collection for national datasets, creating another potential access equity gap (Freeman, 2020; Lillywhite & Wolbring, 2019). There has been a nationwide movement for studies regarding equity improvements to undergraduate research entry (Gentile et al., 2017). By gaining a better understanding of what students are participating in these HIPs, we can also learn more about those that may be missing and focus UREs to better meet their needs.

One frequently suggested method of increasing equity in research is to create course-based undergraduate research experiences, sometimes called CUREs (Ballen et al., 2018; Bangera & Brownell, 2014; Kirkpatrick et al., 2019; Krim et al., 2019; Sandquist et al., 2019; Szteinberg, 2012). All CUREs are course-based but not all course-based research experiences are classified as CUREs; for simplicity in the discussion below, the CURE acronym will be used to mean all course-based UREs (Auchincloss et al., 2014). Course-based research is a beneficial way to provide research opportunities for a greater number of students than other forms of URE. This is because, though limited by lab regulations and space, one can interact with more students in course-based research setting as opposed to an apprenticeship-style research experience. Other suggestions for expanding the availability of UREs include using mentors from industry as opposed to solely university personnel (Frantz et al., 2017) and moving research labs online, which simultaneously increases accessibility and decreases cost (Kirkpatrick et al., 2019; Lopatto et al., 2014). These all help improve the number of students that are able to participate in research but do little to increase student awareness of available opportunities. Additionally, without proper resources and training these methods create unfair burdens on the mentors designing the CUREs or virtual research experiences (Lopatto et al., 2014). This study seeks to improve equity from the standpoint that by helping institutions increase identified opportunities and decrease barriers to research participation, institutions will be able to improve the equity of the availability of their UREs.

Intersectionality in Higher Education

An additional important equity consideration is intersectionality. Intersectionality is an emerging idea that addresses the compounding effect of multiple marginalizing identities has on individuals. It was defined by Crenshaw (1989) in a study considering the intersection of race and gender for Black women in politics. Critical theorist Patricia Collins has further developed the idea and notes that while intersectionality helps shed light on contemporary social issues, it is not yet developed enough to be considered a critical social theory because practitioners have not fully defined its assumptions, epistemologies, and methods (Collins, 2019). Intersectionality originates with race/ethnicity and gender; however, these ideas resonate within this study's purpose of considering individual pathways to undergrad research participation. This is done particularly when considering the experiences of women with concealable identities in Chapter Four.

Theoretical Framework

Science Capital

Students' pathways into undergraduate research experiences are influenced by both individual and institutional factors. This study lies at the intersection of the individual focus of Science Capital (Archer et al., 2015) and the institutional focus of Social Cognitive Career Theory (SCCT; Lent et al., 1994). Capital can be generally defined as assets that individuals "carry" with them. If you picture students in a class with backpacks, they may carry backpacks of different styles, brands, or sizes, they may have been purchased or gifted to them from different places or people, and they may be filled with different resources, but they all serve the purpose of helping the student be prepared for class. Likewise, sociological capital are the "things" that we "carry" with us as we are interacting with the world around us. Elements of sociological capital can include, but are not limited to, individual characteristics (e.g., demographic factors), past mentor relationships, and access to resources (Jones et al., 2020).

Bourdieu (1986) describes four such forms of capital applied to academic settings in his works, economic, symbolic, cultural, and social. These four types are described below and in Table 1.1. Economic capital encompasses tangible assets and financial-related attributes (e.g., the difference in pay an individual receives when they earn an additional degree). Symbolic capital is qualifications, honors, and reputation; this form of capital usually changes, or even loses, its value when taken into other environments (e.g., an honors student transfers institutions). Cultural capital assets are those that contribute to knowledge, tastes, and cultural dispositions (e.g., a student wanting to pursue research to help their community). Lastly, social capital is family, networks, and relationships that contribute to engagement in academic activities (e.g., a friend informing a student of a research opportunity). The capital forms are designed to help consider situations around us, and many assets and situations are a part of more than one form of capital at the same time (Bourdieu, 1986).

An important weakness in Bourdieu's conceptualization of cultural capital is that it carries a deficit approach and views differences in culture as a hierarchical structure. This

Table 1.1. Definitions of forms of capital. Definitions based on descriptions in Bourdieu (1986) and DeWitt et al. (2016).		
	Capital Type	Definition
Bourdieu (1986) Forms of Capital	Economic Symbolic Cultural Social	Financial and other tangible assets Qualifications, honors, and reputation Knowledge, tastes and cultural dispositions Family, networks and relationships
DeWitt et al. (2016) Science Capital	How you think Who you know What you know What you do	How an individual understands science Scientific social capital Individual values in science contexts How individuals talk about science and participation in science-related activities

often leads to the assumption that individuals from some cultures "lack" the social and cultural capital required for social mobility. In response to this, Yosso (2005) developed the theory of Community Cultural Wealth (CCW). Community Cultural Wealth is related to Bourdieu's theory and involves all four types of capital. However, the focus is on cultural capital with an asset-based mindset. Instead of what do certain cultures lack, the focus is on what each culture teaches the members of their communities, and what other cultures have to learn from them.

Another weakness is that the educational experiences Bourdieu described primarily apply to students' interactions with the arts and humanities. Science-related events such as labs, interest in nature, and science camps or museums are not fully captured in the conceptualization (Archer et al., 2015). In response to this, Archer et al. (2015) developed Science Capital as a reframing of several forms of sociological capital to explicitly describe how individuals navigate to and through science experiences. Science Capital is not a form of capital on its own, rather it is encouragement to researchers to rethink forms of capital previously described in science-focused contexts to address inequalities in science participation (Archer et al., 2015). DeWitt et al. (2016) further described four major elements of Science Capital as: *How you think*, *Who you know*, *What you know*, and *What you do* (Table 1.1). *How you think* science capital is described as how an individual understands science (e.g., knowing how to use evidence to make an argument). *Who you know* science capital is related to science-related social capital (e.g., interactions with instructors that encourages one to pursue science). *What you know* science capital is related to what individuals value in science contexts (e.g., pursuing science related qualifications in search of a job). Lastly, *What you do* science capital describes how individuals talk about science and their participation in science related activities (e.g., reading books about science outside of class assignments; DeWitt et al., 2016). Combinations of these four elements describe how individuals interact in science contexts.

Social Cognitive Career Theory

Social Cognitive Career Theory (SCCT; Lent et al., 1994) is frequently used in studies describing the outcomes of undergraduate research participation (Carpi et al., 2017). Though originally designed to describe career choice, SCCT has been used to understand decision making in a variety of contexts. This theory describes how self-efficacy, an individual's self-confidence in their ability to accomplish a given task, and outcome expectations, what an individual believes they will come away from a task having accomplished or gained, are influenced by demographic and background factors, and play a role in an individual's career choice. Contextual factors, those pertaining to individual's backgrounds, and environmental factors, those pertaining to their current surroundings. These characteristics can be distal, meaning they do not have an immediate effect on individuals (e.g., visiting science museums as a child having an influence on participation in UREs), or proximal, having a more immediate effect on individuals (e.g., pursuing undergraduate research because a roommate recommended it). Environmental factors can also provide affordances or barriers. Because of the consideration of contextual and environmental factors, SCCT allows further consideration of the institutional rules and norms that may be in effect as students enter UREs in addition to the individual capital they are bringing with them. When considering SCCT in undergraduate research settings, it is helpful for understanding how a student's future goals, such as their career aspirations, may affect their research participation.

The Intersection

This study lies at the intersection of the individual focus of Science Capital and the institutional focus of SCCT. Jones et al. (2020) developed a model which displays the intersection of Science Capital and SCCT and served as the starting theoretical framework for this study (a reproduction in Figure 1.1). Their model was the starting theoretical framework for this study and allowed consideration of potential influencing factors and their relationship to entry into UREs.

Research Questions

The overarching research question of this study is:

How do undergraduate science students initialize research experiences such that departments and institutions can improve access to these high-impact practices?

This question will be addressed by analyzing the following research questions:

1) Among science students, who is and is not participating in undergraduate research experiences?
2) What influencing factors are identified by science students as impactful for participation or non-participation in undergraduate research experiences?
3) In what ways do influencing factors differ across model identified groups?
4) How do science students describe: a) proximal support and/or barriers they experienced and b) self-efficacy and outcome expectations related to participating in undergraduate research?

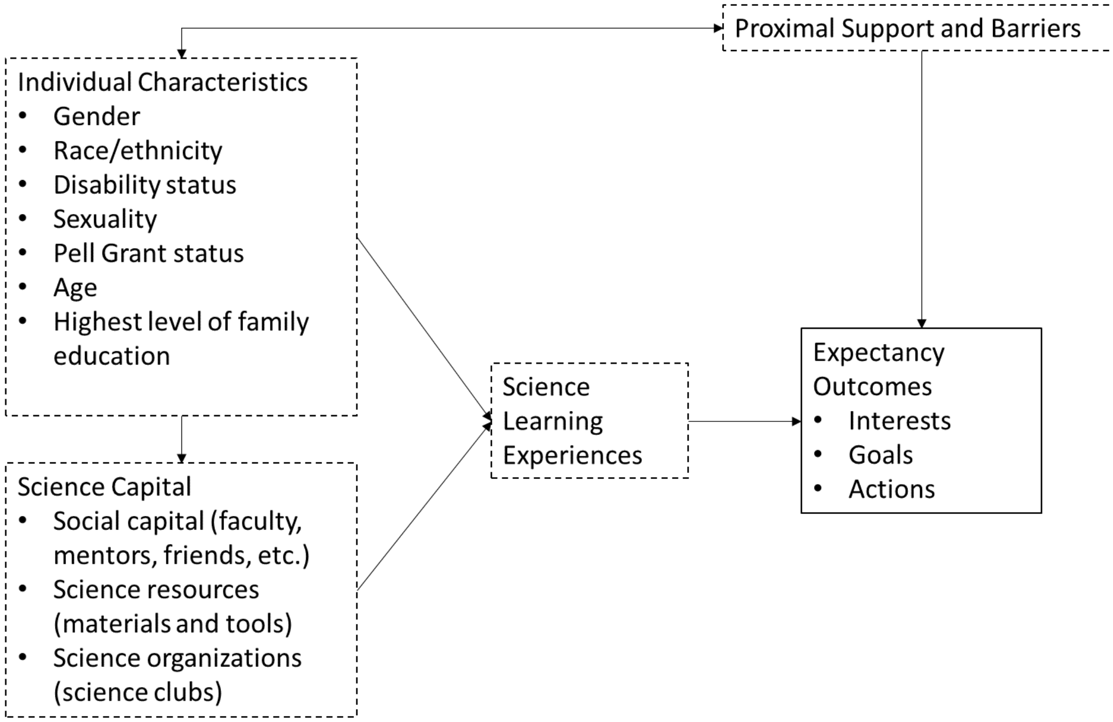


Figure 1.1: Adaptation of Jones model (2020). Items focused primarily on Science Capital are outlined with dotted lines, items focused primarily on Social Cognitive Career Theory outlined in solid lines.

Research Design

A survey that addressed students' influencing factors surrounding their participation in undergraduate research experiences was deployed at four public R1 institutions in the Southeastern United States (Appendix A & B). Descriptive factors of each institution are displayed in Table 1.2.

This study was approved exempt from review by University Alpha's Institutional Review Board (IRB). Additionally, permission for research was obtained from each participating institution's IRB and surveys distributed following their guidelines for research administered by external personnel. The selection criteria for participation in the study was to be at least 18 years old, a student at one of the four participating institutions, and a science major. The National Center for Education Statistics Classification of Instructional Programs (NCES CIP; *CIP User Site*, n.d.) was utilized to determine what majors would be included in the study as science majors. This classification is a taxonomic scheme designed to support the accurate tracking and reporting of fields of study for institutions of higher education. The CIP codes Biological and Biomedical Sciences (26), and Physical Sciences (40) are included as they are the CIP codes that are entirely science. Participants who identified more than one major needed only one of their majors to fall within the established CIP codes to be classified as a science major for participation in the study. Appendix C includes a full listing of the majors included within these CIP codes.

Surveys were administered according to each institution's distribution protocol. Eight hundred thirty-three completed responses met the selection criteria and were included

Table 1.2. Participating institution descriptive factors

University Alpha	University Beta	University Gamma	University Delta
Large	Large	Large	Large
Public	Public	Public	Public
R1	R1	R1	R1
Land Grant	Land Grant	Non-Land Grant	Non-Land Grant
Predominantly White Institution	Predominantly White Institution	Predominantly White Institution	Hispanic Serving Institution

as the population of the study. Ten interview participants were intentionally selected based on survey responses. All interview participants were selected from the same institution to control for institutional differences. All interview participants, and fifty randomly selected survey participants, received a $20 incentive card for their participation.

Research Question 1 (RQ1) of was addressed using solely quantitative data and is discussed in Chapter Two (Fig 1.2). For this portion of the study, demographic information from 833 survey responses was compared to publicly available data (Greathouse et al., 2018; Hamrick, 2022; U.S. Department of Education, n.d.). This comparison aids in identification of potential over or underrepresentation of undergraduate researchers at these institutions as compared to national trends in representation. By identifying who is participating in research, recruitment efforts can be altered to target groups that are underrepresented in research. The remaining research questions are addressed using a sequential explanatory (QUANT→qual) mixed methods approach (Creswell & Plano-Clark, 2007). Research Question 2 (RQ2) is addressed by analysis of both the quantitative and qualitative portions of the survey sent out to students at participating institutions in Chapter Two. Results from the survey were used to perform Topological Data Analysis (TDA), a method that allows for clustering of complex data with more nuance than traditional clustering methods (Doyle, 2017). The TDA results address RQ3 and are described in Chapter Three. The groupings formed by the TDA led to purposeful selection of ten participants for qualitative interviews (Fig. 1.2). Survey results revealed the opportunity to further explore an area underexplored in the literature, individuals with concealable identities and their participation in UREs. As such, concealable identities revealed in the demographic section of the survey were considered during the purposeful interview selection. Interview data were analyzed and results are presented to address RQ4 in Chapter Four. Four main data sources were used in this dissertation study: a survey with both qualitative and quantitative data, the quantitative TDA mappings formed by the survey responses, quantitative publicly available datasets, and qualitative interviews. A mixing diagram for the study including these four data sources is presented in Figure 1.2.

Legitimation

As a mixed methods study, it is important to consider the legitimation, or strength of mixing throughout the study design, data collection, analysis, and interpretive phases. (Onwuegbuzie et al., 2011) describe several legitimation types, all of which were considered throughout this study. Six of these legitimation types, sample integration, inside-outside, weakness minimization, sequential, conversion, and multiple validities, were implemented and described in Appendix D. Creamer's (2019) Mixed Method Evaluation Rubric (MMER) was also considered to ensure that the mixing occurred sufficiently throughout the study.

Quality Considerations

All research studies should assess their methods to acknowledge the strengths and threats to the quality of their research methods and to limit bias. The validity and reliability of all methods must be considered throughout the study. Walther et al. (2013) describes a framework for ensuring six kinds of validation which are considered both during the data creation and data handling stages of the research. These six validation types are theoretical

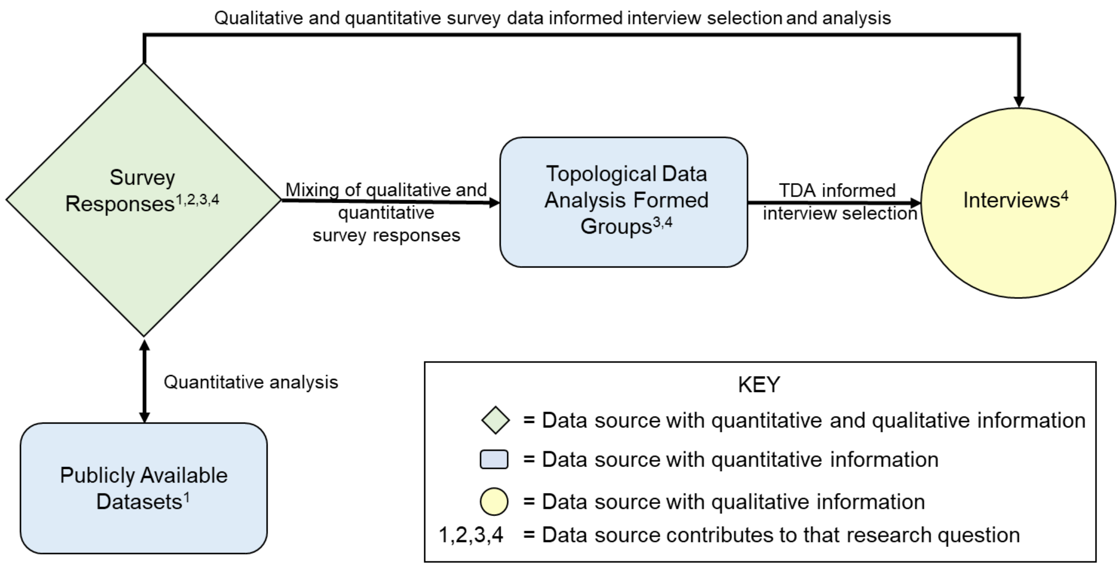


Figure 1.2: Research design and mixing of methods.

(how well concepts of the theoretical frameworks match the study), procedural (how the study matches theory and reality), communicative (how the study is presented to the intended audience), pragmatic (how well the study withstands exposure to reality), ethical (consideration of research ethics in the study), and process reliability (insurance that results of the study that are due to random chance are limited). All studies will have some threats to these validation types, and some research decisions are both opportunities and threats to the same or multiple types of these validities simultaneously. Descriptions of all six, as well as how research decisions in the planned data creation and handling stages pose opportunities or threats to each, are detailed in Appendix E.

CHAPTER TWO

UNDERGRADUATE RESEARCH SCIENCE CAPITAL: MEASURING CAPACITY TO ENGAGE IN RESEARCH

This chapter is being prepared for submission to a scholarly journal. Individual survey responses are available upon request. The following modifications were made to include the article in this dissertation: 1) tables and figures were renumbered, 2) all references were moved to the full list at the end of the document.

Abstract

Undergraduate research has been identified as a high-impact educational practice. However, despite the body of evidence on the outcomes of undergraduate research, few studies have focused on the influences students face regarding participation. Influenced by Science Capital and Social Cognitive Career Theory, a survey comprised of potential factors influencing undergraduate science research participation was disseminated to science majors at four R1 institutions in the Southeastern United States. Free response questions allowed for the addition of missed influences. Participation rates across several demographic factors and effect of participation influences were analysed; significant differences were found for two demographic factors and fourteen participation influences. The Undergraduate Research Science Capital (URSC) scale was developed that can help science departments and their respective institutions develop equitable entry into these engaged learning practices.

Introduction

High-impact educational practices (HIPs; Kuh, 2008) are efforts in higher education that are proven to help increase student retention and engagement in students from varying backgrounds. Undergraduate research experiences (UREs) have been listed as one of these HIPs and the positive outcomes for students have been well documented. Undergraduate research experiences are well established in the literature as highly beneficial academic experiences for all students. In the sciences, UREs have been identified as so vital that professional societies such as the American Chemical Society (ACS, 2015) and American Physics Society (APS, 2008) have called for the incorporation of UREs across their disciplines. Despite the growing evidence of the impact these experiences have on students, few studies have analyzed the recruitment practices that bring students into UREs (Haeger et al., 2021; Morales et al., 2017; Thompson & Jensen-Ryan, 2018; Bingham, 2021). Additionally, there is little information available about representation of underrepresented groups within URE spaces, and there are a growing number of calls to improve equity of these opportunities (Haeger et al., 2021; Vincent-Ruz et al., 2018; Krim et al., 2019; Hewlett, 2018; Lopatto, 2009; Bangera & Brownell, 2014).

Literature Review

Types of UREs

There is not a single understanding of what constitutes a URE, however most lead to similar student outcomes. The most common URE in the sciences is apprenticeship-style research in which a student is paired with a research mentor (usually a faculty member or graduate student) and works with them in their research space (Lopatto, 2009). These experiences are often at least a semester in length with students participating in research a few hours a week. Another common form of UREs are summer research experiences. Though these are also often apprentice-style in nature, they are usually full time and typically last 8-12 weeks. Course-based undergraduate research experiences, sometimes called CUREs, are another common format for UREs. All CUREs are course-based but not all course-based research experiences are classified as CUREs; for simplicity the CURE acronym will be used to mean all course-based UREs that are incorporated into the curriculum of a class (Auchincloss et al., 2014). By nature of being course-based, CUREs are usually one semester, but can sometimes be carried out in multiple semesters as a sequence of courses. Students will always receive course credit for these UREs; however, they are not always clearly labeled as research experiences.

Those three forms of research experiences described above are likely not an exhaustive list of the many forms of research experiences students may encounter but are the three most common in the sciences. Within a specific URE each individual's experience will likely be different; these individual experiences are further differentiated by institutional differences that impact available opportunities, supports (e.g., major requirements, offices of undergraduate research), and requirements for research participation.

Outcomes of UREs

Since the Boyer Commission on Educating Undergraduates in the Research University's original report in 1998 (Boyer Commission on Educating Undergraduates in the Research University, 1998), many institutions have increased focus on improving student engagement, including through undergraduate research (Katkin, 2003). While undergraduate research is beneficial to students in all fields, it is arguably most prevalent in the sciences (Haeger et al., 2021; Kuh, 2008). A vast body of research has been generated demonstrating the positive outcomes of UREs for science students, including increased student interest in the discipline, increased persistence to graduation with a science degree, enhanced career preparation, clarification of future goals, improved technical and professional skills, critical thinking gains, improved science literacy, improved confidence and self-efficacy in science abilities, and increased sense of belonging in the sciences. Relevant work related to each of these impacts is shown in Table 2.1.

These outcomes have been shown to be amplified for students the longer (or the earlier) in their academic careers that they begin participation in their URE (Russell et al., 2007; Sandquist et al., 2019; Vincent-Ruz et al., 2018). Additionally, benefits are increased for students that are traditionally underrepresented in STEM (Carpi et al., 2017; Castillo & Estudillo, 2015; Chemers et al., 2011; Eagan et al., 2013; Haeger et al., 2021), transfer students (Haeger et al., 2021), and those that have already struggled academically (Kirkpatrick et al., 2019). Despite this information being available to departments, many UREs are not reaching these students who would benefit the most. These students have been found to participate in research experiences at lower rates and for shorter timespans within their college careers (Haeger et al., 2021).

One frequently suggested method of increasing equity of research experiences is to create CUREs (Ballen et al., 2018; Bangera & Brownell, 2014; Kirkpatrick et al., 2019; Krim et al., 2019; Sandquist et al., 2019; Szteinberg, 2012). Course-based research is a

Table 2.1. Outcomes of undergraduate research experiences

Outcome	References
Increased interest in the discipline	Seymour et al., 2004; Russell et al., 2007; Harsh et al., 2011; Kirkpatrick et al., 2019
Increased persistence to graduation within a science major	Auchincloss et al., 2014; Chemers et al., 2011; Harsh et al., 2011; Hewlett, 2018; Sandquist et al., 2019; Seymour et al., 2004; Vincent-Ruz et al., 2018
Enhanced career preparation	Auchincloss et al., 2014; Frantz et al., 2017; Harsh et al., 2011; Krim et al., 2019; Seymour et al., 2004; Vincent-Ruz et al., 2018
Clarification of future goals	Carpi et al., 2017; Chemers et al., 2011; Eagan et al., 2013; Harsh et al., 2011; Hewlett, 2018; Russell et al., 2007; Seymour et al., 2004; Thompson & Jensen-Ryan, 2018
Improved science technical skills	Auchincloss et al., 2014; Castillo & Estudillo, 2015; Eagan et al., 2013; Harsh et al., 2011; Seymour et al., 2004; Szteinberg, 2012
Improved professional skills	Auchincloss et al., 2014; Castillo & Estudillo, 2015; Seymour et al., 2004; Thompson & Jensen-Ryan, 2018
Critical thinking gains	Ballen et al., 2018; Castillo & Estudillo, 2015; Seymour et al., 2004; Vincent-Ruz et al., 2018
Improved science literacy	Auchincloss et al., 2014; Castillo & Estudillo, 2015; Krim et al., 2019; Seymour et al., 2004; Szteinberg, 2012
Improved confidence and self-efficacy in science skills	Auchincloss et al., 2014; Ballen et al., 2018; Carpi et al., 2017; Chemers et al., 2011; Eagan et al., 2013; Frantz et al., 2017; Harsh et al., 2011; Seymour et al., 2004; Szteinberg, 2012
Increased sense of belonging in the sciences	Chemers et al., 2011; Sandquist et al., 2019; Seymour et al., 2004

beneficial way to provide research opportunities for a greater number of students than other forms of UREs. Though limited by lab regulations and space, it is possible to fit more students in a CURE lab than can be adequately mentored in an apprenticeship-style research experience. Other suggestions for expanding the availability of UREs include using mentors from industry as opposed to solely university personnel (Frantz et al., 2017) and moving research labs online, which simultaneously increases accessibility and decreases cost (Kirkpatrick et al., 2019; Lopatto et al., 2014). These all help improve the number of students that are able to participate in research but do little to increase student awareness of available opportunities.

Opportunities, Barriers, and Recruitment Practices in UREs

Haeger et al. (2021) investigated opportunities and barriers to research participation in which they included undergraduate students, faculty members, and academic advisors at a midsized institution. Many of their identified barriers fall into the categories of institutional barriers (e.g., finding a mentor, fitting it into one's curriculum), other commitments (e.g., having to use that time for an outside job, familial commitments), and affective concerns (e.g., lack of sense of belonging). Bangera and Brownwell (2014) described many of these opportunities and barriers but also issues of student awareness regarding URE opportunities, how to pursue them, and the benefits of UREs.

Theoretical Framework

This study lies at the intersection of the individual focus of Science Capital (Archer et al., 2015) and the institutional focus of Social Cognitive Career Theory (SCCT; Lent et al., 1994). Capital can be generally defined as assets that individuals "carry" with them. If you picture students in a class with backpacks, they may carry backpacks of different styles, brands, or sizes, they may have been purchased or gifted to them from different places or people, and they may be filled with different resources, but they all serve the purpose of helping the student be prepared for class. Likewise, sociological capital are the “things” that we “carry” with us as we are interacting with the world around us. Elements of sociological capital can include, but are not limited to, individual characteristics (e.g., demographic factors), past mentor relationships, and access to resources (Jones et al., 2020).

In academic settings, discussions of capital are traditionally primarily applied to arts and humanities settings; in response to this, Archer et al., (2015) developed Science Capital as a reframing of several forms of sociological capital to explicitly describe how individuals become involved in science experiences and the differing paths they may take to get there. They further described four major elements of Science Capital as: *What you know, Who you know, How you think*, and *What you do* (DeWitt et al., 2016). Combinations of these four elements describe how individuals interact in science contexts.

Social Cognitive Career Theory (SCCT) is frequently used in studies describing the outcomes of undergraduate research participation (Carpi et al., 2017). Though originally designed to describe career choice, SCCT has been used to understand decision making in a variety of contexts. This theory describes how self-efficacy (an individual’s self-confidence in their ability to accomplish a given task) and outcome expectations (what an individual believes they will come away from a task having accomplished or gained), are influenced by demographic and background factors and play a role in an individual’s career choice (Lent et al., 1994). When considering SCCT in undergraduate research settings, it is helpful for understanding how a student's future goals, such as their career aspirations, may affect their research participation.

Jones et al. (2020) developed a model which displays the intersection of Science Capital and SCCT. Their model (Fig. 1.1) was the starting theoretical framework for this study and allowed the authors to consider potential influencing factors and their relationship to UREs. This has led to the following research questions (RQs):

1) Among science students, who is and is not participating in undergraduate research experiences?
2) What influencing factors are identified by science students as impactful for participation or non-participation in undergraduate research experiences?

Methodology

Population

Surveys were administered at four participating institutions to measure students' undergraduate research-related science capital. Institutions were selected via a random number generator from a list of public Carnegie Doctoral Universities with Very High Research Activity (Carnegie Classifications, n.d.; "R1") in the Southeastern United States. Public R1s specifically were studied due to the difference in funding styles between public and private, potentially impacting the URE recruitment strategies. These Very High Research Activity institutions tend to be large and have a higher student to faculty ratio that could influence student recruitment into UREs. These institutions also by definition of being an R1 have large amounts of research funding and opportunities taking place.

However, despite the research activity level that is happening at these institutions, there is no guarantee that the undergraduates are getting the full benefit of the opportunities available. Since 2015, the Council of Undergraduate Research (CUR), a national organization developed to promote undergraduate research participation, has issued the Campus-Wide Award for Undergraduate Research Accomplishments (AURA) to institutions that exemplify CUR's characteristics of excellence in undergraduate research. To date, twenty institutions have received this award, but only three of them (15.00%) are public R1 institutions (CUR AURA, n.d.). This indicates potential room for improvements in the allocation of resources for UREs at these institutions. Identifying this space for improvement would allow these public R1 institutions to better their efforts in equitable recruitment into UREs. Participating institutions include one Hispanic serving institution (HSI) and three predominantly white institutions (PWI). Half of the institutions are land-grant institutions.

This study is approved exempt from review by the supporting university's Institutional Review Board (IRB). Each participating institution's IRB approved of participation and dissemination followed individual institutional guidelines. To reach a broad spectrum of science majors regardless of class standing and best control for potential sampling bias, surveys were disseminated via email. This was performed either by the research team or by faculty at the institution depending on institutional guidelines for external researchers. Where applicable, department chairs identified potential instructors to assist with survey dissemination, science related clubs, and listservs were utilized, and flyers were posted in locations near where science courses meet across campuses. Upon survey completion, fifty participants were randomly selected to receive a $20 incentive card for their participation, additionally individual instructors were permitted to offer extra credit for survey completion at their discretion.

In this study, science major was defined as those falling within the Classification of Instructional Programs (CIP) codes for physical and life sciences (CIP User Site, n.d.; Appendix C). Participants who identified more than one major needed only one of their majors to fall within the established CIP codes to be classified as a science major for participation in the study. Further discussion of CIP codes is available in Chapter One. Student transfer status, gender, race/ethnicity, GPA, Pell Grant status, and time since matriculation were also self-reported in the survey.

Description of Survey

To measure undergraduate students' research-related science capital, a survey with both Likert-style and free response questions was administered to science majors at all four participating institutions. Participants were asked about the number and type of research experiences they had participated in. Research experiences were divided into four categories: lab for credit and/or pay, course-based, summer, and volunteer to further distinguish between some of the research factors that may be present for each experience. All survey instruments explained to the students that not all research opportunities are in a lab and any mentored research experience would be applicable to the study (Appendix A). Additionally, students that identified not having participated in research experiences were asked to identify any reasons for their lack of participation.

To develop a scale for measuring science capital and its influence on URE participation, twenty-five influencing factors were included on the survey for participants to rate on a Likert-type scale from 1 (extremely negative impact) – 7 (extremely positive impact), with a not-applicable option. Twenty-three of the twenty-five items were identified from the literature (Szteinberg, 2012; Haeger et al., 2021); the influence of COVID-19 and travel to/from research sites were identified as potential factors that were not previously found in literature and added before survey dissemination. Factors were presented on the survey neutrally so as to not steer the respondents towards opportunity or barrier (e.g., “Work-Jobs outside of your research responsibilities”). The survey was reviewed by educational researchers and undergraduate students prior to dissemination with a focus on pragmatic and communicative validation (Walther et al., 2013). In addition to the Likert-style factors, free response questions were included to allow students to include additional factors that may have been missed by the survey and allow for further elaboration. Qualitative analysis of free response questions responses was magnitude coded by two researchers and found to have a Cohen’s Kappa value of 0.77 (Landis & Koch, 1977).

Population and Survey Analysis

Due to varying availability of population demographic data for comparison, publicly available data were collected from three different databases. National proportions within science majors were collected for student disability status and students that are members of the lesbian, gay, bisexual, transgender, and queer plus (LGBTQ+) community. Disability status is available through the National Science Foundation (NSF) and National

Center for Science and Engineering Statistics (NCSES) database "Women, Minorities, and Persons with Disabilities in Science and Engineering" (Hamrick, 2022). Sexuality status data were calculated from Greathouse et al.'s (2018) report on national survey data. Institutional data as an average of the four participating institutions was collected as the population data wherever possible from the National Center for Education Statistics Integrated Postsecondary Education Data System (NCES IPEDS; NCES, 2022). Data are available at the major level for transfer student status, gender, and race/ethnicity. Pell grant status was available for each of the institutions but not disaggregated by major. Once collected the population proportions were compared to the survey responded proportions using unpaired *t*-tests. Sample means and standard deviations were calculated from the quantitative survey responses and compared as applicable using two proportion z-tests and unpaired *t*-tests to understand the differences between the sample and population proportions and differing groups within the sample.

Free response survey responses were qualitatively coded using magnitude coding by two researchers based on coding methodologies described in Saldaña (2016). Intercoder reliability was checked to ensure qualitative coding reliability following methodologies described in O'Connor and Joffe (2020) and Landis & Koch (1977).

Results

RQ1: Population Demographics

An estimated 12,442 students are science majors across the four participating institutions based on publicly available enrolment statistics. One thousand three hundred ninety-five survey responses were obtained resulting in an overall response rate of approximately 11.21%. After completion and inclusion criteria were applied, 833 responses were included in the study resulting in a response rate of approximately 6.70%. Response rates are approximated due to IRB limitations preventing surveys from being distributed to all science majors at all four institutions and exact enrollment statistics are not publicly available.

Population demographics (P) in comparison to the study population (s) are presented in Table 2.2 to determine differences between the population and the study sample size. The only significant differences between the P and s populations are present in the proportion of students who identified as members of the LGBTQ+ community (z=11.66, p<.001) and Pell Grant recipients (z=3.05, p=.002; Table 2.2). The remaining demographic characteristics (identification of a disability, proportion of transfer students, individuals with traditionally marginalized genders [female, non-binary, and more than one gender selection], and race/ethnicity underrepresented in science [those reported in this study include American Indian or Alaskan Native, Black or African American, Hispanic and/or Latino/a/x, Native Hawaiian or Pacific Islander, and Middle Eastern]) are all statistically similar between the overall population and the study population (Table 2.2).

Of the 833 respondents, 240 (28.81%) had participated in research (R) and 593 (71.19%) had not yet participated (NR). Two demographic characteristics exhibited significant differences between the R and NR groups. Researchers had a significantly larger proportion of students (1) identifying as a member of the LGBTQ+ community (z=4.35, p<.001) and (2) self-reporting having a disability (z= 2.86, p=.004; Table 2.2).

Table 2.2. Study demographic characteristic representation. Population proportions collected from: (Greathouse et al., 2018; Hamrick, 2022; NCES, 2022). Population proportions are collected as National Proportions in Science Majors (‡), Institutional proportions of all majors (^), and Institutional Proportions in science majors (†). Further explanation of gender and race/ethnicity categories described in Population Demographics. Researcher and Non-Researcher proportions are calculated by number of (non-)researchers belonging to category/total number of (non-)researchers. Significant results (p<.05) of the z-test comparing proportions of R and NR groups indicated by an *.

Demographic Characteristic	Population Proportion (%)	Study Proportion (%)	*z* score	*p*-value
Members of the LGBTQ+ community‡	4.87	13.67	-11.66	< .001*
Pell Grant recipients ^	19.43	24.00	3.05	.002*
Disability‡	8.18	8.58	-0.42	.674
Transfer students†	21.43	19.24	1.53	.276
Genders traditionally marginalized in science†	71.42	73.60	-1.39	.226
Race/Ethnicity traditionally marginalized in science†	21.79	19.68	1.47	.197

Demographic Characteristic	Researcher Proportion (%)	Non-Researcher Proportion (%)	*z* score	*p*-value
Members of the LGBTQ+ community	22.03	10.25	4.35	< .001*
Disability	12.92	6.77	2.86	.004*
Pell Grant recipients	27.36	22.57	1.33	.184
Transfer students	20.34	18.79	-0.34	.728
Genders traditionally marginalized in science	73.66	73.60	-0.05	.960
Race/Ethnicity traditionally marginalized in science	16.74	20.87	-1.29	.197

Year in college when responded	Study Proportion (%)
1	51.89
2	19.57
3	14.02
4	10.73
5+	1.52

Approximately half of students were in their first year of college when they completed the survey (51.89%) with the remaining 48% of student proportions decreasing as students continued along their college experience (Table 2.2). This large proportion of first year students may have influenced proportions of research participation as students may not yet have had the opportunity to engage. Table 2.3 contains the breakdown of participants in this study by major. Majority of participants are life science majors (87.27%) with the largest proportion majoring in general biology (67.59%; Table 2.3). This approximately matches national enrollment statistics which suggest 80% of science majors are life science majors with the remaining 20% enrolling in life science majors (Hamrick, 2022).

Scale Development

In order to create a scale as a measure of student undergraduate research related science capital, factor analysis of the Likert-style survey items was performed. An exploratory factor analysis (EFA) with maximum likelihood extraction and direct oblimin rotation was used to validate the Undergraduate Research Science Capital (URSC) scale for our sample resulting in a five-factor structure (Table 2.4). In addition to factor loadings, parallel analysis of the data confirms the inclusion of five factors (Appendix F), and scree plot analysis suggests four to five factors (Appendix G). Eigenvalues greater than one suggest the inclusion of six factors, however in conjunction with all other analyses and theoretical underpinnings of this work, a five-factor structure was deemed to be the most appropriate model for the data. Statistical analysis for factor analysis was carried out in SPSS (IBM, 2021). The Kaiser-Meyer-Olkin (KMO) measure was .889, exceeding the recommended value of .600 (Field, 2013). Reliability analysis resulted in a Cronbach's alpha of .888.

Table 2.3. Distribution of responses by CIP code	
CIP Code Category	Number of responses (% of population; n=833)
Life Sciences – 727 (87.27%)	
General Biology	563 (67.59%)
Microbiology and Immunology	63 (7.56%)
Biochemistry, Biophysics, and Molecular Biology	45 (5.40%)
Genetics	36 (4.32%)
Marine Sciences[1]	3 (0.36%)
Plant or Animal Biology	2 (0.24%)
More than one within Life sciences	12 (1.44%)
Physical Sciences – 100 (12.00%)	
Chemistry	41 (5.14%)
Geological and Earth Sciences/Geosciences	30 (3.76%)
Astronomy and/or Physics	17 (2.13%)
Environmental Sciences	3 (0.38%)
Atmospheric Sciences and Meteorology	2 (0.25%)
Marine Sciences[1]	2 (0.25%)
More than one within Physical Sciences	5 (0.63%)
More than one category - 6 (0.72%)	
More than one major in physical and life science	6 (0.72%)
1- Marine Science CIP code is an interdisciplinary science; coding is based on department.	

Table 2.4. Scale items and their factor loadings. Significant loadings (>0.32) are indicated by * and are shaded gray. Items presented as they were categorized on the original survey instrument. [Research abbreviated items]. Research abbreviated items will continue to be used for the remainder of the chapter. Survey instructions shown in box below; full survey instrument is presented in Appendix A. Likert anchors: NA, (1)Extremely Negative, (2)Very Negative, (3) Negative, (4)Neutral, (5)Positive, (6)Very Positive, (7)Extremely Positive

Survey instructions: The next questions will help us identify influences that could be considered opportunities or barriers to undergraduate research participation. On a scale of (1) Extremely negative impact to (7) Extremely positive impact, how much of an impact did the following things have on your ability to participate in undergraduate research? Please use NA to indicate any that did not have an effect on you.

Subscale (Cronbach's alpha)	How you think (0.925)	What you dream (0.927)	Who you know (0.845)	What you know (0.757)	What you do (0.720)
Survey Items [Research shortened name]					
Survey Category: OUTSIDE RESPONSIBILITIES – Responsibilities that may influence your ability to participate in undergraduate research.					
Family Obligations- Family can be biological or chosen. (e.g., care responsibilities, driving family members places, etc.) **[Family Responsibilities]**	.010	.067	.004	-.020	.618*
Work- Jobs outside of your research responsibilities **[Jobs]**	.004	.047	-.037	.041	.555*
Athletics – School sponsored athletic obligations (NCAA, intramural, club, etc.) **[Athletics]**	.003	-.002	.115	.073	.466*
Religious Obligations **[Religious responsibilities]**	.000	.080	.017	-.023	.561*
Social Obligations – Activities outside of those already mentioned that may influence your ability to participate in undergraduate research (e.g., Greek life, clubs, friends) **[Social Responsibilities]**	-.043	.071	.106	.108	.350*
Survey Category: INFLUENTIAL PEOPLE - Interactions with others that may influence your participation in undergraduate research.					
Professors – Interactions in or outside of class **[Professor Influence]**	.136	.036	.606*	-.126	-.087
Teachers (from K-12) **[K-12 Influence]**	.097	-.035	.475*	-.022	.139

Table 2.4 (Continued)					
Subscale (Cronbach's alpha)	How you think (0.925)	What you dream (0.927)	Who you know (0.845)	What you know (0.757)	What you do (0.720)
Survey Items [Research shortened name]					
Survey Category: INFLUENTIAL PEOPLE - Interactions with others that may influence your participation in undergraduate research.					
Academic Advisors - Interactions in or outside of official advising time **[Academic Advisor Influence]**	-.004	-.024	.785*	-.022	-.064
Professors – Interactions in or outside of class **[Professor Influence]**	.136	.036	.606*	-.126	-.087
Teachers (from K-12) **[K-12 Influence]**	.097	-.035	.475*	-.022	.139
Academic Advisors - Interactions in or outside of official advising time **[Academic Advisor Influence]**	-.004	-.024	.785*	-.022	-.064
Other Students **[Peer Influence]**	.043	.027	.661*	-.021	-.012
Office of Undergraduate Research - If your school has one, if not, or you don't know, mark NA **[Office of Undergraduate Research]**	-.011	-.086	.610*	.010	.063
Family Members – Family can be biological or chosen **[Family Influence]**	-.085	.157	.566*	-.004	.132
Other Mentors- Anyone you consider a mentor that has not been previously listed **[Other Mentors]**	-.003	.020	.702*	-.008	.001
Survey Category: INTEREST - Your interest in participating in undergraduate research.					
Interest in Science Generally **[Interest in Science]**	.724*	.093	.088	-.081	-.069
Interest in Solving Real-World Problems **[Interest in Solving Real-World Problems]**	.826*	.020	.048	-.041	-.035

Table 2.4 (Continued)					
Subscale (Cronbach's alpha)	How you think (0.925)	What you dream (0.927)	Who you know (0.845)	What you know (0.757)	What you do (0.720)
Survey Items [Research shortened name]					
Survey Category: INTEREST - Your interest in participating in undergraduate research.					
Interest in Exploring New Ideas [Interest in Exploring New Ideas]	.950*	-.032	-.022	.012	.012
Interest in Learning New Skills [Interest in Learning New Skills]	.895*	-.015	-.028	.030	.006
Interest in Questioning Misconceptions [Interest in Questioning Misconceptions]	.819*	-.011	-.045	.037	.089
Interest in Research [Interest in Research]	.534*	.191	.160	-.020	-.045
Survey Category: FUTURE GOALS - Goals that may influence your decision to participate in undergraduate research.					
Career Goals [Career Goals]	.119	.842*	.019	-.074	.051
Grad/Professional School Goals [Graduate/Professional School Goals]	.091	.867*	-.022	-.108	.001
Survey Category: OPPOURTUNITY - Impacts on your ability to participate in research experiences.					
Your GPA [GPA]	-.005	.023	.087	-.647*	.042
Your Major [Major]	.062	.155	.047	-.780*	.039
COVID-19 – Anything COVID related in the past or current (e.g., research being online, restrictions in place because of COVID) **[COVID-19]**	.016	-.126	-.045	-.051	.410*

Table 2.4 (Continued)					
Subscale (Cronbach's alpha)	How you think (0.925)	What you dream (0.927)	Who you know (0.845)	What you know (0.757)	What you do (0.720)
Survey Items [Research shortened name]					
Survey Category: OPPOURTUNITY - Impacts on your ability to participate in research experiences.					
Disability Limitations – Any disability you identify with **[Accessibility]**	.025	-.097	.071	-.125	.383*
Travel - Transportation to and from research locations **[Travel]**	.025	-.010	-.013	-.123	.484*

Values above .800 are widely considered adequate, with values above .500 appropriate for psychological constructs and initial development of scales (Field, 2013). Four items (student awareness of research opportunities, finding research opportunities, and influence of coursework both inside and out of the major) were included in the original survey but removed from the overall scale based on factor analysis results. Two items, influence of coursework both inside and outside of the major, did not load with a value of at least .32 on any factors. The remaining two items loaded onto a factor of their own, however, parallel and scree plot analysis along with theoretical underpinnings did not recommend the inclusion of this factor, so it was removed from the scale. It is hypothesized that these items may influence research participation in a different way than the rest of the factors, resulting in their lack of fit in the overall scale.

DeWitt et al. (2016) defines a four-factor model to describe science capital amongst secondary school students in England. This model was adapted here as a measure of science capital and SCCT surrounding undergraduate research experiences using the four suggested areas of *How you think*, *What you do*, *What you know*, and *Who you know* with the addition of future goals as a fifth factor named *What you dream* (Fig. 2.1).

RQ2: Research Participation

The majority of the R group had participated in one research experience (54.35%), which was most often in the first year of their college experience (49.13%; Table 2.5). When asked their primary reason(s) for not participating in research, 68.67% of the NR group said that they were not aware of available opportunities and 27.08% reported they

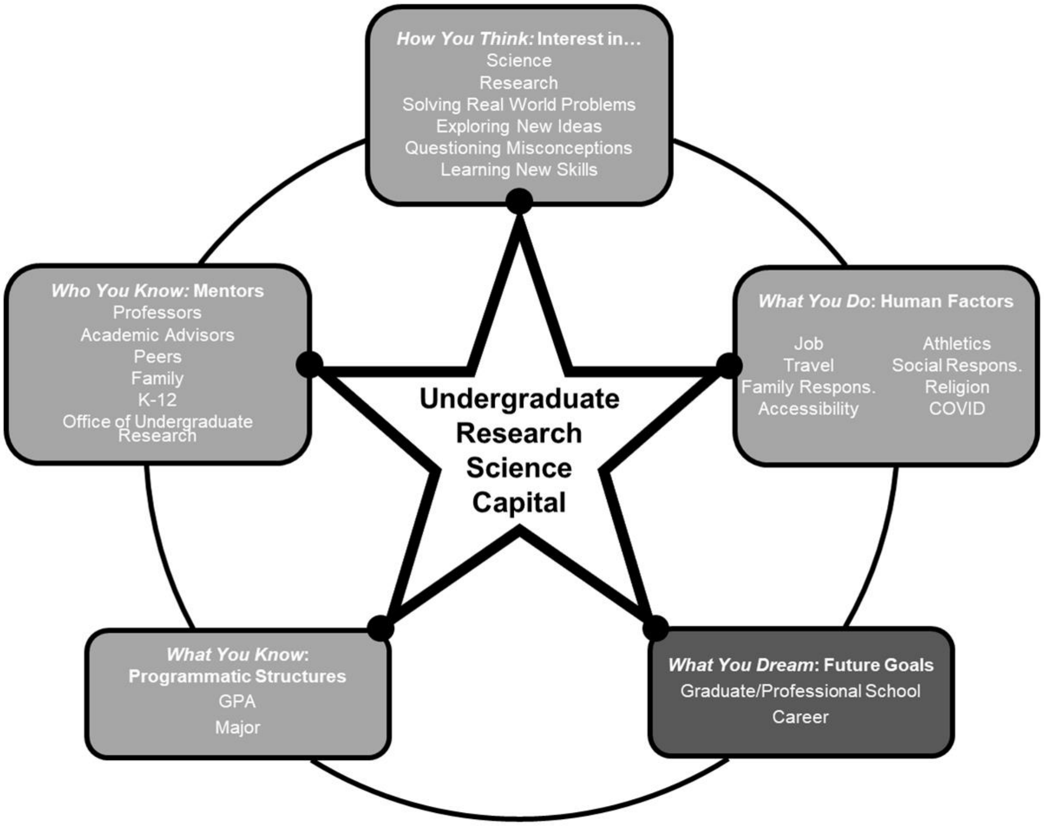


Figure 2.1: Undergraduate Research Science Capital Scale. Light gray boxes indicate elements drawn from Science Capital (Archer et al., 2015). Dark gray from Social Cognitive Career Theory (Lent et al., 1994).

Table 2.5. Researcher characteristics (% of researchers)	
Number of research experiences participated in	
1	54.35
2	28.26
3	10.00
4+	7.39
Year in college of first research experience	
1	49.13
2	26.52
3	18.70
4	7.39
5+	0.87
Type of experiences (select all that apply)	
Course based	50.00
Lab for credit or pay	43.48
Summer	35.65
Volunteer	16.96

had never considered research (Table 2.6). Of all respondents, 68.07% said that they hope to participate in research in the future. Student participation changes over time through college. Though the most prevalent research experience is course based, lab/research group-based experiences for credit and/or pay become much more common starting in the second year of college (Fig. 2.2). Additionally, there were differences in the year in college in which students began researching (Table 2.5) and the mean score of several of the participation factors was found to be significantly different between first-year (beginning of college) and fourth-year (end of college) students (Appendix H).

Mean scores of participant responses to the factors of the URSC Scale were calculated; these means were used to determine ten opportunities ($M \geq 5.00$), nine neutral factors ($M=4.00$-4.90), and six barriers ($M \leq 3.90$; Table 2.7). Ten factors presented significant differences between the R and NR groups (Professor Influence, Major, Interest in Research, Interest in Science, Family Responsibilities, Academic Advisor Influence, Peer Influence, Family Influence, Other Mentors, and GPA). Additionally, four items were not a part of the URSC Scale but were included in the survey; all demonstrated a significant difference between the R and NR groups (Table 2.7). The three opportunities with the highest sample mean were Graduate/Professional School Goals ($M=5.64$, $SD=1.22$; reported by 59.90% of students), Career Goals ($M=5.59$, $SD=1.21$; reported by 63.27% of students), and Interest in Learning New Skills ($M=5.58$, $SD=1.13$; reported by 77.07% of students). The three barriers with the lowest sample means were Accessibility ($M=3.50$, $SD=1.37$; reported by 18.61% of students), COVID-19 ($M=3.54$, $SD=1.35$; reported by 32.89% of students), and Athletics ($M=3.79$, $SD=1.20$; reported by 22.21% of students; Table 2.7).

Table 2.6. Influences for non-participation in research (Select all that apply; % of non-researchers)	
Not aware of opportunities	68.67
Time	52.92
Prefer to participate in an internship	33.10
Not interested	22.30
Not considered	27.08
Research opportunities do not pay well/at all	7.61

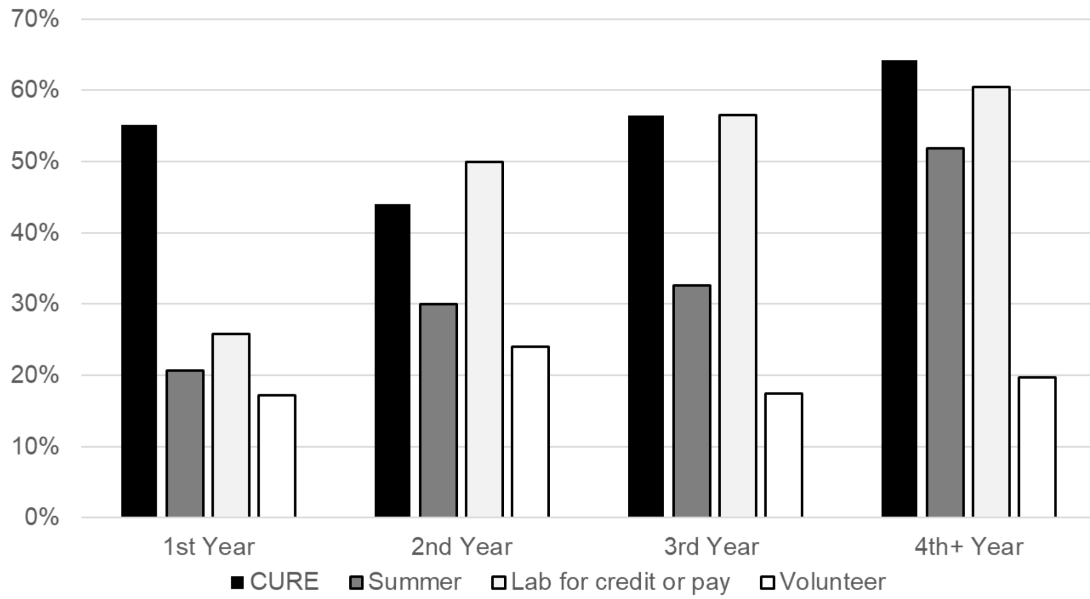


Figure 2.2: Percentage of students who have engaged in research by type of research experience.

Table 2.7. Survey responses. Responses given on a scale of 1 (extremely negative impact) – 7 (extremely positive impact). Factors sorted by Sample Mean response as opportunities (M≥5.0), neutral (M= 4.0–4.9), and barriers (M≤3.9). Significant results (p<.05) of the t-test comparing R and NR groups indicated by an *. Standard deviations are shown in parenthesis after each mean.

Factor	Sample Mean	Researcher Mean	Non-Researcher Mean	R/NR *p*-value
Opportunities (M≥5.0)				
Professor Influence	5.11 (1.18)	5.48 (1.14)	4.95 (1.16)	<.001*
Major	5.21 (1.16)	5.37 (1.18)	5.15 (1.15)	.007*
Interest in Research	5.18 (1.36)	5.56 (1.22)	5.03 (1.39)	<.001*
Interest in Science	5.56 (1.16)	5.71 (1.16)	5.49 (1.16)	<.001*
Interest in Solving Real World Problems	5.51 (1.13)	5.62 (1.08)	5.47 (1.14)	.113
Interest in Exploring New Ideas	5.46 (1.13)	5.51 (1.19)	5.45 (1.11)	.634
Career Goals	5.59 (1.21)	5.64 (1.18)	5.56 (1.23)	.122
Graduate/Professional School Goals	5.64 (1.22)	5.68 (1.21)	5.62 (1.22)	.440
Interest in Learning New Skills	5.58 (1.13)	5.57 (1.30)	5.58 (1.06)	.710
Interest in Questioning Misconceptions	5.20 (1.23)	5.15 (1.41)	5.22 (1.15)	.518
Neutral (M= 4.0– 4.9)				
Family Responsibilities	4.24 (1.28)	4.38 (1.30)	4.18 (1.27)	.023*
Academic Advisor Influence	4.83 (1.16)	5.06 (1.19)	4.74 (1.13)	<.001*
Peer Influence	4.79 (1.15)	5.05 (1.19)	4.68 (1.12)	<.001*
Family Influence	4.81 (1.22)	5.00 (1.22)	4.73 (1.21)	.001*
Other Mentors	4.72 (1.16)	5.02 (1.13)	4.59 (1.15)	<.001*
GPA	4.78 (1.23)	4.95 (1.32)	4.71 (1.18)	.015*
Social Responsibilities	4.08 (1.28)	3.98 (1.17)	4.12 (1.32)	.124
K-12 Influence	4.63 (1.27)	4.58 (1.12)	4.65 (1.31)	.479
Office of Undergraduate Research	4.53 (1.15)	4.59 (1.27)	4.50 (1.10)	.312
Barriers (M≤3.9)				
Job	3.92 (1.24)	3.87 (1.24)	3.94 (1.24)	.711
Athletics	3.79 (1.20)	3.77 (1.12)	3.80 (1.23)	.727
Religious Responsibilities	3.83 (1.15)	3.94 (1.00)	3.79 (1.20)	.061
COVID-19	3.54 (1.35)	3.54 (1.49)	3.55 (1.29)	.895
Accessibility	3.50 (1.37)	3.61 (1.30)	3.46 (1.39)	.196
Travel	3.90 (1.28)	3.87 (1.23)	3.90 (1.30)	.837
Items not in the Undergraduate Research Science Capital Scale				
Coursework in the Major	5.24 (1.28)	5.44 (1.21)	5.16 (1.29)	.002*
Coursework Outside of the Major	4.41 (1.16)	4.56 (1.18)	4.35 (1.15)	.026*
Finding Opportunities	4.45 (1.37)	4.70 (1.43)	4.35 (1.34)	.004*
Awareness of Opportunities	4.19 (1.45)	4.54 (1.46)	4.05 (1.42)	<.001*

On average students reported 12.38 influences as opportunities towards their participation in UREs and 3.01 influences as barriers.

Free Response Analysis

Free response questions were coded for additional factors utilizing magnitude coding and intercoder reliability measures. The three most frequently mentioned opportunities leading to research participation are students seeking out research experiences, coursework influences, and professor influences. These influences are directly opposed to the three most commonly mentioned barriers to research participation: students not knowing how to get involved, the amount of time a research commitment requires, and coursework as a hinderance to participating in research.

New Participation Factors. The potential effect of student travel to research sites (M=3.90, SD=1.28) and the COVID-19 pandemic (M=3.54, SD=1.35; Table 2.7) were both included in the survey as they were anticipated to have an effect during survey validation despite little literature support. Travel was supported in the free response by many students mentioning transportation to/from campus or to research sites off campus being a concern. Several other factors were mentioned in the free responses that had not been prompted in the survey. Students seeking their own opportunities into research is not frequently mentioned in URE literature, however it was the most commonly mentioned opportunity and barrier in the free response. Additionally, 68.67% of NR students selected this as one of the primary reasons they had not yet become involved (Table 2.6).

Additionally, many students mentioned the impact their mental health has on their research participation, as exemplified by this student's reflection.

"Mental Health as both a[n] outside responsibility to deal with as well as an obstacle for entering research. Having obligations outside of work and school to also take care of mental health in college is time consuming. It's also an obstacle as my feelings of imposter syndrome and anxiety definitely held me back from participating in research. I often felt unqualified to get involved and the rejection and silence one gets from professors/PIs etc. when first inquiring after research opportunities can be very discouraging especially when dealing with these two issues."

-Biochemistry Major, Research Participant (R)

Several other students (n=15) described mental health concerns as well as imposter syndrome being a barrier to research, both explicitly as with this student and more implicitly as with this biology major who had not participated in research (NR) when asked for the largest barrier to participation, "*I feel I am not competent enough, whereas other students are more applicable* [sic] *to doing research*." A last newly mentioned influence was a student indicated their citizenship status as the largest barrier to their participation. These newly mentioned influences provide further insight into the paths students navigate contributing to their participation in UREs and are beneficial in further developing a scale for URSC.

Discussion

Significantly larger proportions of research students reported having a disability (z= 2.86, p=.004) and being a member of the LGBTQ+ community (z=4.35, p<.001; Table 2.2) than their peers who had not previously participated in research. In a national study, Hughes (2018) found a disproportionately large number of students that were members of the LGBTQ+ community to be participating in research, however, that did not translate to persistence to a STEM degree.

Efforts to support retention of LGBTQ+ students in science majors are hampered by current data collection and survey methodologies (Freeman, 2018; 2020). This is in part because data on these students' experience is not collected by the majority of nationally representative datasets, and when these data are collected, students may be concealing that portion of their identity (Freeman, 2020). In response to a 2022 presidential executive order (Biden, 2022), recent efforts have been made by education organizations such as the American Educational Research Association (AERA) and the American Association for the Advancement of Science (AAAS) urging the inclusion of sexual orientation and gender identity (SOGI) indicators on surveys conducted by the National Science Foundation (NSF; AERA, 2022).

Additionally, the data in this study was self-reported. With regards to disability status, as many as two-thirds of students with disabilities do not report them to their universities. This creates a support gap for student accessibility services and underrepresentation in institutional datasets (NCES, 2022b) while directly affecting the construct of "*What you do*" (Fig. 2.1). Studies on how to support students with commonly concealed identities (such as sexuality status, disabilities, and mental health struggles) are a needed avenue of research (Bingham, 2021; Cooper, Gin, Barnes, et al., 2020; Freeman, 2018, 2020).

Citizenship was mentioned by one student in the free response as the largest barrier to their participation. Though often an overlooked equity issue, nationality has been found to significantly affect students' access to resources and thus their participation in educational opportunities such as UREs (Cacciatore, 2021; Gonzales, 2016). This lack of access to proper resources has major impacts on their *"What you do"* forms of capital. An example of a barrier presented by a students' citizenship status is that several U.S. funding agencies require U.S. citizenship to receive their funding. Citizenship status was not explored in the demographic questions, and it is possible that its effect on participants was greater than captured by this analysis.

Student Search for UREs

Students seeking and finding their own research experiences, an influence that draws on students' "*What you know"* form of URSC, was the most mentioned opportunity in the free response questions for students entering UREs (n=55). However, many students described struggles with not knowing how to access research opportunities or get involved. As described by this physics major (NR), "*I would love to participate in research. I just don't know how. I haven't the faintest idea how to begin that process."* Students not knowing how to get involved in research is the most commonly mentioned barrier in the free response, and this is further supported by 68.67% of NR students selecting this as one of the primary reasons they had not yet become involved (Table 2.6). This juxtaposition is exemplified by both influences of finding opportunities (M=4.45, SD=1.37) and awareness of opportunities (M=4.19, SD=1.45; Table 2.7) presenting a neutral sample mean, indicating that it was an opportunity for some students, but a barrier for others. Science departments and faculty could help mitigate this barrier by communicating with their students what opportunities are available and how to access them early and often. In this study, respondents indicated that CUREs were the most common research experience overall, especially for first year students (Fig. 2.2). However, other forms of research experiences became more common as students progressed through their collegiate careers, indicating that as students progress through college and gain more science capital, how they participate in UREs also shifts from CUREs to more lab-based UREs (Fig. 2.2).

The effect of coursework on students' participation in UREs was listed among both the most frequent opportunities and barriers. Students, like this geology major (R) described conducting research within courses themselves, "*In Geology, we have research classes each semester starting our sophomore year. These classes really prepare us to take on our own research. We are lucky that the geology program gets us so involved in research so early*." There were also accounts where courses and curriculum provided a means for students to learn about available opportunities and become involved outside of class.

> *"After this one class in which we listened to different people from the department discuss their research, I looked into different people in a research area that interested me. I ended up reaching out to a professor and we wrote a proposal for a project."* -Physics Major, (R)

However, a biology major (R) described how coursework could be a hinderance to participation, "*My school and work schedules impact my ability to participate in undergraduate research the most.*" Science curricula are often inflexible and outside research experiences may be difficult to schedule while balancing coursework and other outside responsibilities.

Studies have suggested increasing the incorporation of research experiences into courses to increase availability of opportunities and participation (Bangera & Brownell, 2014). By providing opportunities for research and coursework simultaneously, CUREs can assist with many of the commonly mentioned barriers to URE participation, including students searching for opportunities, the amount of time research takes, and coursework preventing students from participating in research. Bingham (2021) also found that significantly fewer students viewed logistics (including travel, a newly added barrier in our study) as a barrier to URE participation when the research was course-based. However, the coursework code in the free responses also applies to students describing finding their research opportunities through classes, through modes such as course content, guest speakers, and graduate teaching assistants (TAs). Students interact with course material daily, and it is a natural means to disseminate information about beneficial opportunities such as UREs.

Professor Influence

Instructors often serve as mentors to their students both inside and outside of their courses and contribute to their "*Who you know*" capital. The third most frequently mentioned opportunity in the free response was professor influence (n=49). Additionally, 65.55% of students quantitatively responded with this as an opportunity (M=5.11, SD=1.18; Table 2.7). This code consisted of instances where students described being directly invited to participate in research by a professor, or where the student cited a professor as the major influence leading to their research participation. In addition to assisting students, there were accounts of professors inspiring students to want to participate in UREs as was the case for this genetics major (NR), *"My desire to participate in research largely came from the excitement that I saw in my professors in my department. I wanted to challenge myself with problem-solving tasks to find solutions to unanswered questions in human medicine."* Haeger et al. (2021) found that professor influence was the leading opportunity into research experiences in their study at a midsized institution. The differences could be attributed to considering different factors into research and/or the difference in size between the institutions leading to less opportunity for individualized professor interactions.

Despite being a positive influence overall, some respondents (n=13) shared accounts of a professor creating a negative research environment for them, or their friends dissuading them from participating in a URE. This was the case for a chemistry major (NR) who indicated their largest barrier to participation, *"…not liking how professors treat the undergrads researching with them, especially in my major…."* Research experiences have the potential to be unhappy and unsafe environments for any student, yet those who are members of communities underrepresented in science based on gender, race/ethnicity, sexuality, and those with disabilities are particularly vulnerable (Marín-Spiotta et al., 2020; Santana & Singh, 2021; Traxler, 2020). Students' relationship with professors as their research advisors greatly impacts the student experience and research group environment (BrckaLorenz et al., 2017; Cooper, Gin, Barnes, et al., 2020). It is important that institutions acknowledge the potential for these situations to become negative and implement guidelines to ensure that UREs are a safe and positive environment for students (Ackerman et al., 2018; Demery & Pipkin, 2021).

Accessibility & Student Health

How students access and interact with their UREs is a portion of their *"What you do*" capital. Accessibility had the lowest sample mean, indicating that students viewed it as the largest barrier to their participation in UREs (M=3.50, SD=1.37; Table 2.7). A science curriculum often requires long labs and field environments that can be difficult to navigate (Batty & Reilly, 2022; Carabajal et al., 2017). In addition to their own accessibility needs, several students, such as this chemistry major (R), described situations where they could not meet the needs of their service dogs in research settings which limited their participation, *"I have a psychiatric service dog so it makes participating in research more difficult as the labs are not the most accommodating environment for a dog."*

Students described how their accessibility needs were not met in research settings, however there were cases where students' accessibility needs increased their interest in research.

> *"For disabilities, I put negative. As I continue to think about my disability (type 1 diabetes) I've realized that I shouldn't have marked negative and should've marked positive. It has made me rethink many aspects of life and helped my desire to possibly one day find a cure."* - Biology major (NR)

Another student also described the potential for research as a means of learning more about their own health, *"I have a joint condition, I would like to participate in research so that I could learn more about it.… I am currently on crutches and recovering from a surgery so I can't access many places easily."* - Biology major (NR)

The impact of student mental health on participation in research, a newly identified factor in this study, is a related understudied area. Cooper et al. (2020) have explored student depression and its effect on persistence in UREs as well as a student's relationship with their research advisor.

They found that students' depression negatively affected their motivation to participate in UREs and their engagement and productivity while participating (Cooper, Gin, Barnes, et al., 2020). Though coded separately from mental health, students also described imposter syndrome as a barrier to participation. In science fields, imposter syndrome has been found to be more prevalent in highly achieving students, women, and members of traditionally marginalized racial, ethnic, and religious groups (Chrousos & Mentis, 2020).

Lastly, the COVID-19 pandemic has significantly impacted science and higher education as a whole (Cameron et al., 2021; Myers et al., 2020) and engaged learning opportunities such as UREs are not exempt (Erickson et al., 2022; Utah, 2020). Effects of the COVID-19 pandemic on participation in research was one of the highest barriers to students both by average (M=3.54, SD=1.35; Table 2.7) and prevalence (32.89% of students reporting it as a barrier),

"I've had some really good experiences with my research advisors and some really negative experiences with some research advisors. I'm not

entirely put off by research, but I've had two extremes of the spectrum. Also, Covid-19 decimated my opportunities for research in undergrad (fresh-soph years for me) and [I] was completely unable to get anything. This has lead [sic] to a sense of desperation for me to get more experiences before applying to grad school." - Genetics Major (R)

However, several students also described changes in curriculum and research opportunities made due to pandemic response as being an opportunity for them.

> *"COVID helped my research opportunities because my 'big break' happened after taking a field class that was offered over Spring break in the [Local Research Area] due to travel restrictions. It was here that I met my current research advisor and he offered me the opportunity to participate in research with him over the summer."* -Geology Major, (R)

As institutions continue to adapt, there are lessons that can be learned from the pandemic response that can improve the accessibility of these experiences for students in the future, including the possibilities of online UREs (Barber et al., 2021) and considering field research sites that are closer to campuses to limit travel needs. Ensuring the accessibility of UREs is an important consideration for departments in an effort to make science available to all (Bingham, 2021; Gin et al., 2022; Lillywhite & Wolbring, 2019).

URE Impact on Science Recruitment and Retention

It is also of note that approximately 55% of students who participated in research, indicated participation in one or more research experiences outside of their declared major

(Fig. 2.3). Undergraduate research experiences have been closely linked with students' major choice and persistence to graduation as well as entering the STEM workforce (Chemers et al., 2011; Harsh et al., 2011). Additionally, the benefits of UREs have been found to cross research experience types and disciplines (Lopatto, 2009). Encouraging students to pursue UREs outside of their immediate discipline could expose them to other ideas within science which could lead to further solidification of major choice and career goals.

Operationalizing the URSC

The URSC Scale allows students to identify the impacts of influences URE participation in the five areas of "*How you think*, *What you do*, *What you know*, *Who you know*, and *What you dream*." Science departments and their respective institutions could use this scale as a means of identifying areas for supporting the growth of their students' science capital to promote participation in engaged learning activities such as UREs. The scale was developed and EFA performed, however dissemination of the scale to similar populations followed by confirmatory factor analysis (CFA) is needed for further development of the scale and is an avenue for future work.

Conclusion

Previous studies have identified four main areas of Science Capital: *What you know*, *Who you know*, *How you think*, and *What you do* (DeWitt et al., 2016). When considering educational practices such as UREs, the effect of students' goals on their participation is an additional important area of consideration. This led to the inclusion of the SCCT-influenced factor, *What you dream*, to the construction of a scale measuring URSC. The

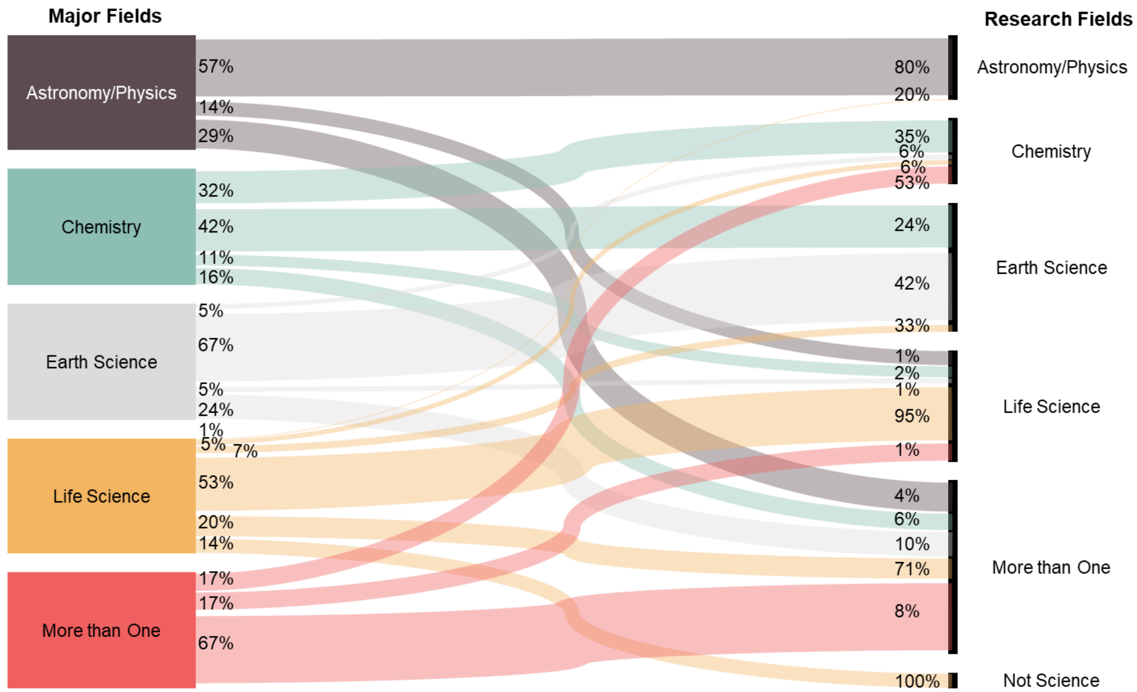


Figure 2.3: Comparison of major with the research field of students who conducted undergraduate research. Percentages represent the proportions of researchers in each group. Ex: 42% of chemistry majors (left side) conducted research in Earth science (right side); 24% of all Earth science research was conducted by chemistry majors.

combination of these five factors captures many of the influences that are present for student participation in UREs. The majority of the most common and strongly identified opportunities and barriers into research are related to science capital. This holds true for the factors that are significantly different between the R and NR groups and supports the validity of the scale developed by this study. Institutions and departments can leverage this and utilize the scale to measure their students' science capital related to UREs to improve programming and the equity of availability of experiences. A prominent way to do so is making sure the paths into UREs are made clear to students. There are many ways to participate in research experiences, but by increasing students' capital and making them aware of research opportunities and the potential benefits, students can make informed decisions about their future research participation.

Acknowledgments

This work was reviewed and approved exempt under the institution's IRB office and supported by a Geological Society of America Graduate Student Research Grant. The authors would like to thank our participants, the faculty who assisted with survey dissemination, as well as Shannon Conner, Cole Bowman, Gavin Gleasman, Dr. Brian Dominy, Dr. Marian Kennedy, and Dr. Bridget Trogden.

CHAPTER THREE

EXTRACTING MEANING FROM MULTI-DIMENSIONAL EDUCATION DATA: TOPOLOGICAL DATA ANALYSIS OF UNDERGRADUATE SCIENCE RESEARCH PARTICIPATION

This paper is currently under review in the Proceedings of the National Academy of Sciences of the United States of America. The following modifications were made to include the article in this dissertation: 1) tables and figures were renumbered, 2) all references were moved to the full list at the end of the document, 3) research questions were renumbered.

Abstract

The benefits of undergraduate research experiences (UREs) are well documented, particularly in science fields. However, there is little research focused on the recruitment strategies involved in bringing science students into undergraduate research positions. Additionally, researchers have called for improvements of the equity of the process. This study leverages a big-data approach to science education research, topological data analysis (TDA), to identify student influences on entry to undergraduate research. Topological data analysis is a powerful quantitative methodology that has yet to be widely applied within a science education context and allows researchers to group participants with more nuance than traditional clustering methods. Here we demonstrate the application of this methodology, the resulting common characteristics of student groups formed by TDA, and the influences identified as opportunities or barriers to participation in undergraduate research. This study adds to the growing body of educational research that utilizes big-data

approaches like TDA to understand student pathways through higher education. These results will help science departments and institutions gear their research recruitment efforts in ways that will reach the students that may currently be underserved by these highly impactful experiences.

Significance

Improving the equity of opportunity to enter undergraduate research experiences may help fulfill the ever-growing need for science students in the STEM workforce. This study leverages topological data analysis (TDA) to identify influences on research entry amongst undergraduate science students. Here we show the resulting common characteristics of student groups formed by TDA and the influences identified as opportunities or barriers. This study contributes to the conversation of increasing student involvement in higher education and explores a big-data approach not frequently utilized in education research. These results will help science departments and institutions gear their research recruitment efforts in ways that will reach the students that may currently be underserved by these highly impactful experiences.

Introduction

Efforts to improve recruitment and retention in the sciences, particularly for those from traditionally marginalized groups, is of paramount importance as the need for scientists to join the workforce is ever-growing (Zilberman & Ice, 2021). Engaged learning practices such as undergraduate research experiences (UREs) have been found to be highly beneficial in promoting science students' persistence toward degree, exploration and actualization of career goals, growth of technical and professional skills, as well as many other academic and career related benefits (Harsh et al., 2011; Russell et al., 2007; Seymour et al., 2004). Despite the body of research surrounding the benefits of UREs, few studies have analyzed the influences surrounding student participation. Further understanding of the opportunities and barriers students face surrounding this highly influential educational practice will lead to further promotion of equity within UREs (Krim et al., 2019) by empowering institutions to increase the opportunities and lower the barriers presented.

Since 2017, the National Science Foundation (NSF) has promoted its ten Big Ideas (*NSF*, n.d.); two of these, Harnessing the Data Revolution and Growing Convergence Research between Disciplines, are critically important for STEM education research to address modern challenges in higher education. Fields within STEM education in higher education spaces are often limited by participant sample sizes from utilizing big-data methods. This study bridges the gap between topological data analysis, a highly quantitative methodology, and STEM education in the understudied space of improving the equity of entry into undergraduate research experiences.

Topological Data Analysis in Discipline-Based Education Research

Topological data analysis (TDA), an offshoot of machine learning, is an emerging statistical methodology used to visualize structure in data that may not be readily apparent (Carlsson, 2009; Wasserman, 2018). Originally designed for use with image analysis, it is used in a variety of fields, both theoretical and applied, to understand relationships between data present in complex datasets (Serrano et al., 2020; Wasserman, 2018). Despite its widespread use, TDA has only been applied to educational data in a handful of studies (Boyd et al., 2023; Doyle, 2017; Godwin et al., 2019). Expanding the use of the method in education research opens opportunities for visualization of data and data networks in new and innovative ways with the potential to transform the discipline.

Topological data analysis works through persistent homology (Fig 3.1). To carry this out, a researcher-designated filtering lens is applied to the data as a focus for cluster analysis. The algorithm then creates a network of the similarity relationships between the data. In the network, similar datapoints (in the case of education research, these are typically participant responses) are clustered together into nodes. When a given response demonstrates enough similarity that it could be part of more than one node, an edge is drawn to tie the two points together. Groups are formed based off interconnected nodes (e.g., the network in Figure 3.1 would have two groups). The network can be further analyzed to explore groupings of datapoints with more nuance than traditional clustering methods because greater dimensionality is considered in the formation of the network (Doyle, 2017). Further explanation of TDA and its applications in survey methodology can be found in the works of Doyle (2017) and Godwin et al. (2019) and in Appendix I.

Theoretical Framework Guiding Research Design

The survey for this study was developed utilizing an adaptation of Jones et al.'s (Jones et al., 2020) model for the intersection of Science Capital (Archer et al., 2015) and Social Cognitive Career Theory (SCCT; Lent et al., 1994). Science Capital, a science-focused extension of traditional Bourdieusian forms of sociological capital, has a primary focus on the individual (Archer et al., 2015; Bourdieu, 1986). Meanwhile, SCCT is an educational theory designed to analyze big decisions in a student's academic career. In this

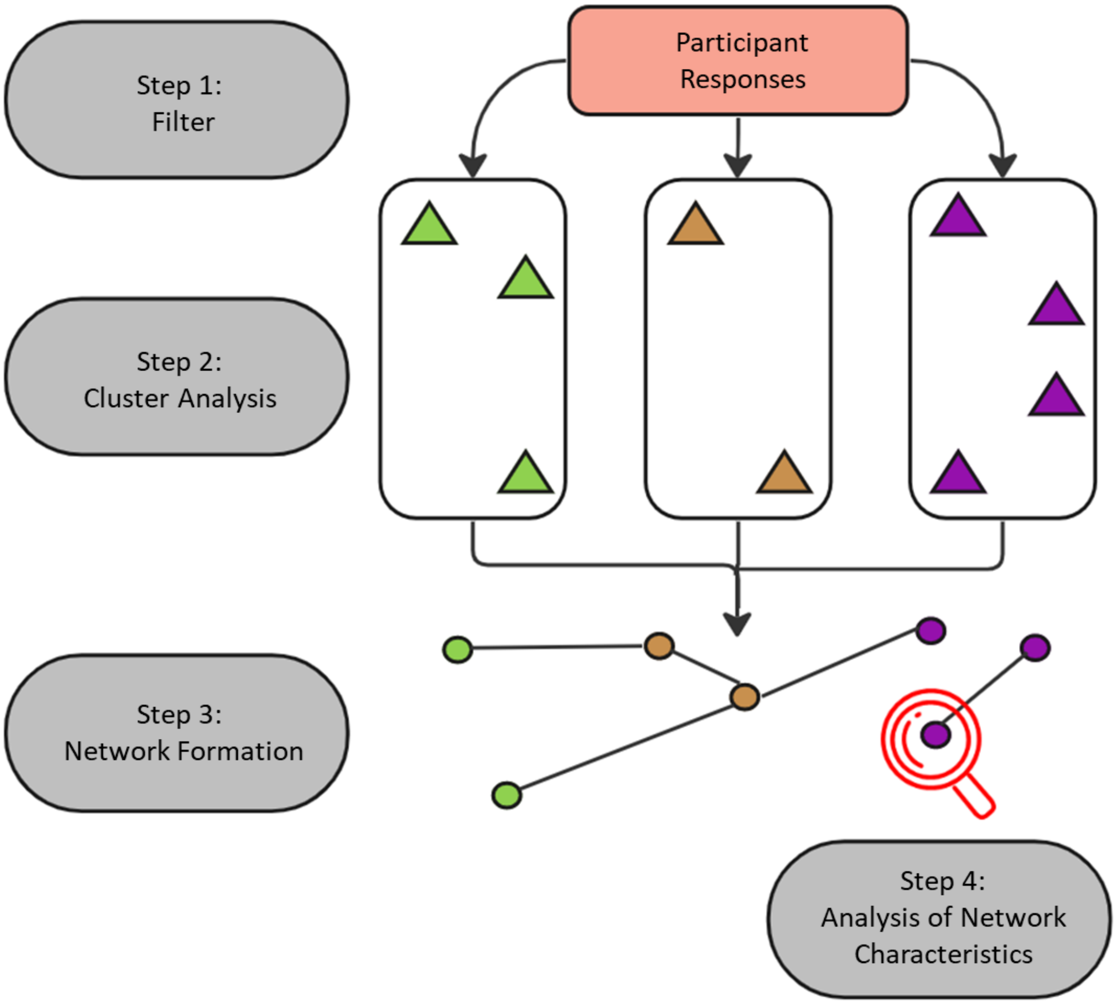


Figure 3.1: Overview of topological data analysis.

study, it is being used to focus on institutional impact on a science student's participation in undergraduate research. Combining theories with both an individual and institutional focus allows researchers to explore the influences that students face regarding their participation in UREs at multiple levels. This theoretical framework guides the exploration of the research questions for this study:

2) What influencing factors are identified by science students as impactful for participation or non-participation in undergraduate research experiences?

3) In what ways are factors different across model-identified groups?

Materials and Methods

Survey Development

A survey with both Likert-style and free response questions was developed to capture the level of influence known factors had on each student's participation in undergraduate research and identify new influencing factors. Likert-style responses were collected on a scale from 1 (extremely negative impact) – 7 (extremely positive impact), with a not-applicable option. Twenty-five influencing factors to undergraduate research participation were included (Figure 2.4). Twenty-three of the twenty-five items were identified from the literature (Haeger et al., 2021; Szteinberg, 2012). Potential influences were presented neutrally so as to not bias the respondents. The survey was reviewed by educational researchers and undergraduate students prior to dissemination with a focus on pragmatic and communicative validation (Walther et al., 2013). The influence of COVID-19 and travel to/from research sites were identified as potential factors that were not previously found in literature and added before survey dissemination.

Study Population

Surveys were administered at four public, doctoral granting, highest research activity (R1) institutions in the Southeastern United States. One institution is a Hispanic-serving institution (HSI) while the remaining three are predominantly white institutions (PWI). Half of the institutions are land-grant institutions. In this study, science major was defined as those majors falling within the Classification of Instructional Programs (CIP) codes for physical and life sciences (*CIP User Site*, n.d.; Appendix B). Eight hundred thirty-three responses were included in this study.

Mapper Algorithm

The Mapper algorithm (Singh et al., 2007) was developed for carrying out TDA in R (R Core Team, 2017). When adapting the algorithm for use with survey data, each datapoint represents one participant's responses. To cluster datapoints, Mapper combines parameters that inform the network formation. These parameters include the desired clustering method, filtering lens, number of slices (*n*), amount of overlap, and cut height (ε). The clustering method parameter describes how the distances between points and group memberships will be determined. For these models, k-nearest neighbors (KNN) classification was chosen. In KNN, the algorithm determines the approximate distance between points and groups them accordingly into the class that is nearest to them (Pandey & Jain, 2017). Mapper then utilizes the indicated lens to focus the data into a network of vertices. Researchers have many options for the choice in lens including mathematical operations, machine learning formulas, or alternative datasets. However, the lens selection should be informed by the research questions to be most impactful (Bak, 2014). This study presents two mappings of the survey data, one utilized the number of opportunities identified in student responses (Likert-type responses >4 [neutral]) and the second utilized the number of barriers (responses <4) identified. Researchers also choose the number of slices, percentage of overlap of lenses in vector space, and ε, the approximate distance each lens covers. These parameters are largely variable and can be adjusted until a reasonable, resolvable model of the data is formed. Approximate guidelines for parameter choice include: [1] approximately 100 datapoints should be allotted for each slice, and [2] $\varepsilon \approx \sqrt{d}$, where d is the number of items included in analysis (Equation 3.2; Godwin et al., 2019). In this study, 26 slices, 50% lens overlap, and a cut height of 4 were utilized for the opportunity mapping. Nine slices, 50% lens overlap, and a cut height of 4 were utilized in the barrier mapping. Algorithm outputs indicate in which vertex each datapoint (in this case, study participant) is located. However, there are duplicates where one participant can be a member of more than one vertex. Vertices that contain duplicates are connected by a line called an edge. In this study, interconnected vertices were kept in the same group. Additional applications of TDA in education research that describe model decisions and further examples can be found in the works of Doyle (2017), Godwin et al. (2019), and (Boyd et al. (2023).

Input data must meet specific characteristics for the algorithm to sort effectively. Survey responses must use approximately the same number of response options and there must be as few missing responses as possible (Doyle, 2017). To address these requirements, survey items were on a Likert-style scale of 1 (extreme barrier) – 7 (extreme opportunity) and missing survey responses were assigned a score of "4" to signify a neutral response as neutral responses do not affect the chosen clustering methods.

Topological Data Analysis Input Variables

Topological data analysis is designed to condense highly dimensional data into a viewable network. However, this visualization is limited by the relationship between the number of data points and the number of dimensions: the greater the number of datapoints the more dimensions are condensable by the algorithm. When applying TDA to educational survey data, the survey items are the dimensions and the participant responses are the datapoints, therefore there is a limit to the number of survey items that can be included in the TDA based on the number of participants. This relationship is guided by the suggestion $N \approx 2^d$ (Equation 3.1), where d = the number of items included in the analysis and N is the total number of respondents needed to resolve the TDA for a given value of d (Formann et al., 1980). With 833 participants, TDA is able to resolve approximately ten dimensions; ultimately eleven were selected based on their variance, uniqueness, and theoretical interest determined by the research team. Selection criteria was based on those used by Doyle (Doyle, 2017). Exploratory factor analysis of the twenty-five-item scale resulted in a Cronbach's alpha .888. The factor structure included one factor that contained all six interest items with a subscale Cronbach's alpha of .925 (Table 2.4). As such, the six interest items were averaged together to result in one input variable. Ten additional items were incorporated as the *input variables* to the TDA algorithm and are presented in Table 3.1; five variables had a mean response greater than a neutral response (4) and are identified as

Table 3.1. Input variables for TDA
COVID-19; Accessibility; Job Obligations; Travel Requirements; Athletic Commitments; Religious Responsibilities
Avg. Interest; Major Requirements; Professor Influence; Career Aspirations; Graduate/Professional School Aspirations

opportunities, six had a mean response lower than neutral and are identified as barriers. Topological data analysis code utilized in this study is available in Appendix J.

Statistical analysis was used to compare the groups present in the TDA mapping across demographic and input variables. Student's *t* test, *z* tests, and chi-square analysis was used where applicable. In all cases, significance was indicated by a $p \leq 0.05$.

Results

Topological Data Analysis Mappings

The TDA output map with a lens focused on the of the number of barriers experienced by students did not resolve clear groupings because the vast majority of participants are represented by nodes that are interconnected by edges (Appendix K). This indicates that there is greater variation in the number of opportunities presented by participants than the number of barriers. Due to this, the remainder of the analysis was carried out only considering the opportunity mapping (Fig 3.2).

Three distinct groups were identified from the TDA output that utilized a lens of the number of opportunities for participation in undergraduate research (Fig 3.2). These groups are: [1] students that had a significantly higher number of opportunities presented (HO; 465 students; $t(773)=35.15$, $p<.001$), [2] those that identified a significantly lower number of opportunities (LO; 310 students; $t(773)=35.15$, $p<.001$)), and [3] those that had mainly neutral responses (58 students). groups are: [1] students that had a significantly higher number of opportunities presented (HO; 465 students; $t(773)=35.15$, $p<.001$), [2] those that identified a significantly lower number of opportunities (LO; 310 students; $t(773)=35.15$, $p<.001$)), and [3] those that had mainly neutral responses (58 students).

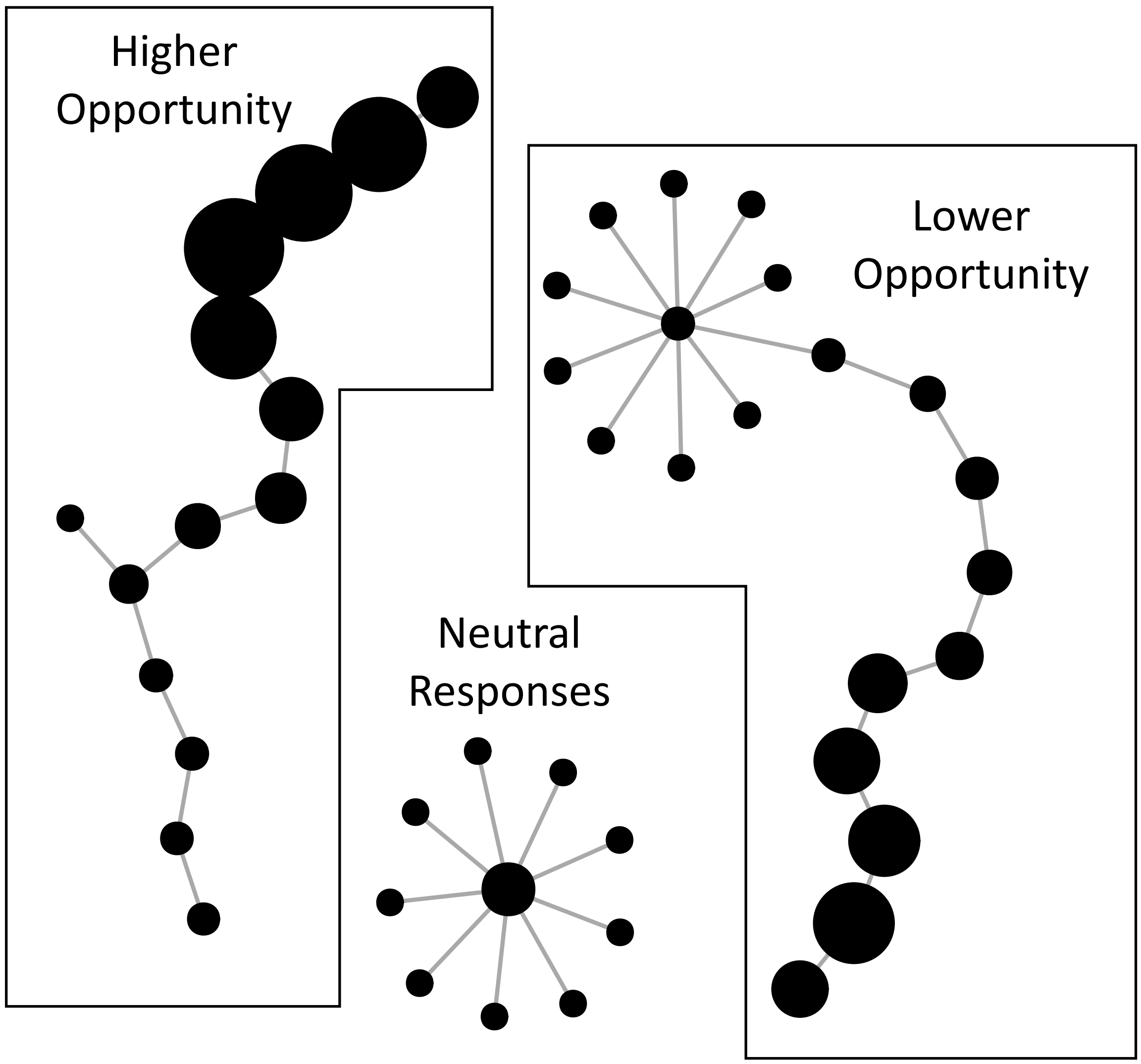


Figure 3.2: TDA map using number of opportunities as a filter. Higher Opportunity (HO), and Lower Opportunity (LO) groups are separated by boxes. Neutral group is in between the HO and LO group. Nodes (circles) represent many participants; lines represent edges within the data structure.

The mean responses and *t*-test results for each input variable of the TDA-derived groups are presented in Table 3.2. Ten of the eleven chosen input variables demonstrated significant differences between the HO and LO groups; accessibility was the only input variable that did not significantly differ between these groups. Despite five of the items presenting a population average of an opportunity (≥"5.0"), the LO group did not have any items that were determined to be opportunities. Two items, travel requirements and job obligations, were determined to be neutral items for the HO group but barriers (<"4") when considering the entire population (Table 3.2).

A series of *z*-tests of demographic variables did not present significant differences (Table 3.3). However, a significantly larger proportion of students in the HO group participated in research than those students with fewer opportunities that comprise the LO group (z=3.55, p<.001). Chi-square testing of the year in college in which participant responded to the survey revealed a higher proportion of first year students in the LO group (X^2 (3,9.074, p =.028). The HO group was found to have significantly greater numbers of opportunities (t(773)=35.15, p<.001) and the LO group was found to have significantly greater numbers of barriers ((t(773)=9.63, p<.001). This evidence supports the methodological choice of using a lens of the number of opportunities to complete the TDA. Chi-square analysis of respondent's majors revealed that physical science majors (Chemistry, Geological/Earth Sciences, and Physics/Astronomy) were more likely to be members of the HO group than their life science peers (Biology, Genetics, Biochemistry; (X^2 (7,28.567, p <.001)); Table 3.4).

Table 3.2. Input Variable Results. Mean responses on a scale of 1 (extreme barrier) – 7 (extreme opportunity) and *t*-test outcomes. * Indicates statistical significance between the higher and lower opportunity groups. Standard deviations are shown in parenthesis after each mean.

	Higher Opportunity (HO) Group	Lower Opportunity (LO) Group	*t* Score	*t*-test *p* Value	Neutral Group	Study Population
Participants	465	310	-	-	58	833
Participant %	55.82	37.21	-	-	6.96	100.00
Accessibility	3.77 (1.11)	3.65 (0.98)	1.54	.123	3.76 (0.73)	3.50 (1.37)
COVID-19	3.69 (1.39)	3.46 (1.13)	2.43	.015*	3.78 (0.73)	3.54 (1.35)
Athletics	3.94 (1.28)	3.58 (1.11)	4.04	<.001*	3.72 (0.81)	3.79 (1.20)
Religious Responsibilities	3.97 (1.23)	3.63 (1.07)	4.49	<.001*	3.81 (0.69)	3.83 (1.15)
Travel Requirements	4.09 (1.26)	3.69 (0.98)	4.72	<.001*	3.76 (0.73)	3.90 (1.28)
Job Obligations	4.18 (1.27)	3.57 (1.17)	6.76	<.001*	3.67 (0.94)	3.92 (1.24)
Professor Influence	5.53 (1.00)	4.51 (1.10)	13.36	<.001*	3.79 (0.67)	5.11 (1.18)
Major	5.59 (1.02)	4.77 (1.09)	10.67	<.001*	3.78 (0.73)	5.21 (1.16)
Avg. Interest	5.84 (0.78)	4.90 (1.05)	14.28	<.001*	3.80 (0.66)	5.42 (1.19)
Career Goals	5.77 (1.12)	4.79 (1.11)	11.98	<.001*	3.76 (0.73)	5.59 (1.21)
Graduate/Professional School Goals	5.71 (1.18)	4.72 (1.15)	11.56	<.001*	3.95 (0.39)	5.64 (1.22)

Table 3.3. Analysis Variable Results. * Indicates statistical significance between the higher and lower opportunity groups. Standard deviations are shown in parenthesis below each mean where applicable. LGBTQ+ = lesbian, gay, bisexual, transgender, and queer plus community. Genders traditionally underrepresented in science include female, non-binary, and more than one gender selection. Race/Ethnicity underrepresented in science reported in this study include American Indian or Alaskan Native, Black or African American, Hispanic and/or Latino/a/x, Native Hawaiian or Pacific Islander, and Middle Eastern.

	Higher Opportunity (HO) Group	Lower Opportunity (LO) Group	Statistical Test Value	*p*-Value	Neutral Group	Study Population
Average number of opportunities (TDA Lens)*	16.00 (2.74)	9.00 (2.67)	35.14	<.001*	0.16 (0.37)	14.38 (1.93)
Average number of barriers (TDA Lens)*	2.00 (2.24)	4.00 (3.54)	9.63	<.001*	2.76 (6.39)	3.75 (4.06)
Research Participants (%)*	34.84	22.90	3.55	<.001*	12.07	28.45
LGBTQ+ (%)	12.26	10.00	0.97	.332	1.72	13.67
Disability (%)	9.68	7.42	1.09	.276	3.45	8.58
Transfer students (%)	17.63	12.69	1.85	.054	24.14	19.24
Genders traditionally underrepresented in science (%)	79.35	73.55	1.88	.060	67.24	73.60
Race/Ethnicity underrepresented in science (%)	18.92	20.97	-0.70	.484	18.97	19.68
Pell Grant recipients (%)	18.71	20.32	-0.56	.575	12.07	24.00
Year in college (%)*			X^2 (3,9.074, p =.028			
1	44.95	54.84			53.85	49.70
2	19.78	16.77			38.46	18.97
3	16.99	11.29			7.69	14.17
4+	17.20	15.48			0.00	14.53

Table 3.4. Participant majors. * Indicates statistical significance between the higher and lower opportunity groups. Life science majors include General Biology, Microbiology and Immunology, Biochemistry, Biophysics, and Molecular Biology, and Genetics. Physical science majors include Chemistry, Geological and Earth Sciences/Geosciences, and Astronomy and/or Physics.

		Higher Opportunity (HO) Group	Lower Opportunity (LO) Group	Neutral Group	Study Population
	Major (%)*	X^2 (7,28.567, p <.001)			
Life Science	General Biology	61.29	76.77	88	67.59
	Microbiology and Immunology	9.68	4.84	3.45	7.56
	Biochemistry, Biophysics, and Molecular Biology	8.17	3.55	1.72	5.40
	Genetics	5.38	3.55	1.72	4.32
	More than one within Life sciences	1.51	1.29	1.72	1.44
Physical Science	Chemistry	5.59	5.16	0.00	5.14
	Geological and Earth Sciences/Geosciences	5.16	2.90	1.72	3.76
	Astronomy and/or Physics	4.09	0.97	0.00	2.13
	More than one within Physical sciences	0.65	0.00	0.00	0.63
Both	More than one major in Physical and Life science	0.43	1.29	0.00	0.72

Discussion

The TDA mapping which utilized the number of barriers identified by participants as a filter did not result in meaningful groups (Appendix K). This lack of differentiation may indicate that though there are differences in the number of barriers (M=3.01, SD=3.27) presented by students, there are more differences in the number of opportunities (M=12.38, SD=5.39; Table 3.3). Topological data analysis was originally designed for use with continuous variables, therefore the closer together the data values are (i.e., such as the spacing between ordinal items), the less difference there is in vector space for separation. This may contribute to the lack of differentiation present in the barrier mapping (Doyle, 2017).

Characteristics of topological data analysis-derived groups

In the opportunity mapping, the HO group responded with a higher-than-average number of opportunities (HO) and a higher proportion of students that had participated in research than the LO group (z=3.55, p<.001). Encouragingly, there were no significant differences across the HO and LO groups for the demographic variables of LGBTQ+ status, disability status, transfer status, gender, race, or Pell Grant status (Table 3.3). Not all students will participate in UREs, but by increasing the number of opportunities available to students, more students have the *choice* to participate, rather than not participating because they do not know how, or they cannot access the opportunities.

The HO group had a significantly higher proportion of second, third, and fourth or more-year students compared to the LO group (X^2 (3,9.074, p =.028; Table 3.3)). This highlights the effectiveness and importance of programming and activities throughout the college experience that increase undergraduate research science capital and awareness of UREs. Students enter college with differing levels of science capital, but such programming can become an equalizer for many students, particularly marginalized students (Carpi et al., 2017; Pierszalowski et al., 2021; Ries & Gray, 2018).

Additionally, there was a significantly lower proportion of life science students and higher proportion of physical science students in the HO group compared to the LO group (X^2 (7,28.567, p <.001; Table 3.3)). Both the American Chemical Society (ACS, 2015) and the American Physics Society (APS, 2008) have viewed URE participation as so vital that they formally recommend its incorporation into the undergraduate curriculum. Additionally, of the four participating institutions, three had URE requirements for graduation in one or more of their physical science degrees while only one had a similar requirement for graduation in the life sciences. The difference in research requirements between the disciplines, at least in this study, may in part be due to the size of the departments and "space" (both in terms of lab space and mentorship opportunities with faculty) for students within UREs. Nationally, 80% of science majors are within the life sciences, leading to these departments being larger and increasing the logistical difficulty for departments to provide UREs for their large numbers of students (Hamrick, 2022). Encouraging students to pursue UREs outside of their immediate major and implementation of course-based research experiences (CUREs; for a richer discussion of these experiences see Bangera & Brownell, 2014; Kirkpatrick et al., 2019) are some ways to control for large amounts of students in UREs.

Model-identified opportunities and barriers to research participation

Responses for ten of the eleven input variables indicated a significant difference between the HO and LO groups (Table 3.2). This confirms literature that suggests these items may have an effect on undergraduate research participation (Haeger et al., 2021; Szteinberg, 2012). Several outside commitments (i.e., influences unrelated to academic commitments) presented mean scores indicating they are likely to be more of a barrier for the LO group than the HO group. These include the influences of job obligations (($t(773) = 6.76$, $p < .001$), athletics ($t(773) = 4.04$, $p < .001$), religious responsibilities ($t(773) = 4.49$, $p < .001$), and travel requirements ($t(773) = 4.72$, $p < .001$; Table 3.2). These results suggest that LO students may have more demands on their time that makes scheduling UREs difficult. Efforts to incorporate UREs into curriculum CUREs and encouraging mentor flexibility in scheduling would help lower these barriers to participation (Bingham, 2021).

Influences of the COVID-19 pandemic and accessibility are primarily affected by institutional structure and are closely linked. Adaptations made to UREs for COVID-19, such as decreasing the required travel for participation and allowing students to work remote as needed, helped improve the accessibility of the experiences (Erickson et al., 2022). This is a potential explanation for why, although the respondents' mean scores for the influence of COVID-19 were found to be a barrier for both the HO (3.69) and LO (3.46) groups, there was a significant difference in the responses, with the HO group's responses averaging higher (($t(773) = 2.43$, $p = .015$). This indicates that for some students the flexibility demanded by COVID-19 had the potential to be an opportunity. Accessibility, however, was a barrier for both groups and was the only input variable that did not have a significant difference between the HO and LO groups. University accessibility offices are typically dedicated to courses and curricular accessibility, however, assisting students with navigating entry into opportunities such as UREs may be beyond the scope of their offerings (Hall & Belch, 2000). Accessibility of UREs is an often overlooked topic and one of important consideration when assessing the equity of access to experiences (Gin et al., 2022).

The influence of a professor was another input variable that exhibited significant differences between the HO and LO groups. Students in the HO group indicated professor influence was an opportunity (M=5.53, SD=1.00) while those in the LO group indicated it as a neutral influence (M=4.51, SD=1.10; $t(773) = 13.36$, $p < .001$). Previous studies indicate that faculty interactions are amongst the most impactful influence of research participation for students (Haeger et al., 2021). However, high student-to-faculty ratios at large institutions can inhibit these interactions. Nonetheless, that students in the HO group responded with faculty interaction being an opportunity makes sense, as students who are more engaged and involved in their institutions (i.e., have more opportunities) are more likely to have faculty interactions. Events that allow for students to interact with faculty and hear about potential UREs may be beneficial for students, especially those who are not already well-connected within their departments.

Six survey items were related to interest and were determined to be factoring together through exploratory factor analysis. Due to the high correlations that emerged during factoring, they were averaged together for inclusion as one input variable (see Materials and Methods and Table 2.4). Students in the HO group responded significantly higher for this mean aggregated interest item than their LO peers (M=5.84; SD=0.78; $t(773) = 14.28$, $p < .001$; Table 3.2). This is consistent with previous studies that suggest that students with higher early (K-12) levels of student science interest were correlated with greater participation in science opportunities later in their academic careers (Alexander et al., 2012). Additionally, Archer et al. (Archer et al., 2012) found that students with higher levels of science capital were more interested in science and that an important contributor to science capital is engagement in science-related practices that can support sustained interest in science (such as UREs). Participation in UREs has been shown to contribute to increases in students interest in their discipline (Harsh et al., 2011; Kirkpatrick et al., 2019; Russell et al., 2007; Seymour et al., 2004), and a larger proportion of the HO group had participated in UREs than the LO group (Table 3.2). As part of this study, it is not possible to determine the exact relationship between these students' interest, the number (and which) opportunities they have access to, and their participation in UREs. However, the data suggest that activities found to increase student science interest may also contribute to opportunities related to UREs.

Similar conclusions can be applied to students who identified a higher-than-average number of opportunities also perceiving their career aspirations ($t(773) = 11.98$, $p < .001$), graduate and professional school goals ($t(773) = 11.56$, $p < .001$), and major ($t(773) = 10.67$, $p < .001$; Table 3.2) to have a positive influence on their URE participation. Studies have found that students with differing levels of cultural and science capital utilize that capital in different ways as they engage in science experiences and continue towards a science career (Archer et al., 2015; Habig et al., 2021; McPherson, 2014). DeWitt et al. (2016)

found that while aspects of cultural capital contributed to science participation, science capital had a larger effect on participation, specifically for items involving students' perceived transferability and utility of science. This suggests that one way to increase participation in UREs may be to make the benefits of UREs more apparent to students, particularly those that contribute to student professional development (Seymour et al., 2004; Thompson & Jensen-Ryan, 2018)).

Implications for STEM education research

Topological data analysis allows for examination of quantitative data with more dimensionality than traditional clustering methods (Doyle, 2017). For STEM education research, this provides an avenue for introduction of big-data approaches in a way that meets the sample sizes available (e.g., hundreds to thousands rather than millions of data points). Within STEM education research, this study (n=833) is amongst the smallest sample size, with larger studies considering responses of thousands of students; this demonstrates the scalability of the methodology within education research (Boyd et al., 2023; Doyle, 2017; Stevens, 2016). By using TDA, this study was able to consider a greater number of influences to undergraduate research, which provides a more complete approach to the research question: *What influencing factors are identified by science students as impactful for participation or non-participation in undergraduate research experiences?* Students are complex and have many influences on their participation in undergraduate research. Additionally, not every student has the desire to participate. However, by considering a greater number of influences, science departments have a fuller picture of ways to improve the equity of entry into these beneficial experiences.

Conclusion

Bringing TDA to education research is a means of incorporating a scalable big-data approach to the field. By allowing researchers to consider data with high dimensionality, they can reveal results that would not be considered with other methods, allowing researchers as well as faculty and departments opportunities to make evidenced-based improvements to the student experience. This study, focusing on quantitative survey data exploring science student entry into UREs, provides promise for further exploration of data science methodologies in education research.

Respondents who were identified as HO students reported professor influence, interest, major, career goals, and graduate/professional school goals as opportunities to their undergraduate research participation. The average scores of their LO peers were significantly lower across all of these factors and did not have any influences that could be identified as an opportunity. Additionally, the HO group had significantly higher responses than the LO group for the influences of job obligations, athletics, religious responsibility, COVID-19, and travel requirements. Accessibility did not present significant differences between the HO and LO group responses (Table 3.2) and was seen as a barrier to research participation by both groups. A focus on accessibility could be a place to start improving the equity of UREs as it was a barrier across all groups.

Science students face many factors surrounding their participation in UREs, some individual and some institutional. These results suggest that science departments encouraging faculty to share available UREs with students, as well as developing communication with students about interest, motivation, and career related benefits of participation in UREs, may be impactful in promoting access to UREs (Ceyhan & Tillotson, 2020; Chemers et al., 2011; Frantz et al., 2017).

As previous studies involving science capital suggest, students entering college with higher levels of capital are more likely to report interest in, see the benefits of, and participate in science learning experiences (Archer et al., 2012, 2015; DeWitt et al., 2016; Moote et al., 2020). However, students can increase their science capital throughout their college experiences, and by making the availability and outcomes of UREs more apparent, the number of opportunities present for students to potentially pursue participation in UREs also increases. By increasing these opportunities, the number of students that are interested in science and pursue science careers may also increase, which can help meet the ever-growing need for STEM professionals in the workforce (Zilberman & Ice, 2021).

Acknowledgments

The authors would like to thank our participants, the faculty who assisted with survey dissemination, as well as Dr. Allison Godwin, Dr. Stephen Moysey, Dr. Brian Dominy, Dr. Marian Kennedy, Dr. Bridget Trogden, Cole Bowman, Shannon Conner, and Gavin Gleasman. This work was reviewed and approved exempt under the institution's IRB office. This work was supported in part by a Geological Society of America Graduate Student Research Grant.

CHAPTER FOUR

"I'M STILL HERE AND I WANT THEM TO KNOW THAT": STUDENT EXPERIENCES AND EFFECTS OF CONCEALABLE IDENTITIES ON UNDERGRADUATE RESEARCH SCIENCE CAPITAL

This chapter is currently under review for the Journal of Research in Science Teaching. Anonymized transcripts are available upon request. The following modifications were made to include the article in this dissertation: 1) tables and figures were renumbered, 2) all references were moved to the full list at the end of the document, 3) research questions were renumbered.

Abstract

Concealable identities, those which are not always visually apparent, and individuals must navigate choosing to reveal have been found to impact individuals' psychological well-being. The effect of concealable identities on student participation in engaged learning activities is an understudied avenue of research. In a phenomenographic approach rooted at the intersection of the theories of Science Capital and Social Cognitive Career Theory (SCCT), this study analyzes the experiences of ten women as they navigate their visible and concealable identities surrounding the entry into undergraduate research experiences. Though all ten students described similar levels of Undergraduate Research Science Capital, they did so in different ways. These results highlight the need for multiple approaches to equity efforts to ensure that high-impact practices such as undergraduate research are accessible to all students.

Introduction

Engaged learning experiences such as undergraduate research have been found to have highly beneficial outcomes for undergraduate students (Krim et al., 2019; Sandquist et al., 2019). However, few studies explore the pathways which students take to enter undergraduate research experiences (UREs), and even fewer of those examine the potential impact of concealable identities on student participation (Bingham, 2021; Cooper, Gin, Barnes, et al., 2020; Haeger et al., 2021). Concealable identities are those which are not visually apparent and include sexual orientation and disabilities amongst other identity characteristics. Concealable identities are often marginalizing. Students can be faced with the choice of revealing these pieces of their lives or leaving them concealed to their peers and mentors, a decision which often causes stress (Cooper, Gin, & Brownell, 2020). Additionally, the effects of marginalized identities, visible and concealable, lead to intersectionality which individuals must navigate, compounding on the mental toll (Robinson, 2018).

In a phenomenographic approach rooted at the intersection of Science Capital and Social Cognitive Career Theory (SCCT), this study describes the experiences of ten women surrounding their undergraduate research participation. As women in STEM, these participants all carry at least one marginalized identity, this study explores the effect of their concealable identities on their participation. Thus, offering an understudied perspective, and new potential approaches to improvements of equity in UREs.

Literature Review

Pathways into UREs

Though the positive outcomes of UREs are well established particularly for underrepresented students, few previous studies have examined the pathways by which students enter UREs (Castillo & Estudillo, 2015; Eagan et al., 2013; Lopatto, 2009). Haeger et al. (2021) investigated opportunities and barriers to research participation at a public, primarily undergraduate, Hispanic Serving Institution (HSI) in the Western United States. Their study included undergraduate students, faculty members, and academic advisors; many of their identified barriers fall into the categories of institutional barriers (e.g., finding a mentor, fitting it into one's curriculum), other commitments (e.g., having to use that time for an outside job, familial commitments), and affective concerns (e.g., lack of sense of belonging). Bangera and Brownwell (2014) described many of these opportunities and barriers but also issues of student awareness regarding URE opportunities, how to pursue them, and the benefits of UREs, they called for implementation of research experiences into courses (sometimes called CUREs) to improve student accessibility to opportunities. Additionally, Cooper et al. (2021) identified "rules to research" that students progress through when entering research experiences and recommended publicizing the hidden curriculum behind UREs to improve the equity of entry.

Other studies have analyzed opportunities and barriers to UREs at a programmatic level instead of an individual level. Their identified positive and negative influences include institutional financial resources, faculty availability, limited student preparation, faculty support with curriculum development, and department/administrative support of UREs (Frantz et al., 2017; Hewlett, 2018; Kirkpatrick et al., 2019; Lopatto et al., 2014; Morales et al., 2017). These influences primarily impact the institution, and its employees rather than directly impacting the students. However, they can all serve as opportunities or barriers (or even both) to research participation depending on implementation and have an influence on students' pathways to URE participation.

Concealable Identities in STEM

Identities can either be visible, meaning they are apparent without one knowing or speaking to the individual or concealable, meaning they are pieces of that individual's life that they may choose not to reveal to others. Individuals may choose to conceal different aspects of their identities for many reasons. Oftentimes this concealment occurs because of an identity that is stigmatized in their community or is minoritizing to individuals. Several concealable identities such as religion, disabilities, coming from a low socioeconomic background, and being a member of the LGBTQ+ community impact student experiences within STEM communities (Busch, 2022; Cooper, Gin, & Brownell, 2020; Scheitle & Dabbs, 2021). Previous studies have established a connection between possessing, and attempting to conceal, these identities and increased psychological distress (Quinn et al., 2014). Representation is beneficial for both visible and concealable identities to help improve students' sense of belonging (Lewis et al., 2016). However, it can be especially difficult for students to find mentors who share their concealable identities as the mentors themselves must navigate their choice to reveal their identities in their workplace and with their students (Cooper et al., 2019; NASEM, 2019; Yoder & Mattheis, 2016). Additionally, concealable identities are not always clearly defined, leaving individuals uncertain if they are eligible for available support (Santuzzi et al., 2014) and further exacerbating the underrepresentation of these individuals in large scale demographic datasets (Freeman, 2020). As Cooper, Gin, & Brownell (2020) proposed, understanding how concealable identities relate to URE participation is an important step toward creating more inclusive research experiences with the potential to improve the retention of underserved undergraduates in STEM.

Research spaces are not always safe or welcoming spaces for women and those with other identities marginalized in STEM (Clancy et al., 2014; Giles et al., 2020; Kuchynka et al., 2018; St. John et al., 2016). Many women find themselves having to outperform their male colleagues only to receive lesser recognition (Bloodhart et al., 2020). As individuals balance multiple marginalized identities, this intersectionality adds to the complexity of their potentially concealed identities. Studies have found that individuals that carry more marginalized identities, both visible and concealable, are more likely to experience microaggressions and generally lower well-being (Robinson, 2018). Ramirez and Paz Galupo (2019) found that lesbian, gay, and bisexual persons of color (LGB-POC) reported significantly more incidents of stress, and symptoms of depression and anxiety than their LGB white peers. When designing research spaces, this potential should be considered to ensure a welcoming learning environment for all students (Demery & Pipkin, 2021; Diamond & Alley, 2022).

Theoretical Frameworks

Science Capital. Sociological capital can be defined as the identity and personality aspects that an individual "carries" with them as they go about their lives. Archer et al. (2015) describes a specific framing of sociological capital pertaining to interactions with science activities, ideas, and concepts. Science Capital has been previously described as having four constructs that help individuals navigate science fields, *How you think*, *Who you know*, *What you know*, and *What you do* (DeWitt et al., 2016). This study included a fifth construct, *What you dream,* which was revealed in factor analysis of a quantitative Undergraduate Research Science Capital Scale developed by the authors (Chapter Two; Figure 4.1).

Social Cognitive Career Theory. Though originally designed to describe students career choice, Social Cognitive Career Theory (SCCT) is frequently adopted to consider choices in a student's academic career (Lent et al., 1994). Social Cognitive Career Theory considers influences both internal to the individual and external such as institutional rules and structures that may impact student decisions. As such, it is beneficial to include when considering student participation in undergraduate research due to the external factors that are involved in the process. In particular, this study uses the construct of *expectancy outcomes*. Expectancy outcomes describe the effect of what an individual thinks will happen (their expected outcome) has on their decision to participate in a given experience. An example of an expectancy outcome is when students pursue experiences because they expect them to benefit their career, even if that outcome is not guaranteed. This forward-thinking mindset can have large impacts on student participation in undergraduate research.

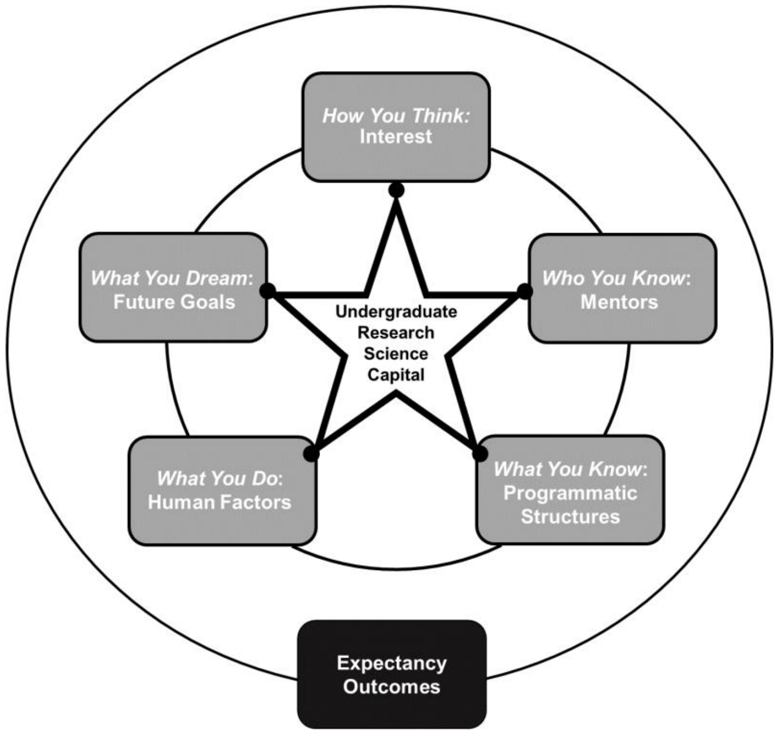


Figure 4.1: Theoretical framework. The inner ring contains the five Undergraduate Research Science Capital constructs used. The outer ring contains the SCCT constructs used.

The theoretical model for our study was influenced by the model created by Jones et al. (2020) combining Science Capital and Social Cognitive Career Theory (SCCT) to measure student Science Capital, and future science interests (Fig 1.1). Likewise, this study considers both theories when exploring student entry into undergraduate research experiences. However, this study deepens the exploration of the five areas of science capital to further explore the effect of concealable identities on students' participation in UREs. The theoretical model for this study is presented in Figure 4.1.

Research Question

The research question explored by this study is:

4) How do science students describe: a) proximal support and/or barriers they experienced and b) self-efficacy and outcome expectations related to participating in undergraduate research?

This study is addressing this research question while focusing on students with concealable identities due to the likelihood that these identities have an effect on student participation in engaged learning spaces such as UREs.

Methodology

This study is the phenomenographical, qualitative portion of a sequential explanatory (QUANT→qual) mixed methods study. In the quantitative portion, a survey pertaining to students' undergraduate research related Science Capital was administered at four R1 institutions in the Southeastern United States. Survey items were on a Likert-style scale of extreme barrier (1) – extreme opportunity (7). Topological data analysis (TDA) was utilized to sort the responses of 833 participants into groups using methods similar to those described by Godwin et al. (2019). The TDA mapping resulted in three groups, participants who responded with a higher-than-average number of opportunities (HO; N=465), participants who responded with a lower-than-average number of opportunities (LO; N=310), and participants whose responses were mainly neutral (Neutral; N=58). These outcomes are discussed in detail in Chapter Three. Ten interview participants were purposely selected from these respondents based on their position within the mapping and their reported demographic characteristics. Due to the differences in size between the HO, LO, and neutral groups, six participants were chosen from the HO group, three from the LO group, and one from the neutral group to ensure perspectives of members from each TDA group were included in qualitative analysis. All ten participants are women, as are approximately 75% of the survey respondents (Table 2.2). This provides an often-overlooked perspective of women with concealable identities in science spaces. This study was approved for exempt-level review by the institution's Institutional Review Board.

The interview protocol was designed to address the constructs of SCCT (Appendix L). Interviews were semi-structured, and approximately thirty minutes in length. Participants earned a $20 incentive card for their participation, awarded after the completion of the interview. Interviews were transcribed, cleaned, and verified by a researcher listening to interview recordings and checking the content of the transcripts. Then, two cycles of coding were performed in accordance with suggestions in Saldaña (2016). The first coding cycle included six passes to encompass each of the theoretical framework's constructs (Fig. 4.1). Themes and sub themes were developed from these codes as a transition between the first and second coding phases. A second cycle of coding then followed to confirm the themes and sub themes (Fig 4.2). Codes, code definitions, and themes were kept in a codebook (Appendix M) and consensus coding of both cycles was executed by an additional researcher familiar with qualitative coding to ensure validity. Consensus coding continued until both coders had reached full agreement.

Participant Characteristics

All ten participants identify as women and are science majors at an R1 institution in the Southeastern United States. Participants selected pseudonyms and designed icons for visual representation following protocols described in Boyd et al. (2023). Five participants, Britana, Susan, Emily, Camryn, and Cee had previously participated in research and the remaining five had not yet participated. Two participants are Black/African American, seven are White, and one is both Black/African American and White. Four participants reported disabilities, six reported being members of the LGBTQ+ community. Additionally, one participant was a Pell Grant recipient, an example of another potentially concealable identity (Kallschmidt & Eaton, 2019). However, due to her being the only participant who revealed that, it was not considered further in the analysis. Participant demographics are presented in Figure 4.3.

Results

There is no statistical difference presented between the average responses of the ten interview participants in total compared to those that identified as being a member of the LGBTQ+ community or having a disability (Table 4.1). However, in interview responses it becomes apparent that though those that possess these potentially hidden identities

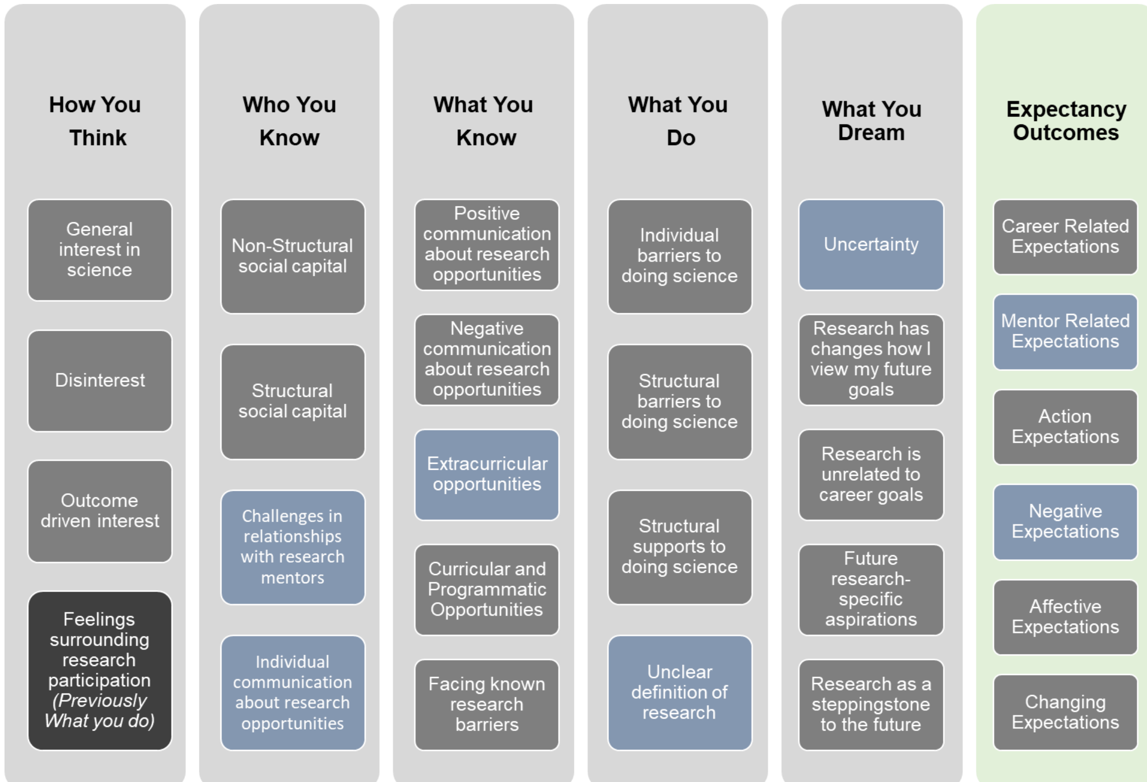


Figure 4.2: Coding themes and subthemes. Major themes are represented by the six vertical boxes. Subthemes are represented by smaller boxes within each theme. Subthemes removed between the first and second rounds of coding are shaded in blue and the subtheme that was moved between themes is shaded in black. Light gray shaded themes are constructs of Undergraduate Research Science Capital; green shaded theme is a construct of SCCT.

High Opportunity Group	Britana	Jay	Josie	Morgan	Susan	Tee
Major	Genetics	Microbiology	Biology	Biology	Geology	Chemistry
Participation in Research	Researcher	Not yet participated in research	Not yet participated in research	Not yet participated in research	Researcher	Not yet participated in research
Race	White	Black/ African American, White	White	White	White	Black/ African American
Disability Status	Student with a disability	Student with a disability	No reported disability	No reported disability	No reported disability	No reported disability
LGBTQ+ Status	LGBTQ+ Student	LGBTQ+ Student	LGBTQ+ Student	Not LGBTQ+ Student	LGBTQ+ Student	Not LGBTQ+ Student

Neutral and Low Opportunity Group	Emily	Camryn	Cee	Monday
Major	Biochemistry	Biochemistry	Biochemistry	Chemistry
Participation in Research	Researcher	Researcher	Researcher	Not yet participated in research
Race	White	Black/ African American	White	White
Disability Status	No reported disability	No reported disability	Student with a disability	Student with a disability
LGBTQ+ Status	LGBTQ+ Student	Not LGBTQ+ Student	LGBTQ+ Student	Not LGBTQ+ Student

Fig 4.3. Participant characteristics. Icons and pseudonyms were participant designed following guidelines in (Boyd et al., 2023). Emily is the Neutral group participant.

Table 4.1. Interview participant survey averages across areas of Undergraduate Research Science Capital. Study mean is comprised of all ten interview participants survey responses. Unpaired t-tests performed between the study mean and the responses of those who self-identified as being members of the LGBTQ+ community and those that reported a disability. * Indicates significance. Few values obtained significance however, differences between communities were revealed in qualitative survey responses.

		How you think	Who you know	What you know	What you do	What you dream
Interview	Study Mean (N=10)	5.57	4.97	5.55	3.86	5.85
	LGBTQ+ Students (N=6)	5.69; t(15)=0.15, *p*=.882	4.76; t(15)=0.53, *p*=.216	5.57; t(15)=0.04, *p*=.966	3.64; t(15)=0.64, *p*=.533	5.86; t(15)=0.01, *p*=.990
	Students with Disabilities (N=4)	6.22; t(11)=1.05, *p*=.318	5.00; t(11)=0.05, *p*=.958	5.33; t(11)=0.34, *p*=.741	3.86; t(11)=0.00, *p*=1.00	6.00; t(11)=0.03, *p*=.980
Survey	Study Mean (N=833)	5.35	4.73	4.95	3.89	5.22
	LGBTQ+ Students (N=114)	5.58; t(945)=2.07, *p*=.039*	4.59; t(945)=1.42, *p*=.156	5.03; t(945)=0.75, *p*=.451	3.79; t(945)=1.46, *p*=.146	5.35; t(945)=0.96, *p*=.339
	Students with Disabilities (N=71)	5.36; t(902)=0.73, *p*=.464	5.01; t(902)=1.44, *p*=.150	5.02; t(902)=0.53, *p*=.595	3.89; t(902)=0.00, *p*=1.00	5.20; t(902)=0.12, *p*=.903

present similar levels of each area of science capital, they do so in differing ways. By identifying the ways in which students with concealable identities express science capital, science departments and their respective institutions can better promote student participation in UREs. These results demonstrate the ways that these representations differ across all six theoretical framework constructs (Table 4.2).

How You Think

Capital related to *How you think* about undergraduate research mainly pertains to students' expressed dis/interest in undergraduate research and its influence on student participation. The students that did not reveal any potentially concealed identities (i.e., students without concealable identities; white boxes in Table 4.2) did not describe interest in research in their interviews. This does not mean that research interest is not present, as they did express it when specifically asked on the survey but could indicate that interest is less of a driving factor for research participation for these students than their peers who are members of the LGBTQ+ and/or disabled communities.

Several of the LGBTQ and students reporting a disability expressed areas of *How you think* capital explicitly in their interviews. Emily and Britana both stated how a general interest in science was a driving factor for their research participation, as exemplified by Emily's quote *"I really love science, and I always like was interested in research…like in high school, I didn't really exactly know what it meant but there was always something I wanted to try."* Josie and Cee described interest in specific outcomes such as experience and skill gains as part of their reasoning for potential research participation. Lastly, Emily,

Table 4.2. Student responses across subthemes. Boxes shaded in red represent students who are members of the disabled community, boxes shaded in yellow represent students who are members of the LGBTQ+ community, boxes shaded in orange represent students who are members of both the LGBTQ+ and disabled communities, boxes shaded in white are students that are neither members of the LGBTQ+ nor disabled communities.

Subtheme	Camryn	Morgan	Tee	Monday	Emily	Josie	Susan	Britana	Cee	Jay
How you think										
General Interest in Science					X			X		
Disinterest	X		X					X		
Outcome Driven Interest						X			X	
Who you know										
Non-Structural Social Capital			X			X	X		X	X
Structural Social Capital					X		X	X	X	X
What you know										
Positive communication about research opportunities	X	X	X				X			
Negative communication about research opportunities	X	X	X	X		X	X		X	X
Curricular and Programmatic Opportunities			X		X	X	X			
Facing known research barriers			X				X	X	X	
What you do										
Individual barriers to doing science	X		X			X				
Structural barriers to doing science		X	X	X		X	X	X	X	X
Structural supports to doing science					X			X	X	

Table 4.2 (Continued)

Subthemes	Camryn	Morgan	Tee	Monday	Emily	Josie	Susan	Britana	Cee	Jay
What you dream										
Research has changed how I view my future goals	X				X		X	X	X	
Research is unrelated to career goals				X						
Future research-specific aspirations			X		X	X	X			X
Research as a steppingstone to the future							X			
Expectancy Outcomes										
Career related expectations	X			X	X	X		X	X	X
Action Expectations						X	X	X		
Affective Expectations	X								X	
Stated Career Goals										
Exiting STEM	X									
Medical Field		X				X			X	
STEM Industry				X			X			
STEM Research			X		X		X	X		X

Josie, Cee, and Jay all described feelings that surrounded their decisions. These feelings were usually negative and connected to peer descriptions "*Just from having like basically like no motivation to like engage.*" Feelings were sometimes positive though, as with Cee's reflection on her own research experience, *"But if I knew that it wasn't something that has to be like twenty hours a week of me like doing this busy work, or it could be something that actually is like stress relieving and therapeutic and inspiring to me, I would have joined a lot sooner."*

Who You Know

Who you know capital pertains to mentors and other influential people in students' lives that may have an impact on their undergraduate research participation. These descriptions were sorted into structural social capital, meaning influences coming from individuals who worked for or represented their university, and non-structural social capital, pertaining to individuals who are not directly connected to the university. The students that are not members of the LGBTQ+ nor disabled communities mentioned very little *Who you know* capital in their interviews, with the exception of Tee who described non-structural social capital in her statement that she found research opportunities through her peers but did not participate yet because she did not find an opportunity she was interested in.

All of the students who reported a potentially concealed identity described some form of *Who you know* capital with the exception of Monday. Josie, Susan, Cee, and Jay all described non-structural social capital, as Susan did here when describing how she got involved with her first research experience.

"I had just come as a freshman. It was not even my first month here, and a girl that worked not with my Ph.D. candidate, but with one in the same lab, was like, 'Oh, I know this guy who's looking for a student to help him do work, and that does like environmental chemistry sort of stuff. So, if you're interested here's his contact info', and that's how I kind of got my foot in the door."- Susan

Emily, Susan, Britana, Cee, and Jay all described instances of structural social capital. Cee in particular explained how her research mentor was openly gay and knowing that before she even joined the research group was a large driver in her desire to participate so that she could make connections with a mentor with a shared identity.

"…And so, it was like cool to know that there was like a gay professor on campus like obviously, there is…but someone who's like very much out about it, someone that like created a space for people to talk about these things that I haven't ever really been able to talk about. Yeah. So that was like immediately, like, I really want to be a part of this. So, I'm very thankful that he responded to me, and I'm glad I emailed in a coherent way." - Cee

What You Know

Sub-themes within *What you know* include positive and negative forms of research-related communication, curricular and programmatic opportunities, and facing known barriers to research. This theme appears to have a larger impact on the students who did not report a potentially concealed identity as all three reported both positive and negative instances of communication, and Tee had descriptions that met all four subthemes. Positive communication about research opportunities included instances in which participants described communication that led to opportunities for research. None of the students with a disability, and only one LGBTQ+ student (Susan), described instances of positive communication of research opportunities. Far more frequent was instances in which participants described communication struggles or lack of communication which led to a barrier to research, or a lack of participation entirely. Often participants described both, as was the case with Tee who had not participated in research yet but was graduating soon and it is a requirement for her graduation.

> *"Usually, the department chair would send emails out about research opportunities, so they are out there…{The} concept of undergrad research is a great one. It's just that I feel like it should just be more easily accessible to students to get into research. I feel like I could send an email, but sometimes these professors are not seeing the email. So now I have to figure out a time to actually go to their office and set up an appointment and talk to them, like, hey, look, I got to graduate. Can we make something happen? I just feel like it should be a little easier to communicate with professor or the TA that you're interested in their research that they're doing."* – Tee

Additional sub-themes of *What you know* capital include curricular and programmatic opportunities and navigation of known research barriers. Tee, Emily, Josie, and Susan all described curricular and programmatic impact on their ability to participate in research. Susan and Tee are both in majors which require research participation for graduation, and they described that as leading to opportunities for them to participate. Josie, a first-year student, wished research could be required in her major, biological sciences, so that it would be easier to fit into her curricular schedule. She also described her participation in a STEM-focused mentoring program specifically designed to increase diversity in STEM majors at the university. Josie had not yet participated in research, but felt the experience opened opportunities for her as she described here, and several other places throughout her interview.

> *"Yeah, [mentoring experience] it's like a one hour credit class that I signed up for that basically paired me with a mentor in like the biological sciences major, and it's like a class and kind of just like a little program to be a part of within the college and that kind of like helped your introduction to comes in and transition….Yeah, we have like a class time, and they would bring in some of the faculty from the college who are hosting like research or doing [course based research], and they would explain, like their projects and kind of present on it, and, like, give us the opportunity to email them. And if we mentioned we were in [mentoring experience]. They said that would be like definitely a factor in choosing who they were gonna work with them so definitely an opportunity to get into some projects there."* -Josie

Josie is a member of the LGBTQ+ community and also explained that this mentoring experience sometimes breaks students into groups based on certain identity characteristics, such as sexual orientation, so they have a space to discuss these particular aspects of their identity.

Tee, Susan, Britana, and Cee described facing known research barriers, a sub theme that encompasses descriptions of students navigating research barriers that they knew would be present before they began searching for research opportunities. This was the case here for Cee, who describes how multiple research barriers led her to pursue a non-lab-based science research experience which she enjoyed more than she expected.

> *"I didn't have this opportunity because of COVID, and because I didn't have these people guiding me through this, I didn't have parents who knew what the hell undergrad was about. It is supposed to look like COVID happened, I have ADHD which makes me forget things exist frequently, I have chronic illnesses which keep me in bed, and it's like none of those are compatible with what people think is 'real research'. So no, I feel like I don't have any experience in that, and I don't. But I still have all of these great experiences… And yeah, they can think they're all superior, but I'm still here and I want them to know that."* -Cee

What You Do

Excerpts that fall into *What you do* themes are can be categorized as individual, meaning they are not required for students' degree, or structural, meaning they contribute to the curriculum and are directly connected to university or departmental involvement. Individual barriers were mentioned by Camryn, Tee, and Josie and most often have to do with non-research related jobs or extracurricular activities, as described here by Camryn when she explained why she chose not to participate in undergraduate research. *"So mainly was I was involved in a lot of extracurricular activities so like organizations, and I was on* *boards. So, having to balance that with classes, and the research project was not beneficial for me."*

Structural impacts were sometimes opportunities, as described by Emily, Britana, and Cee. For example, Emily described how she was questioning her sexuality when she entered her research experience, but her supportive lab group helped her personally, which was an unexpected outcome for her.

> *"I still don't really know [about her sexuality]. Yet a lot of people in my lab actually are part of the LGBTQ+ community, and that has been helpful in figuring stuff out…. It definitely hasn't dissuaded me from my participation in research, because the majority of people actually surprisingly in my lab, are part of the [LGBTQ+] community, which I think is funny. But they're all really nice, and they're really supportive, and they like, talk openly about it. And everyone is just really nice. So that has been helpful."* -Emily

More frequently, structural impacts are barriers, as described by Morgan, Tee, Josie, Monday, Susan, Britana, Cee, and Jay. All students with a disability described potential accessibility concerns such Britana, who had participated in a virtual research experience explained here "*I have a physical disability with my hand… it didn't affect me because I can type relatively well, but I think if it was a lot more finicky, like in terms of like lab equipment, or needing both hands to do very like detail-oriented tasks. Maybe it would be a bit more difficult."* Other common structural barriers included courses not allowing room in a students' schedule for research. Morgan stated that this was the largest contributor to her not yet participating in research *"As of right now, I would definitely say my course* *schedule [is a barrier to research]."* This is a common sentiment, particularly for transfer students or those who came from schools that did not offer many advanced placement (AP) or dual enrollment opportunities, as Monday describes.

> *"I came from a school where they didn't offer any AP classes or dual enrollment that kind of thing which one would think that just puts you ahead for college, it shouldn't put you behind…And so, even though I was doing everything…what I would consider to be normal. It kind of set me behind compared to my peers. So that's been part of why I have more credit hours than some people and don't have time for research."* – Monday

Finally, effects of the COVID-19 pandemic were a commonly described structural barrier. As Susan describes here, "*I think COVID is probably the biggest one, because as a STEM student, it's really hard to do research fully online, especially in such a hands-on physical field, because you can't just pick up rocks through a computer like that's just not an option."* Susan had participated in two research experiences; however, one was cut short unexpectedly due to COVID and she stated that she would have participated in more research experiences if COVID had not been her largest research barrier.

What You Dream

What you dream capital generally refers to students' future goals. When asked about their future career goals only Camryn plans on exiting STEM completely (Table 4.2). Camryn participated in a research experience and described how, though not a negative experience, it showed her she did not want to continue research in the future.

> *"I kind of saw that I didn't really enjoy the research process. I didn't mind it. It just wasn't for me. So, I kind of was like, maybe I do not want to be a direct scientist, maybe have something indirectly to do with science. But I realized that that's okay, too. I realized that I don't think I want to do research."* -Camryn

Students that participated in research described incidences in which the research experience changed their future goals. In addition to Camryn, Emily describes here *"I came in as a health science major and I switch[ed] to biochemistry. And I switch[ed] to I wanna go to grad school instead of going to med school because I loved doing research a lot more than I liked the idea of med school."* Despite pursuing a career in a STEM field, one student, Monday, described how she did not feel that research fit into her future goals because she would prefer a more applied experience. Consistent with their career goals, many students expressed a desire to research in the future, either before their undergraduate graduation, after as they pursue their careers, or both. Finally, one student, Susan, described how their research was a resume boost for them. Several students described this as an expectancy outcome, but Susan described it as a reflection of her participation in research.

Expectancy Outcomes

The expectancy outcome codes are future-focused and distinct from *What you dream* codes because they describe situations the student *expects* to happen if they participate in undergraduate research and have not occurred yet. All of the students with disabilities and most of the LGBTQ+ students, indicated outcome expectations of seeking a mentor. As was the case when Cee explained the main reason for her entering research.

> *"I wanted to find like a mentor. I wanted to find like a professor that will have a relationship with me because that's kind of difficult in all the classes that I've had. They're all like 100 plus people classes, and I'm not the type of person to barge into office hours just to hang out…because I need like letters of recommendation, for wherever I'm going and I was like, I don't really have anyone. Yeah, then, at this point I was a junior it's like I haven't had a single professor that I can really say that like would give me anything."* -Cee

Other remaining career-related expectations included resume boosts and learning more about future careers and methods used in them.

Additional expectancy outcomes included action expectations, described by Josie, Susan, and Britana, and included cases in which the participant described specific outcomes they were seeking in a research experience. Also, affective expectations were described by Camryn and Cee that included desire for research experiences which contribute to affective constructs (e.g., sense of belonging).

Discussion

This study is rooted in a theoretical model which combined Science Capital and SCCT (Fig. 4.1.) Science capital (inner ring of Fig. 4.1) has an individual focus, allowing researchers to explore factors directly affecting the participants, meanwhile, SCCT (and specifically the construct of Expectancy Outcomes utilized in this study) has a broader focus (outer ring of Fig. 4.1). The combination of the two theories captures many of the influences involved in student entry into UREs. This study utilized the combination of these theories to explore women with concealable identities experiences surrounding URE participation. At a surface level, these students exhibit similar levels of each area of Science Capital to their peers (Fig. 4.4; Table 4.1). However, when interviewed and results were further examined utilizing both theories, it becomes apparent that though these students are reporting similar levels of Science Capital, they do so in differing ways, this is demonstrated with the five major subthemes in Figure 4.5. Further understanding of the pathways students with concealable identities take when entering undergraduate research spaces will help institutions improve the equity of their UREs.

Bourdieu (1986) described several forms of capital, which were the basis for Archer et al. (2015) Science Capital, one of the theoretical frameworks utilized in this study. A limitation of Bourdieu's conceptualization of capital is the deficit mindset it presents, particularly when discussing cultural capital. Bourdieu frames cultures as a hierarchy, with some cultures possessing more or less cultural capital than others. In response to this, Yosso (2005) designed a theory of Community Cultural Wealth with an asset-based mindset, where each members of different communities possess *different* capitals, instead of more or less (further discussion in Chapter One). Archer et al. (2015) considered several theories that had reframed the ideology of Bourdieu's capital when designing Science Capital. The results of this study, though CCW was not specifically used, closely match the mindset produced by Yosso's theory.

Navigating Concealable Identities

Students with concealable identities expressed each form of Undergraduate Research Science Capital differently from their peers. Members of the LGBTQ+ community and

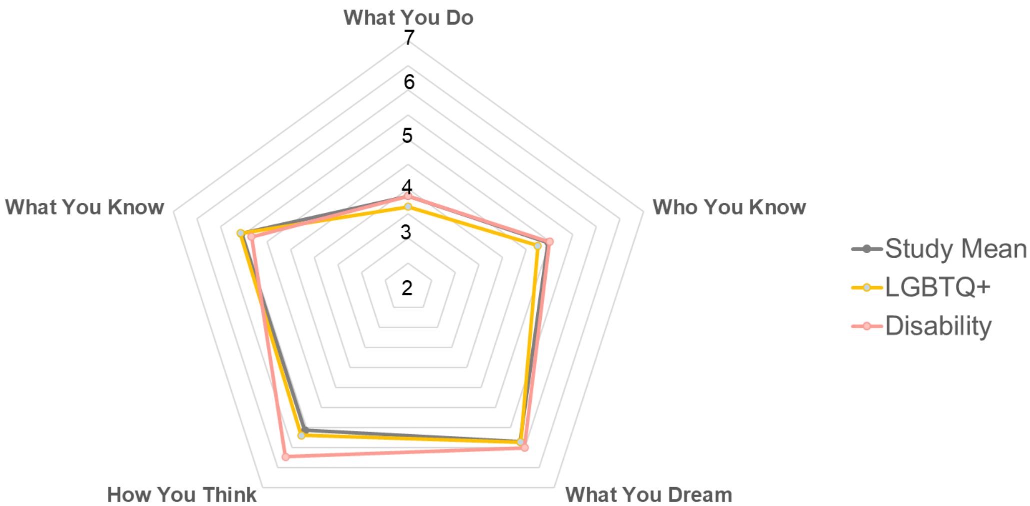


Figure 4.4: Average response of interview respondents to each form of capital across concealable identities.

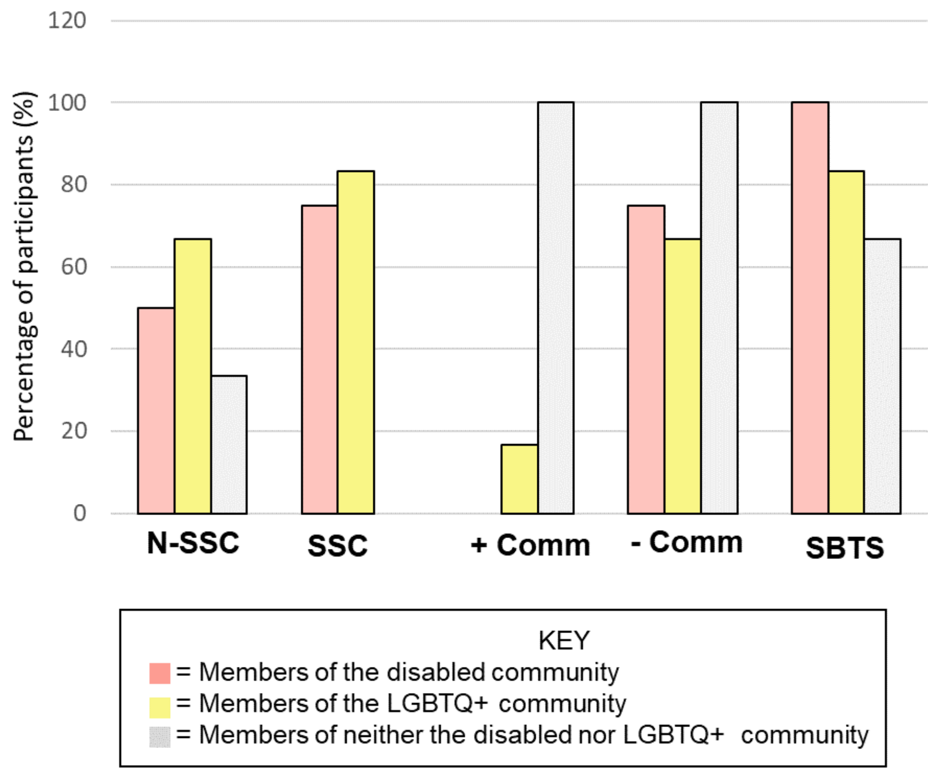


Figure 4.5: Representation of major subthemes across interview population groups. Percentage of response indicate participants who are members of that group who mentioned each code at least once. N-SSC – Non-Structural social capital, SSC – Structural social capital, +Comm – Positive communication about research, -Com – Negative communication about research, SBTS – Structural barriers to science.

Students with disabilities more frequently described instances of using structural social capital as opposed to non-structural. This is consistent with the findings of Whitehead (2019) who interviewed Black LGBTQ+ students about the forms of capital they access when navigating their community college experiences. They found that students often accessed campus resources and clubs and developed social capital in places that they are confident they can feel safe in. This is similar to Josie's description of finding support for her LGBTQ+ identity in the mentoring program she was a part of. For the ten students in this study, the university structure appears to be an overall safe place, however some students stated that their concealable identities were not revealed, particularly in research settings.

Cee specifically mentioned how her disabilities and her status as a first-generation college student prevented her from participating in many research experiences, and she found an alternative option. Additionally, it is possible some of the structural social capital was built by the students navigating their concealable identities. If they had reached out for support from their institution previously, and they were successful in finding it, they may be more inclined to continue to trust and reach out to other forms of structural social capital.

Despite more frequently reporting structural social capital, more LGBTQ+ and students with disabilities described structural barriers to research than their peers. This included all students with disabilities describing incidences of structural barriers. Research experiences in the sciences are not always accessible and research mentors are not always aware of measures that can be taken to make their research spaces more accessible (Batty & Reilly, 2022; Gin et al., 2022). Two of the students with disabilities reported that they only were able to participate in the research experiences because they were virtual, and accommodations were offered to them. Though many research experiences were put online for COVID-19, and students in this study only described COVID as a barrier to research, there are accessibility lessons that can be learned from the pandemic that may make UREs more accessible for students (Erickson et al., 2022). Additionally, Cee described not being sure if her disabilities "counted".

> *"It's weird because I don't know, like I kind of exist on like the boundary of disability. I know ADHD is a disability, but I still feel fake, saying that even though it is very debilitating sometimes. And then I have, like a few chronic illnesses, and so that definitely influenced my like research experience prior to this, because that first semester that I was in that group, I was sick the entire semester, like I had pneumonia 3 times. I could barely get out of bed some…most days actually and I was like…it was a lot."* -Cee

Because concealable identities are not always clearly defined, students may not be certain if they are eligible to access the resources that are available at their institutions. Individual research mentors, science departments, and supporting institutions must make a conscious effort to ensure the accessibility of these opportunities for science students.

Communication has been found to be an important influence to URE participation for science students (Bangera & Brownell, 2014; Cooper et al., 2021). Students do not always have the same exposure to available opportunities and may need to be shown what research opportunities are available to them. Consistent with the structural barrier findings, students with concealable identities reported positive communication about research opportunities less often than their peers. Additionally, there were high levels of reports of negative communication throughout the study in all groups of students (LGBTQ+, students with disabilities, peers, etc.). Encouraging faculty and research mentors to advertise available research positions and state desired qualifications for research opportunities would be highly beneficial in promoting the equity of availability of research experiences.

Representation in STEM Research

Students who are members of the LGBTQ+ community included descriptions of participating in UREs as a means of finding a mentor and increasing sense of belonging. For students with concealable identities, finding a mentor may be especially difficult. Effective mentorship includes providing students with psychosocial support, which Cee and Susan both described. This support has been shown to increase STEM recruitment and retention rates (NASEM, 2019). Additionally, studies have suggested that students with stigmatized race and/or gender identities may benefit from mentors that share those identities (NASEM, 2019). However, when identities are concealed it can be difficult to locate a mentor with that shared identity (as Cee expressed), especially since both students and mentors have been found to be less likely to reveal their concealed identities in professional settings (Cooper et al., 2019; NASEM, 2019; Yoder & Mattheis, 2016). Britana described that she never revealed her sexuality to her research mentor because it did not come up in the research space, and Josie mentioned that though she revealed her sexuality when asked directly in the mentoring experience she participated in, she did not reveal it in research group settings. Representation matters for all minoritized identities, visible or concealable. Studies have shown that openly LGBTQ+ instructors can serve as especially positive potential role models for LGBTQ+ students (Cooper, Brownell, et al., 2019). Students seeing and interacting with mentors that share identities with them can help them see their successful potential in STEM fields.

Conclusion

When compared to their peers, students with concealable identities reported similar levels of each area of Science Capital. However, they did so in differing ways from their peers without concealable identities (Figure 4.6). These differing levels of capital remained consistent across identity groups, rather than research participation or number of opportunities presented. This highlights the need for multiple approaches of recruitment efforts into UREs. Additionally, students with concealable identities expressed more SCCT related expectancy outcomes as being an influence on their potential research participation than their peers. The findings of this study suggest methods by which students with concealable identities, particularly LGBTQ+ students and students with disabilities can be supported to provide entry into UREs and a gateway to all of the positive outcomes they can provide.

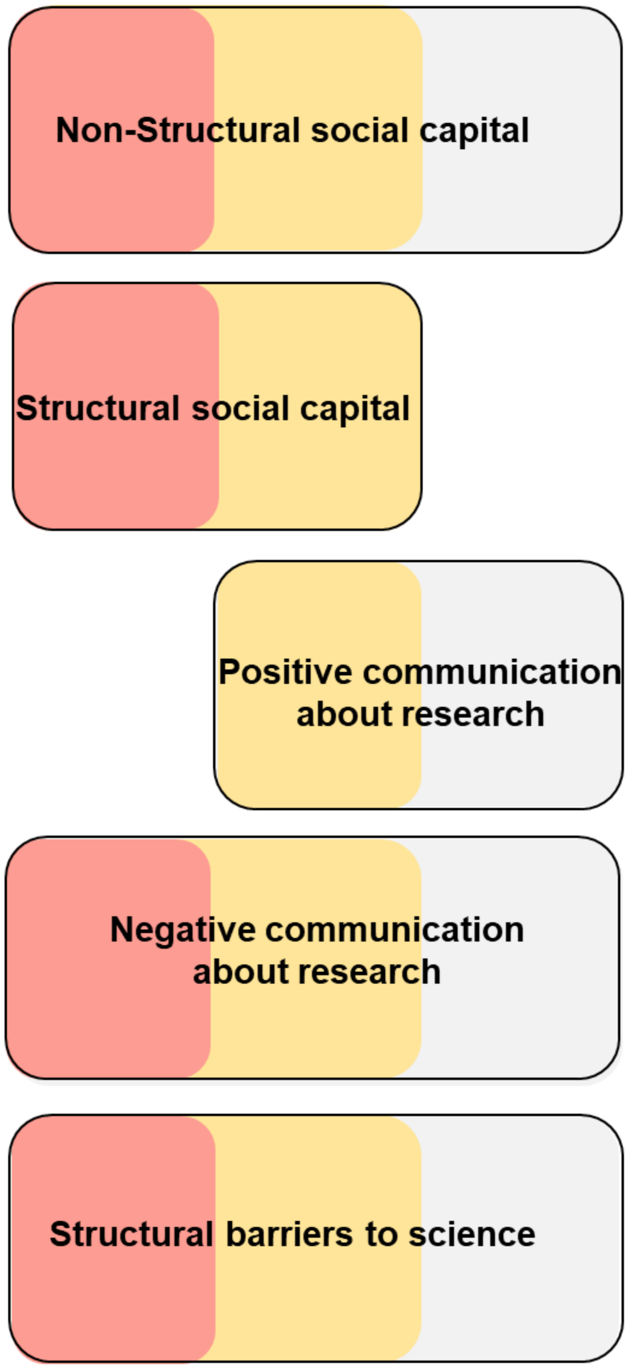


Figure 4.6: Distribution of responses for each responding community across major subthemes. Students with disabilities are represented by rose shading, LGBTQ+ students are represented with yellow shading, and students that not members of either community are represented with white shading.

Acknowledgments

This work was approved exempt by the university's IRB office. It was supported in part by a Geological Society of America Graduate Student Research Grant. The authors would like to thank Shannon Conner, Cole Bowman, Gavin Gleasman, Dr. Brian Dominy, Dr. Marian Kennedy, and Dr. Bridget Trogden for their support.

CHAPTER FIVE

RECOMMENDATIONS AND CONCLUSION

Implications for Practice

Many of the findings of this study can be directly applied by science departments and their respective institutions to improve access to their UREs. A potential outcome of this study is to assist departments in lowering barriers so that students can decide what opportunities they intend to pursue. Amongst the most common influencing factors identified in this study are requests from students for more communication about the ways to get involved in UREs. Promoting clear communication between instructors and students about what UREs are available, and the potential outcomes for students will be highly beneficial in improving access of these high-impact experiences for all students.

Mixing Quantitative and Qualitative Data

Survey and interview data suggest several influencing factors that are impactful to students' URE participation. Identified influencing factors fall into five categories of Undergraduate Research Science Capital: *How you think, Who you know, What you know, What you do,* and *What you dream*. The areas in which the survey identified opportunities and barriers intersect with the five major interview subthemes are displayed in Table 5.1. This can provide further insight into areas for improvement in the entry process into UREs. For example, professor influence and communication are some of the most frequently mentioned opportunities by both survey respondents and interview participants. However,

Table 5.1. Mixing of results. Areas in which the survey identified opportunities and barriers intersect with the five major subthemes identified in the interviews are indicated by an X. The five major qualitative themes and the influences that were categorized as opportunities and barriers are presented as they are the most likely to contribute significantly to student entry into undergraduate research experiences

		Interview/Qualitative Data				
		Structural social capital	Non-Structural social capital	Structural barriers to science	Positive communication about research	Negative communication about research
Survey/Quantitative Data	Opportunities					
	Professor Influence	X			X	X
	Interest	X	X		X	X
	Goals	X	X		X	X
	Major	X	X	X		
	Barriers					
	Accessibility	X		X		
	COVID-19			X		
	Job		X		X	X
	Travel			X		

none of the interviewed students that were members of the LGBTQ+ or disability community mentioned receiving positive communication about research (Table 5.2). Promoting communication with students about available research opportunities early and often, and the variety of experiences available to them (e.g., UREs in other majors, at other institutions) is a tangible way to create equitable research pathways at higher education institutions.

Departments interested in supporting their students' *Who you know* and *What you know* capital can provide faculty (and/or future faculty) development opportunities for their instructors and graduate students to encourage positive communication about available UREs and assist in the development of research curriculum and mentoring practices. Without proper training, mentors may not be aware of best practices to help their students succeed and can become overwhelmed developing UREs in addition to their other job responsibilities (Baker et al., 2015; Limeri et al., 2019; Lopatto et al., 2014). Additionally, it is important for departments and faculty alike to acknowledge the potential for negative and missed communication about research and promote positive interactions between students and faculty whenever possible.

As discussed in Chapter Four, though students across many demographic groups may exhibit similar levels of Science Capital, oftentimes they do so in different ways (Figure 4.6). Though this study analyzes the concealable identities of being a member of the LGBTQ+ and/or disability community specifically, other communities when viewed individually also reveal specific needs for those students. For example, in a conference

Table 5.2. Mixing of results for students with concealable identities. Areas in which the survey identified opportunities and barriers intersect with the five major subthemes identified in the interviews are indicated by an X. The five major qualitative themes and the influences that were categorized as opportunities and barriers are presented as they are the most likely to contribute significantly to student entry into undergraduate research experiences. Grayed column indicates capital identified by students without concealable identities (Table 5.1) which is not shared by students with concealable identities.

		Interview/Qualitative Data				
		Structural social capital	Non-Structural social capital	Structural barriers to science	Positive communication about research	Negative communication about research
Survey/Quantitative Data	Opportunities					
	Professor Influence	X				X
	Interest	X	X			X
	Goals	X	X			X
	Major	X	X	X		
	Barriers					
	Accessibility	X		X		
	COVID-19			X		
	Job		X			X
	Travel			X		

paper utilizing the survey data from this study, it was found that transfer students seek clear connections to their future goals and careers and find these connections more beneficial than non-transfer students when considering participation in UREs. This is likely due in part to transfer students' especially stringent course schedules inhibiting them from pursuing some engaged learning opportunities (Boyd & Lazar, 2023). These findings highlight the importance of disaggregating data in research and specifically analyzing marginalized groups (Trogden et al., 2022). Without disaggregation, data for marginalized groups can often become hidden by the size of the dataset as a whole. Additionally, these findings promote the necessity for various approaches for equity improvements, as a one size fits all approach will not be beneficial for all students.

Implications for Research

This study examines an understudied area of undergraduate research, student pathways surrounding UREs, with a focus on improving the equity of the process. Planned future work includes further refinement of the undergraduate research science capital scale using a broader undergraduate population. An undergraduate research science capital scale could be utilized by both researchers and practitioners to continue to explore the ways that their students access undergraduate research. Additionally, the scale could be revised to be applicable to other high-impact practices and engaged learning experiences (e.g., internships, service learning, and global learning experiences).

Topological data analysis (TDA) is a quantitative methodology that allows researchers to analyze highly dimensional data with more nuance than many clustering methods (Godwin et al., 2019). Thus far, it has only been used in a handful of educational research studies (Boyd et al., 2023; Doyle, 2017; Stevens, 2016). This study expands upon the use of TDA in discipline-based education research (DBER) and introduces a big data approach to education research.

Summary

The overarching research question of this study is: *How can science departments, and their respective institutions, improve access to research experiences amongst their enrolled undergraduate students?*

Analysis of survey demographic data and publicly available datasets reveals a disproportionately high number of students with a disability and LGBTQ+ students participating in research. Though consistent with literature findings, it is still perplexing given the increase in barriers these students face when navigating their identities (Hughes, 2018). Interview data reveals a potential cause of this to be that these students are seeking a place to belong and view research experiences as a place to strengthen relationships with mentors and a research group.

It is important to note that though previous studies found disproportionately high research participation rates for STEM LGBTQ+ students, they also found that these students did not complete their degrees at higher rates than their peers (Hughes, 2018). A suggested explanation for this has been that negative experiences with mentors and/or research groups may have an opposite effect on students and left them feeling as if they did not belong in STEM (Freeman, 2018). Efforts to ensure a positive and safe research environment should be made by all departments to protect students and continue promoting participation in science for all.

An additional potential explanation for the disproportionately high URE participation rates is the lack of available data across science that accurately represents students' sexuality and disability status contributing to the appearance of overrepresentation. Efforts to improve these national datasets to provide a clearer picture of student participation in science fields are being promoted by STEM researchers and organizations across the United States (AERA, 2022; Freeman, 2020).

Topological data analysis produced two distinct groups of survey participants, those who reported a high number of opportunities (HO) and those who reported a low number of opportunities (LO). Respondents who were identified as HO students identified professor influence, interest, major, career goals, and graduate/professional school goals as opportunities to their undergraduate research participation. The average scores of their LO peers were significantly lower across all of these factors and did not have any influences identified as opportunities.

Survey responses identified an opportunity to analyze the understudied area of those with concealable identities and their entry to participation in undergraduate research. Ten participants were purposefully selected based on their membership in the TDA groups and their self-identified demographic factors. Interview data suggests that though students may express similar amounts of each area of Undergraduate Research Science Capital, they may present their capital in different ways influenced by areas of their identity. Students with potentially concealed identities may not receive support as these identities are not known, highlighting the importance of varied approaches to promoting undergraduate research participation.

Equity scholar Jamila Dugan states “Equity isn't a destination but an unwavering commitment to a journey” (Dugan, 2021). This multi-institutional study fills several literature gaps in the field of undergraduate research experiences for science majors. Analysis of potential opportunities and barriers to research participation is provided along with suggestions for researchers and practitioners in science departments and their respective institutions to promote the equity of entry into UREs. Thereby providing a step in the journey towards greater inclusivity in science for all.

APPENDICES

Appendix A

Undergraduate Research Science Capital Survey

The Council on Undergraduate Research (CUR) defines undergraduate research as:
"A mentored investigation or creative inquiry conducted by undergraduates that seeks to make a scholarly or artistic contribution to knowledge."
Undergraduate research experiences can be any research activities you have participated in your time in college. These are often either in classes (sometimes called CURES); for course credit and/or pay with a professor, graduate student, or other mentor; or summer experiences (sometimes called REUs). The next few questions will ask about your undergraduate research experiences. If you have not had any undergraduate research experiences, please respond as such.

What year in college were you when you started participating in research for the first time? (If you started in the summer, then please select the next college year)
o 1st
o 2nd
o 3rd
o 4th
o 5th+
o I have not participated in research yet, but I plan to in the future.
o I have not participated in research yet and I do not plan to.

Including any you are participating in currently, how many research experiences have you had?
o 0
o 1
o 2
o 3
o 4+

What kind of experiences were they (Select all that apply)? *(Only displayed for students who indicated previous research participation)*
▢ Course-based (research in a class)
▢ Summer
▢ In a lab for credit or pay (not during the summer)
▢ Volunteer (no course credit or pay)
▢ Other. Please describe in textbox

On a scale of 1 (Not at all interested) - 5 (Extremely interested), how interested are you in participating in undergraduate research? *(Only displayed for students who HAD NOT indicated previous research participation)*

o 5 - Extremely interested
o 4 - Interested
o 3 - Neutral
o 2 - Not interested but might change mind
o 1 - Not at all interested

Which of the following has influenced your lack of participation in undergraduate research? Please select all that apply. *(Only displayed for students who HAD NOT indicated previous research participation)*

▢ I would prefer to participate in an internship/Co-op.
▢ I was/am not aware of research opportunities available to me.
▢ I do not have time in my schedule.
▢ I am not interested in doing research.
▢ I have never considered participating in research.
▢ Research opportunities available to me do not pay well or do not pay at all.
Other. Please describe in the textbox

Do you hope to/have plans to participate in any research experiences in the future?

o Yes
o No
o Unsure/prefer not to answer

The next questions will help us identify influences that could be considered opportunities or barriers to undergraduate research participation. On a scale of (1) Extremely negative impact to (7) Extremely positive impact, how much of an impact did the following things have on your ability to participate in undergraduate research? Please use NA to indicate any that did not have an effect on you.

Outside responsibilities – Responsibilities that may influence your ability to participate in undergraduate research.

NA 1 Extremely Negative 2 Very Negative 3 Negative 4 Neutral 5 Positive 6 Very Positive 7 Extremely Positive

- Family obligations - Family can be biological or chosen. (e.g., care responsibilities, driving family members places, etc.)
- Work-Jobs outside of your research responsibilities
- Athletics – School sponsored athletic obligations (NCAA, intramural, club, etc.)
- Religious obligations
- Social obligations – Activities outside of those already mentioned that may influence your ability to participate in undergraduate research (e.g., Greek life, clubs, friends)
- Other (Please describe)

Influential people - Interactions with others that may influence your participation in undergraduate research.

NA 1 Extremely Negative 2 Very Negative 3 Negative 4 Neutral 5 Positive 6 Very Positive 7 Extremely Positive

- Professors – Interactions in or outside of class
- Teachers (from K-12)
- Academic Advisors - Interactions in or outside of official advising time
- Other Students
- Office of Undergraduate Research - If your school has one, if not, or you don't know, mark NA
- Family Members – Family can be biological or chosen
- Other Mentors- Anyone you consider a mentor that has not been previously listed
- Other (Please describe)

Courses – Classes that may influence your participation in undergraduate research (for example, requirements to graduate or courses whose content influenced your abilities to participate.)

NA 1 Extremely Negative 2 Very Negative 3 Negative 4 Neutral 5 Positive 6 Very Positive 7 Extremely Positive

- Major Courses – classes within your major(s)
- Other courses outside of major(s)

Future goals - Goals that may influence your decision to participate in undergraduate research.

- NA 1 Extremely Negative 2 Very Negative 3 Negative 4 Neutral 5 Positive 6 Very Positive 7 Extremely Positive
- Career Goals
- Grad/Professional School Goals
- Other (Please describe)

Interest - Your interest in participating in undergraduate research.
NA 1 Extremely Negative 2 Very Negative 3 Negative 4 Neutral 5 Positive 6 Very Positive 7 Extremely Positive
Interest in research

- Interest in science generally
- Interest in solving real world problems
- Interest in exploring new ideas
- Interest in learning new skills
- Interest in questioning misconceptions

Opportunity - Impacts on your ability to participate in research experiences.
NA 1 Extremely Negative 2 Very Negative 3 Negative 4 Neutral 5 Positive 6 Very Positive 7 Extremely Positive

- Finding a research opportunity
- Awareness of research opportunities
- Your GPA
- Your Major
- COVID-19 – Anything COVID related in the past or current (e.g., research being online, restrictions in place because of COVID)
- Disability Limitations – Any disability you identify with
- Travel - Transportation to and from research locations

Free Response Questions:

- Would you like to share more details about any of your above responses?
- Are there any other influences that impacted your interest and/or ability to participate in undergraduate research? If so please describe them and the extent to

which they had an impact. This is to capture any influences that may have been previously missed in the survey.

- How did you become involved in undergraduate research? Please describe. *(Only displayed for students who indicated previous research participation)*
- What are the major reasons that contributed to you not participating in undergraduate research? Please describe. *(Only displayed for students who HAD NOT indicated previous research participation)*

Appendix B

Undergraduate Research Science Capital Survey Codebook

Participant ID - xxyyzzz

XX - University	YY - Major Type/research type	ZZZ - number
[Redacted]	00 - No research	incomplete - incomplete survey
	01 - Astronomy/Physics	
	02 - Chemistry	
	03 - Earth Science	
	04- Life Science	
	05 - more than 1/other	
	06 - Not Science	

What year in college were you when you started participating in research for the first time?
(If you started in the summer, then please select the next college year)
1 1st
2 2nd
3 3rd
4 4th
5 5th+
6 I have not participated in research yet, but I plan to in the future.
7 I have not participated in research yet and I do not plan to.

On a scale of 1 (Not at all interested) - 5 (Extremely interested), how interested are you in participating in undergraduate research?
1 Not at all interested
2 Not interested but might change mind
3 Neutral
4 Interested
5 Extremely Interested

Yes/No questions

0	No
1	Yes
2	Unsure
3	Prefer not to answer

Matrix

1	Extremely Negative
2	Very Negative
3	Negative
4	Neutral
5	Positive
6	Very Positive
7	Extremely Positive

Demographic Codes

Major

Code	Major	Category
1	Astronomy	Astronomy/Physics
2	Physics	Astronomy/Physics
3	Chemistry	Chemistry
4	Geology	Earth Science
5	Marine Science	Life Science
6	Meteorology	Dependent on college
7	Biological Sciences	Life Science
8	Biochemistry and Molecular Biology	Life Science
9	Botany	Life Science
10	Entomology	Life Science
11	Genetics	Life Science
12	Microbiology	Life Science
13	Zoology	Life Science
14	Other	
15	Undeclared/Undecided	
16	Math	2nd major only
17	Animal Science	2nd major only
18	Biosystems Engineering	2nd major only
19	Computer Science	2nd major only
20	Environmental Science	Earth Science
21	Psychology	2nd major only
22	Anthropology	2nd major only
23	General Sciences	

Sexual Identity

0 Prefer not to say
1 Aromantic
2 Bisexual/Pansexual
3 Gay
4 Lesbian
5 Queer
6 Questioning/Unsure
7 Straight/heterosexual
8 Not listed

Age

1 18-24
2 25+
2 Female
3 Non-Binary
4 Not Stated

Race/Ethnicity

0 Prefer not to say
1 Asian
American Indian/Alaskan
2 Native
3 Black or African American
4 Hispanic and/or Latino/a/x
Native Hawaiian or Pacific
5 Islander
6 White
7 Not Listed

Family Degree

0	Prefer not to say
1	Did not complete HS
2	HS/GED
3	Did not complete post-secondary
4	Technical or professional cert
5	Bachelor's Degree
6	Did not complete grad school
7	Master's Degree
8	Doctorate Non-STEM
9	Doctorate STEM
10	Other
11	Unsure

Appendix C

Science Majors for Inclusion in the Study by CIP Code

CIP Code 26 – Biological and Biomedical Sciences
26.01) Biology, General 26.02) Biochemistry, Biophysics and Molecular Biology 26.03) Botany/Plant Biology 26.04) Cell/Cellular Biology and Anatomical Sciences 26.05) Microbiological Sciences and Immunology 26.07) Zoology/Animal Biology 26.08) Genetics 26.09) Physiology, Pathology and Related Sciences 26.10) Pharmacology and Toxicology 26.11) Biomathematics and Bioinformatics 26.12) Biotechnology 26.13) Ecology, Evolution, Systematics, and Population Biology 26.14) Molecular Medicine 26.15) Neurobiology and Neurosciences 26.99) Biological and Biomedical Sciences, Other
CIP Code 40 – Physical Sciences
40.01) Physical Sciences 40.02) Astronomy and Astrophysics 40.04) Atmospheric Sciences and Meteorology 40.05) Chemistry 40.06) Geological and Earth Sciences/Geosciences 40.08) Physics 40.10) Materials Science 40.99) Physical Sciences, Other

Appendix D

Legitimation

Legitimation Type	Description from Onwegbuzie et al.[32]	Presence in this study
Sample Integration	The extent to which the relationship between the quantitative and qualitative sampling designs yields quality meta-inferences	Interview participants were purposefully selected from quantitative TDA results
Inside-outside	The extent to which the researcher accurately presents and appropriately utilizes the insider's view and the observer's views for purposes such as description and explanation.	A) Survey instruments were vetted by other researchers and members of the participating population before administration. B) All qualitative analysis was checked by the same secondary coder who is familiar with the study, the theoretical frameworks involved, and qualitative coding methodologies.
Weakness minimization	The extent to which the weakness from one approach is compensated by the strengths from the other approach.	A) The survey contains both quantitative and qualitative portions. Quantitative survey questions help participants see options that they might not realize in interviews and qualitative portions allow for free response. B) Interviews of 10 participants allow for more description than the survey alone but would not be feasible for all survey respondents.

Legitimation Continued		
Legitimation Type	Description from Onwegbuzie et al.[32]	Presence in this study
Sequential	The extent to which one has minimized the potential problem wherein the meta-inferences could be affected by reversing the sequence of the quantitative and qualitative phases	Interview participants were selected from TDA formed groups. If the interviews had taken place before the quantitative analysis, there would be no way to incorporate those inferences into the formation of the TDA groups. Therefore, a sequential explanatory design is most appropriate for this study.
Conversion	The extent to which the quantitizing or qualitizing yields quality meta-inferences.	All participants had space for both quantitative and qualitative responses.
Multiple validities	The extent to which addressing legitimation of the quantitative and qualitative components of the study result from the use of quantitative, qualitative, and mixed validity types, yielding high quality meta-inferences.	There are multiple validities present throughout the study.

Appendix E

Quality Considerations

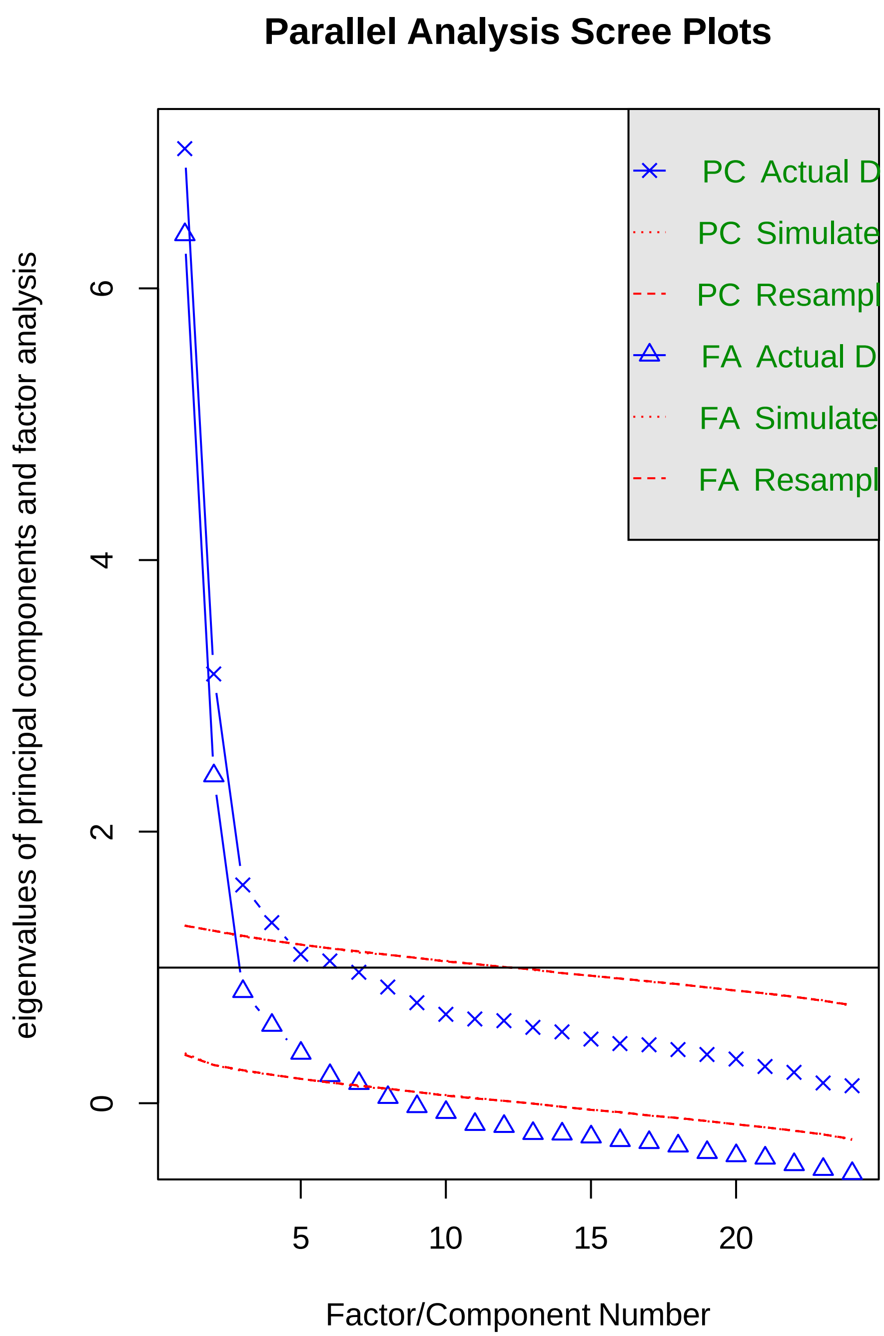
Adapted from Walther et al. (2013)

Research decision or action	Theoretical Validation		Procedural Validation		Communicative Validation		Pragmatic Validation		Ethical Validation		Process Reliability	
	Do the concepts of the theory appropriately match the investigation?		Which features of the study improve the fit between theory and reality?		Are the study constructs appropriately communicated to the intended audience?		Do concepts and claims withstand exposure to reality?		Are ethical practices being considered?		How can the research be made independent from random influence?	
	O	T	O	T	O	T	O	T	O	T	O	T
Making Data												
Modified already established theoretical frameworks	X						X					
Multiple theoretical frameworks used	X		X									
Surveys vetted before administration			X		X		X				X	
Surveys rooted in literature	X				X		X				X	
Mixed-methods design	X		X									
Intended diversity in respondents					X		X		X		X	
Researcher level of training in qualitative and quantitative methods	X	X	X	X	X		X		X		X	

Quality Considerations Continued												
	Theoretical Validation		Procedural Validation		Communicative Validation		Pragmatic Validation		Ethical Validation		Process Reliability	
	Do the concepts of the theory appropriately match the investigation?		Which features of the study improve the fit between theory and reality?		Are the study constructs appropriately communicated to the intended audience?		Do concepts and claims withstand exposure to reality?		Are ethical practices being considered?		How can the research be made independent from random influence?	
Research decision or action	O	T	O	T	O	T	O	T	O	T	O	T
Handling Data												
Member checking of all qualitative coding					X		X		X		X	
Deidentifying all surveys and interviews			X		X				X		X	
Multiple coding passes	X		X		X		X					
Ongoing engagement with data	X				X		X					

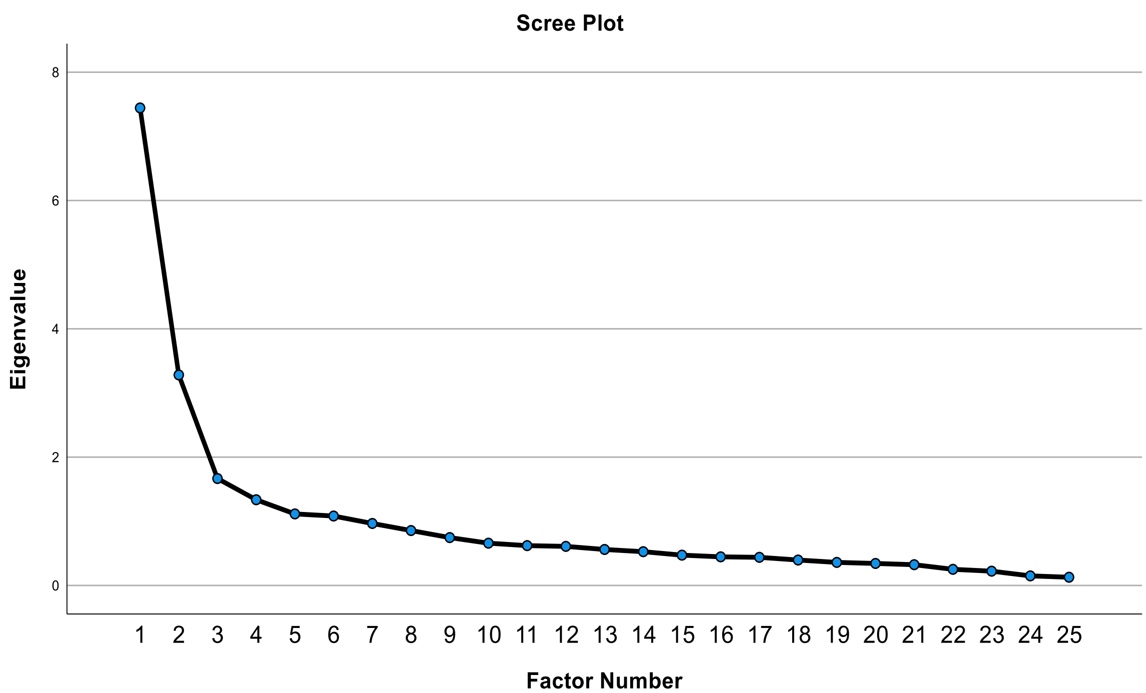
Appendix F

Parallel Analysis Scree Plot of Undergraduate Research Science Capital Survey



Appendix G

Scree Plot of Factoring for Undergraduate Research Science Capital Survey



Appendix H

Comparison of 1^{st} Year and 4^{th+} Year Student Responses to Undergraduate Research Science Capital Scale Items

	Researchers						Non-Researchers					
Influences	*p*-value	Cohen's d	1^{st} Year Mean	SD	4^{tth+} Year Mean	SD	*p*-value	Cohen's d	1^{st} Year Mean	SD	4^{tth+} Year Mean	SD
Professor Influence	.062	-.282	5.13	1.22	5.53	1.09	.272	-.098	4.95	1.17	5.07	1.18
Major	.110	-.221	5.15	1.21	5.40	1.09	.502	-.406	5.11	1.16	5.23	1.26
Interest in Research	.526	.239	5.36	1.03	5.50	1.31	.072	.037	5.06	1.34	4.59	1.72
Interest in Science	.115		5.44	1.15	5.77	1.23	.886	-.321	5.46	1.81	5.49	1.26
Interest in Solving Real World Problems	.083	.041	5.37	1.15	5.71	1.02	.149	-.520	5.38	1.15	5.63	1.19
Interest in Exploring New Ideas	.002*	-.198	4.98	1.28	5.63	1.06	.329	-.448	5.41	1.10	5.57	1.21
Career Goals	.344	.514	5.68	1.05	5.48	1.28	.540	-.217	5.60	1.20	5.48	1.36
Graduate/Professional School Goals	.417	.206	5.49	1.25	5.67	1.18	.041*	.339	5.27	1.56	5.06	1.34
Interest in Learning New Skills	.017	-.082	5.06	1.47	5.66	1.22	.385	-.436	5.52	1.05	5.67	1.19
Interest in Questioning Misconceptions	.182	-.165	4.92	1.41	5.16	1.47	.818	-.036	5.21	1.13	5.25	1.25
Family Responsibilities	.007*	.090	4.33	1.44	4.21	1.28	.313	-.160	4.19	1.40	4.41	1.45
Academic Advisor Influence	.005*	.160	5.35	1.29	4.74	1.05	.366	.099	4.75	1.11	4.59	1.14
Peer Influence	.362	-.192	5.13	1.43	5.93	1.06	.752	-.369	4.59	1.10	4.65	1.17
Family Influence	.007*	.137	5.27	1.06	4.68	1.27	.055	-.007	4.76	1.23	4.37	1.36
Other Mentors	.095	.704	5.23	1.24	4.85	1.09	.313	-.373	4.59	1.12	4.63	1.05
GPA	.494	-.123	4.96	1.37	5.13	1.27	.841	.031	4.21	1.39	3.65	1.59

Appendix H (Continued)												
Social Responsibilities	.220	.334	4.27	1.16	3.87	1.22	.010*	.399	4.72	1.42	5.11	1.16
K-12 Influence	.210	.166	4.63	1.39	4.42	1.16	.011*	.429	4.70	1.38	4.13	1.11
Office of Undergraduate Research	.010*	.123	4.88	1.18	4.24	1.27	.533	-.262	4.55	1.09	4.41	1.45
Job	.224	.230	3.98	1.07	3.69	1.41	.220	-.197	3.86	1.33	4.13	1.75
Athletics	.082	.387	3.75	1.01	3.28	1.39	.786	=.048	3.73	1.72	3.81	1.41
Religious Responsibilities	.087	.385	4.03	1.42	3.54	1.10	.702	-.068	3.62	1.41	3.72	1.52
COVID-19	.029	.026	3.87	1.29	3.27	1.59	.098	.320	3.66	1.24	3.25	1.54
Accessibility	1.00	-.483	3.52	1.28	3.52	1.37	.963	-.010	3.42	1.43	3.43	1.20
Travel	.367	-.207	3.87	1.36	3.64	1.21	.537	.103	3.94	1.27	3.80	1.55

Appendix I

GUIDELINES FOR PARAMETER SELECTION

Abstract

In an increasingly technology-centered world, education researchers seek to adapt new methodologies to gain a clearer understanding of the nuance in their data. Topological data analysis (TDA) is an emerging big data methodology in education research that allows researchers to consider multi-dimensional data. As an emerging methodology, there are few available resources for assisting researchers in adopting these methods. This manuscript provides guidelines and uses examples to promote the use of TDA in education research.

Introduction

Topological data analysis (TDA) is an algorithm used to visualize complex data with more nuance than traditional clustering methods (Chazal & Michel, 2016). Carlsson (2009), the designer of TDA, describes the ideal dataset for use with the methodology as having these characteristics: (1) qualitative information is necessary to fully understand the quantitative mappings, (2) metrics are not theoretically determined, (i.e., the theoretical framework does not indicate what pieces of the dataset need to be analyzed), (3) coordinates are not natural (i.e., there is not a clear mapping of the datapoints), and (4) summaries of the datasets as a whole are more valuable than studying any of the parameters individually. Though TDA has been used in a variety of fields (Serrano et al., 2020), it has only been used in a handful of educational studies to date (e.g., Boyd et al., 2023; Doyle, 2017; Stevens, 2016). Additionally, little guidance is available to assist educational researchers in deciding if their dataset (1) is appropriate for TDA and (2) how to optimize the TDA parameters so that the mapping output best describes their data. Thus far, methods papers have described what TDA is, how it works, how it could theoretically be applied to educational data (Munch, 2017), and some of the basics of application (Godwin et al., 2019). However, more information about the parameters for the algorithm and how to optimize mapping would be beneficial to further the advancement of this novel methodology in educational research settings. This paper aims to inform education researchers on the availability of this methodology and provide guidance on implementation.

The examples described here utilize a dataset from a survey of introductory geology students about their interest in geoscience (results presented in Boyd et al., 2023). The dataset includes 1,681 completed responses that were grouped using the Mapper TDA package in R (Singh et al., 2007) based on respondent interest in pursuing geoscience careers. These data are used only as an example of how different model parameters influence the TDA model outputs to illustrate how choices impact the model solution. Algorithm parameters utilized and recommendations for future use of TDA for survey data are described.

Determining a Suitable Dataset

Mapper, an algorithm which performs TDA in R, collects datapoints and assigns them coordinates in vector space. The algorithm then performs data reduction with those datapoints into points called nodes. Nodes are oftentimes connected by a line called an edge. These edges indicate places where there is overlap between the node membership, meaning there is at least one datapoint that is a part of both nodes. Nodes can be further grouped and analyzed by researchers, similar to the outputs of a Principal Component Analysis (PCA) or other clustering methods (Fig. I.1). Mapper was originally developed for use with numerical data (Singh et al., 2007). As such, the categorical data that results from most Likert-type educational surveys will not always result in a successful TDA mapping. A discussion of how to determine if a dataset is appropriate for TDA is provided below.

Dataset Size

A consistent struggle for quantitative researchers is determining how many survey responses are needed. Topological data analysis recommends a “large” number of data points however there are few guidelines about the lower and upper limits of the algorithm (Doyle, 2017). If there is not enough data (*N* is too small), meaningful maps cannot be created because the underlying distribution is under sampled. However, if *N* is increased too much, the algorithm essentially runs out of space and cannot properly discretize the data (Doyle, 2017). An acceptable dataset size lies in the balance of the number of survey responses, the number of variables (survey questions) selected for inclusion in the analysis, and the distance calculated between participants. The distance calculation is the first step

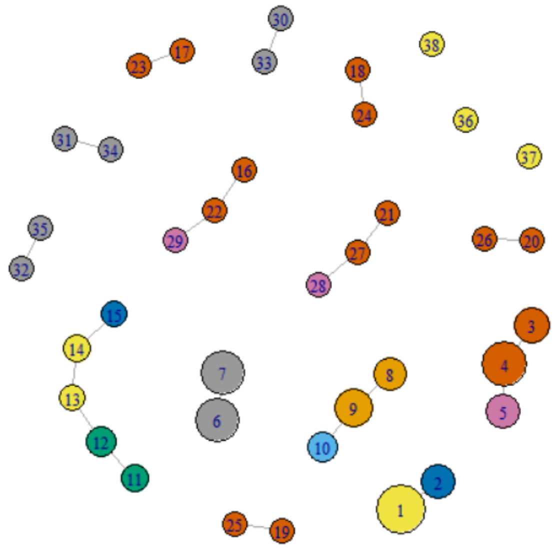


Figure I.1: Example of a successful TDA mapping.

of TDA once inputting the data is uploaded into R (further description in the *Using Mapper* section). A guideline for balancing the number of participants and number of variables included in the TDA has been established by Formann et al. (1980). They state that TDA datasets should follow the following equation:

$$N \approx 2d \qquad \text{(Eq. I.1)}$$

Where N is the total number of respondents needed to resolve the TDA for a given value of d, and d = the number of variables included in the analysis. With 1,681 participants in our example dataset, this formula recommends approximately eleven dimensions; ultimately twelve were selected based on several factors that will be discussed in the next section, *Variable Choice*. The guideline set by Formann et al. (1980) demonstrates the scalability of TDA, as the number of variables considered is influenced by the size of the dataset. This creates an avenue for education researchers to explore this big data approach with datasets of various sizes.

Variable Choice

Once an appropriate number of variables have been decided on based on the size of the dataset, the variables for analysis can be selected. A variable's variance, uniqueness, theoretical interest, and number of missing responses will affect the mapping of the data. Considerations for each of these characteristics are further described in this section.

The variable's variance describes the degree to which participant responses for that item differ from one another. The more variance there is between responses for that item, the better Mapper will be able to discretize that variable (Doyle, 2017). When using Likert-type items as variables, it is best if researchers choose variables that have the same, or at least similar, number of question anchors. This is because Mapper was designed for use with continuous numerical data not categorical data (Singh et al., 2007). When analyzing categorical data, the difference between a response of "3" and "4" carries different meaning on a five-point versus a seven-point scale. However, the algorithm is not designed to differentiate that. If use of items on different scales is necessary, it is recommended the researcher rescores values before inputting them into the algorithm and this should be noted as a potential limitation of that study. Additionally, because the maximum amount of variance is desired, the larger the Likert-scale available, the better. However, the number of anchors on the scale should still be balances with proper survey design techniques; if a scale is too large participants may have difficulty differentiating between responses (Artino et al., 2014). This allows more potential for variance and gives more space for discretization between variables (Doyle, 2017).

Uniqueness, separate from variance, describes if the particular survey item has responses that are different than expected from many of the other survey items. While variance can be calculated numerically (Iacobucci et al., 2022), uniqueness is a subjective measure determined by the researcher. As an example of an item with high uniqueness, if researchers hypothesized that participants would respond with agreeance on a Likert-scale, but the results averaged as "strongly disagree", that is a unique result compared to expectations. Similar to variance, the greater the uniqueness of an item the more fit it is for inclusion in a TDA (Doyle, 2017).

Another consideration when determining items for inclusion is the theoretical interest of the study. Many educational research studies are guided by theoretical frameworks, and though all survey items are of some theoretical interest, certain items may be essential to the analysis or more foundational points of the framework. Theoretical interest can be aided by quantitative measures (such as factor analysis); however, it too is mainly a subjective measure.

The final consideration when determining appropriate variables for a TDA is the number of missing responses. Oftentimes in survey collection participants skip questions or fail to complete a survey. If a certain percentage of the survey is completed, it may be included in analysis depending on the inclusion criteria of the study. For the purposes of TDA, the fewer missing responses, the more accurate the mapping of the data will be. There are algorithms that can approximate missing data responses (Rubin & Little, 2019); however, humans are not fully predictable. The most accurate response is that which comes from the participant themselves rather than from automation. Therefore, when considering items

It can be difficult for researchers to balance these considerations when trying to decide on appropriate variables to include in their analysis. To assist with this, in a dissertation study that involved using survey responses within a TDA, a decision matrix that weights the concepts of variance, uniqueness, and theoretical interest equally (note, the number of missing items was mentioned but not included in the matrix). Researchers would "score" variables of interest for each of those categories and include the items that obtained the highest scores (Table 14 in Doyle, 2017). Doyle (2017) states that this method was used as a starting point to determine items for inclusion, but some items were still included despite their scoring because they were theoretically vital to the study. Likewise, researchers should use their best judgement and consider all aspects of variable inclusion when deciding which items should be used in their TDA.

Methodology: Using Mapper

Once it is determined that a dataset is appropriate for a TDA and imported into R, there are additional research decisions that need to be made to complete the mapping. The base code for the Mapper algorithm requires inputs for the distance between objects, lens values, number of intervals, percent overlap and the number of bins when clustering (Chazal & Michel, 2016). Distance between objects can be calculated based on the researcher's choice of distance matrix. The greater the calculated distances between participant responses, the less data points you will need for a successful mapping. Because TDA was built for numerical data, and survey responses are typically ordinal/categorical, the distance calculation works differently. Boriah et al. (2008) explains the four main things that effect the distance calculations when using categorical data to be: (1) N – the number of participant responses, (2) d – the number of variables included, (3) n_k – the number of datapoints in each attribute (this is best accounted for by having a minimal number of missing datapoints in your dataset), and (4) distribution of $f_k(x)$ – how similar responses are across variables. Some algorithms will place more importance on "rare" values while others will place more importance on those most common in the dataset.

Topological data analysis is an offset of machine learning; as such, there are many possible choices for a distance function (Munch, 2017). A distance function is an objective value that summaries the relative difference between two datapoints. An example of a distance function is k-nearest neighbors (KNN) is the most commonly used classification with a Euclidian method. In KNN classification, an unlabeled datapoint is categorized by comparing it to surrounding datapoints within its k adjacent neighbors (adjacency calculated based on distance). Because k is variable, classification using KNN is not always straightforward (Khan et al., 2018). In TDA this may lead to more nodes being connected by edges because the exact node membership is not known. The KNN distances can be calculated in R and will remain constant regardless of mapping parameter choices as long as the decision of what distance function to use has not changed. The remaining parameters of lens choice, the number of intervals used, percent overlap, and the number of bins when clustering are at the researcher's discretion. Recommended guidelines are discussed below.

Lens Choice

The lens is a critical choice in the mapping of your data as it determines what variables algorithm places priority on when displaying the resulting mapping. Lenses can be mathematical (e.g., 1/distance) or theoretical (e.g., the responses for a survey item that was not included as a variable in the mapping), and more than one at a time can be selected. However, just because Mapper *can* use a certain lens to view your mapping data, does not mean it is a good choice for your analysis. The lens choice should have "intrinsic meaning based on the problem that you are looking at" (Bak, 2014). In other words, lens choice should be consistent with research questions and researchers should be able to explain the structure of the map based on lens. In the case of the example dataset, the research questions and selected variables explored students' geoscience recruitment potential. Therefore, the selected lens was the average response across the selected variables for each participant. A node of students with mostly higher than average responses indicates a higher recruitment potential based on the algorithmic output.

Number of Intervals

The number of intervals, or lens slices (*n*) should result in enough participants in each node to conduct further statistical analysis on your groups. Following Central Limit Theorem, this should be no smaller than 30 participants per interval (Godwin et al., 2019). If there are too many intervals, the mapping results in dyads (two nodes connected by an edge) of repeating participants and the underlying structure of the data becomes obscured. The number of intervals changes the number of participants per node and resolution of the resulting map but does not change the core interpretation of the map structure. Therefore, this parameter has much less impact on the final mapping than other modeling parameters (Godwin et al., 2019).

In the geoscience career interest example, 1681 participants were successfully mapped with 25 intervals resulting in approximately 67 participants per interval (Fig. I.1). Mapping with 30 intervals for approximately 56 participants per interval results in dyads that would be difficult to derive meaning from (Fig. I.2). Lowering *n* to 17, for 99 participants per interval, likewise changes the resolution such that the determination of groups of participants would be difficult (Fig. I.2). As displayed in Figure I.2, most of the nodes in this mapping are connected by edges. When forming groups based on TDA, researchers should include all interconnected points in the same group because it is unclear which node each datapoint (in the case of education research, participant) is closest to. Due to this, Figure I.2 would not have enough meaningful groups to guide further analysis.

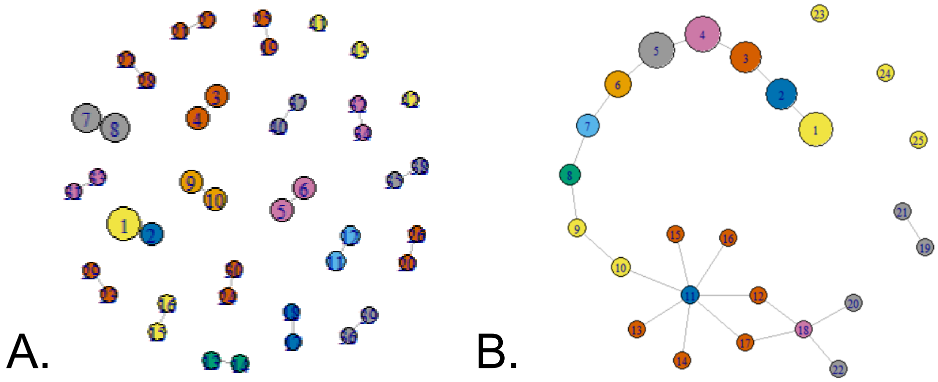


Figure I.2: Mappings with variation in the number of intervals. A) Mapping with 30 intervals, all other parameters remained constant. The majority of the map is dyads of nodes resulting in many potential groups but little meaning for education researchers. B) Mapping with 17 intervals, all other parameters remained constant. The majority of the map is one interconnected line of nodes resulting in few groups and reducing the meaning to education researchers.

Percent Overlap

As TDA applies the selected lens there is a certain amount of overlap between the applied filters. The percentage of overlap is determined by the researcher. Figure I.3 contains an example of four filters being applied to a dataset with 50% overlap and the resulting mapping (figure modified from Murugan & Robertson, 2019). Participant responses are grouped by which sub-ranges their filter value falls within and then clustered into nodes, creating a network. If two nodes have a common participant (due to the overlap in filter ranges), then they are connected in the map with an edge (Godwin et al., 2019). When mapping is complete, researchers can view the population of each node and remove duplicate responses systematically based on their research questions. Percent overlap affects the connectedness of the map. Higher overlap results in more connection between portions of the map, and a more gradual distribution between nodes. A lower overlap should be used if more groups of datapoints (a greater difference between nodes) is desired. When using categorical data in a TDA, a 50% overlap is typically recommended (Godwin et al., 2019).

Number of Bins When Clustering

Mapper's next step after applying the overlapping filters is to cluster the data into the final nodes. Despite the name of the parameter, the number of bins (ε) does not indicate the number of groups your participants will cluster into. The recommendation for the number of bins is to use follows the equation:

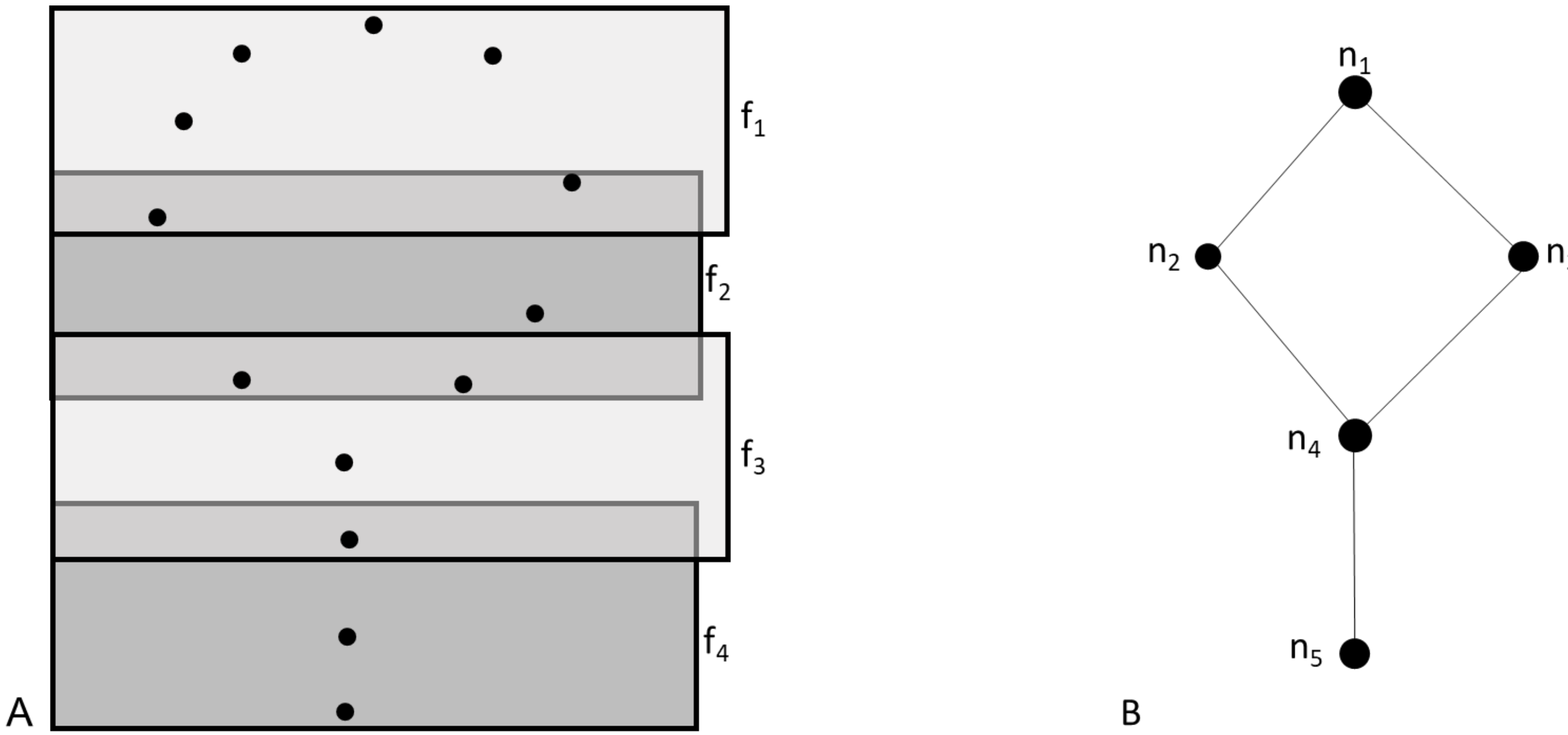


Figure I.3. Application of filters. A) The application of four filters (f_x) with 50% overlap. B) An example of the five nodes resulting from the application of these filters. Datapoints located in the overlap between two filters results in edges between the nodes. Figure modified from Murugan & Robertson, 2019.

$$\varepsilon = \sqrt{d} \qquad \text{(Eq. I.2)}$$

where d = number of variables included into the TDA; Godwin et al., 2019). As a reminder, the equation for the number of variables used in TDA for a given number of participants (i.e., dataset size) is $N \approx 2^d$ (Eq. I.1). Because researchers may have included a different number of variables than indicated by the equation, a more accurate estimate of the connection between these variables is:

$$N = 2^{\varepsilon^2} \qquad \text{(Eq. I.3)}$$

Based on observations with the example dataset, the fewer the number of bins, the more nodes appear which can cause challenges in later analysis of groups. However, this parameter seems to have little to no effect on the overall group membership of participants. The resolved mapping in Figure I.1 utilized five bins, while the same data clustered with three bins is indistinguishable (Fig. I.4), and twelve bins is very similar to five, with the exception of a few nodes in the outer field (Fig. I.4; the field is composed of six participants in this analysis that did not fit with any of the established groups).

Further Analysis

Once all the parameters have been input into Mapper, the algorithm will produce a map and a Mapper file in R. The locations and populations of each of the nodes will be displayed. Researchers should remove duplicates and assign each participant only one node as its location. Likely, several nodes will be combined to make groups for further statistical analysis. It is also possible that some participants will not fit into any of the groups and will make their own field of nodes. Researchers will need to determine if it is best to omit these participants or report their responses in some other manner. This field has the

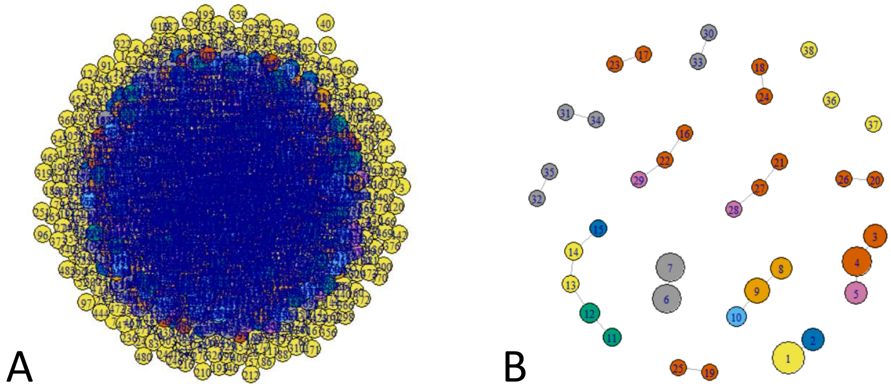


Figure I.4. Mappings with variation in number of bins. A) Mapping with 3 bins, all other parameters remained constant. B) Mapping with 12 bins, all other parameters remained constant.

potential to provide deep insight into education research datasets as it is where the most unique participant responses will sort. In the example of the geoscience recruitment study, it was these students that indicated high levels of recruitment potential but had unique interests compared to the remainder of the dataset. This allows the research team to make recommendations on novel ways to prepare recruitment strategies. When group populations are finalized, further statistical analysis such as t-tests, or ANOVAs can be run to determine the similarities and differences between groups.

Conclusion and Recommendations

Though established for use in many fields, TDA is a novel methodology for educational researchers. Topological data analysis can help education research answer the growing call for big data research methodologies (*NSF*, n.d.) It combines quantitative statistical analysis with qualitative research decisions and allows researchers to consider their theoretical frameworks throughout implementation, making it well suited for many studies involving categorical and/or ordinal data. Topological data analysis requires a fairly large dataset, which can be a limitation for many educational research studies. However, when using the correct parameters for the particular dataset, a successful mapping can be obtained with datasets as small as several hundred (Doyle & Potvin, 2015), and potentially even smaller. A great benefit of the method is that it is run on open-source R, and initial mappings take only moments to produce. If a mapping is unsuccessful, the researcher will know quickly and be able to adjust parameters accordingly. Researchers interested in a clustering method that allows for more nuance than more traditional methods should consider TDA for their work.

Appendix J

R Code Markdown for Topological Data Analysis

Dissertation Markdown

Abby Boyd

Included code is the final code necessary to produce the opportunity and barrier mapping. Several trials as well as statistic codes are available upon request but were not included for simplicity.

Packages

```
library(devtools)

#devtools::install_github("paultpearson/TDAmapper")

library(igraph)

#These are the packages. Do not worry if it looks like there are error messages.
```

Functions

```{r}

sidequestdist = dist(TDA_102722, method =‘euclidian’)

Ndata <- dim(Input_2)[1] #number of observations in the data set knndist <- rep(0, Ndata) #initialize vector of values to store distance to kth nearest neighbor for each point DD <- as.matrix() #use full form of matrix (to allow sorting of values) for (i in 1:Ndata){ #loop through all of the data points knndist[i] <- sort(DD[i,])[k+1] #select row i, sort it and then take the kth value (use k+1 because matrix includes test point with d=0) }

#Function to extract number of points in each vertex: Vsize <- function(map){ vertex.size <- rep(0,map$num_{v}ertices)for(iin1:map$num_vertices){ #points.in.vertex <- map$points_{i}n_{v}ertex[[i]]vertex.size[i] < -length((map$points_in_vertex[[i]])) } return(vertex.size) }

#function to calculate mean filter value for each vertex fvals <- function(map,filtervals){ fv <- rep(0,map$num_{v}ertices)for(iin1:map$num_vertices){ fv[i] <- mean(filtervals[ map$points_in_vertex[[i]] ])
} return(fv) }
```

Codes for Opportunity and Barrier Maps

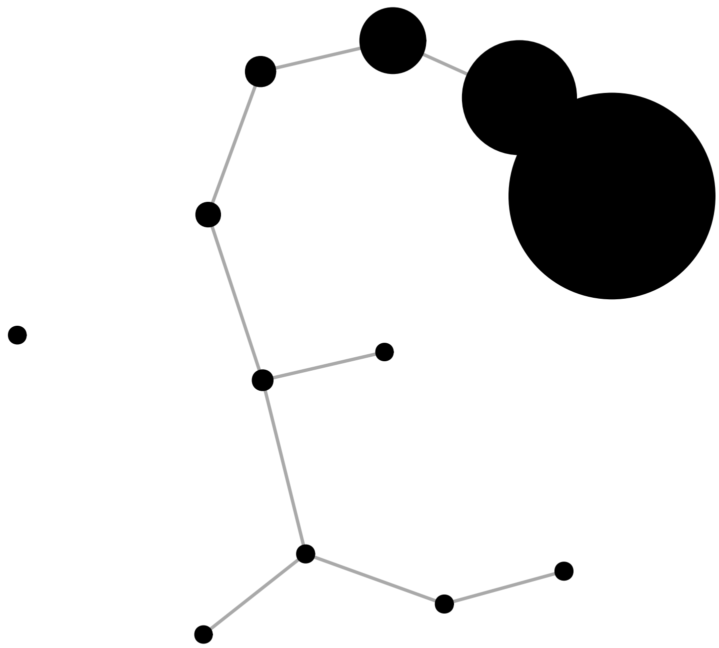
```
```{r}

#Oppourtunity
flens = oppanalysis
Dissertationoppsmap <- TDAmapper::mapper1D(
 distance_matrix = as.matrix(sidequestdist),
 filter_values= flens,
 num_intervals=26,
 percent_overlap=50,
 num_bins_when_clustering=6)
Vsize(Dissertationoppsmap)
fvals(Dissertationoppsmap,flens)
Dissertationoppsgraph <- graph.adjacency(Dissertationoppsmap$adjacency,
mode="undirected")
V(Dissertationoppsgraph)$size <- Vsize(Dissertationoppsmap)/10+5
V(Dissertationoppsgraph)$color <- 'black' #fvals(Dissertationoppsmap,flens)
V(Dissertationoppsgraph)$label<- ''
plot(Dissertationoppsgraph, layout = layout.auto(Dissertationoppsgraph) ); title('')

#Barrier
flens = BarrAnalysis
Dissertationbarrmap <- TDAmapper::mapper1D(
 distance_matrix = as.matrix(sidequestdist),
 filter_values= flens,
 num_intervals=9,
 percent_overlap=50,
 num_bins_when_clustering=4)
Vsize(Dissertationbarrmap)
fvals(Dissertationbarrmap,flens)
Dissertationbarrgraph <- graph.adjacency(Dissertationbarrmap$adjacency,
mode="undirected")
V(Dissertationbarrgraph)$size <- Vsize(Dissertationbarrmap)/10+5
V(Dissertationbarrgraph)$color <- 'black' #fvals(Sidequestmapper12,flens)
V(Dissertationbarrgraph)$label<-''
plot(Dissertationbarrgraph, layout = layout.auto(Dissertationbarrgraph) ); title('')
```
```

Appendix K

TDA map using number of barriers as a filter



Appendix L

Semi-Structured Interview Protocol

SEMI-STRUCTURED INTERVIEW PROTOCOL
Abby Boyd

RQ3: How do science students describe: a) proximal support and/or barriers they experienced and b) self-efficacy and outcome expectations related to participating in undergraduate research?

Interview questions were largely designed to surround Social Cognitive Career Theory (SCCT). However, SCCT's focus is mainly on outcomes, and how those outcomes can lead to continuing in experiences (or not), while the focus of the study is recruitment. Social Cognitive Career Theory is an important piece of the theoretical framework because it allows for the institutional opportunities and barriers that Science Capital misses. However, some interview questions will not focus on SCCT to ask recruitment focused questions.

Key
Researcher Participants
Non-Researcher Participants
All Participants

Semi-structured interview protocol continued

Interview Question	**Justification for Question** *(What do you want to learn from this question? Why are you asking this question? What elements of your theoretical framework(s) are addressed?)*	**Link to Research Question(s)**	**Potential Response Example(s)/Follow- Up questions**
1. What is research?	See what students perceive research to be and research experiences.		Looking for answers to questions
2. What research experiences are open for undergraduates to participate in?			Apprenticeship style, CURES, REUs, etc.
3. What are the costs and benefits you perceive for participating in undergraduate research experiences?			Application improvements
4. Have you participated in research as an undergraduate?			Yes or no
4.1 Tell me about any of these research experiences.	A way to see what types of experiences they have participated in		Apprenticeship style, CURES, REUs, etc.
5. How did you find out about the research experiences you participated in?	Identification/further description of factors that influenced research	Identifying factors for RQ3 but both RQ2&3	Friends Professors Are there other ways that you found out about research experiences?
6. What influenced your decision to participate OR not participate in a research experience?	Identification/further description of factors that influenced research	Identifying factors for RQ3 but both RQ2&3	Professors Graduation Requirements

7. I have listed that you responded on the survey that you identify as [Read off identities], did any of these things affect your participation in research?	Newly added question to help get at identify effects		
8. Was there anything that hindered your ability to participate in undergraduate research experiences? If so, how large of a barrier was this for you to overcome?	What barriers are present? Could be either SCCT or Science Capital	Do these barriers affect their expectations/participation?	Not enough time Needed a job Not "qualified" (actually or didn't think they were qualified)
9. Is there anything about your college experience that you feel had an impact on your ability to participate in research? (Exs: Student athlete, transfer student, had to work, etc)	Specifically seeing if any of these identified influences had a potential impact on participation	Identifying factors	Transfer students' schedules being tight
10. What are your future/career goals?	Outcome Expectations	RQ3	Med-School, grad school, straight to job, unsure
11. Has undergraduate research changed your future/career goals from what they were before you started?	Has the URE affected the expectations?	RQ3	Yes, now want to go to grad school No, stayed

12. Do you plan on pursuing any future research positions?	Outcome Expectations; expected persistence	RQ3	No, but continue in the current one Yes, want to do an REU
13. Is there anything about undergraduate research you wish you had been told beforehand? Any changes you wish the department or university would make regarding recruitment into research experiences?	Outcome expectations	RQ3	Didn't know some opportunities are paid Wish the department would advertise differently
14. Is there anything else you would like to share with us regarding your entry into undergraduate research experiences?	Good to close with anything else you would like to share		

Semi-structured interview protocol continued

Interview Question	**Justification for Question** *(What do you want to learn from this question? Why are you asking this question? What elements of your theoretical framework(s) are addressed?)*	**Link to Research Question(s)**	**Potential Response Example(s)**
1. What is research?	See what students perceive research to be and research experiences.		Looking for answers to questions
2. What research experiences are open for			Apprenticeship style, CURES, REUs, etc.

undergraduates to participate in?			
3. What are the costs and benefits you perceive for participating in undergraduate research experiences?			Application improvements
4. Have you participated in research as an undergraduate?			Yes or no
5. Do you plan on participating in undergraduate research before you graduate?	Outcome Expectations; expected persistence	RQ3	Yes, graduation requirements No, not interested
6. Have you heard about research opportunities? How have you heard about them?	Identification/further description of factors that influenced research	Identifying factors for RQ3 but both RQ2&3	Professors Graduation Requirements
7. **(7.1) IF YES**: Why didn't you participate in the research opportunities?	Identification/further description of barriers that influenced research	Identifying factors for RQ3 but both RQ2&3	Not enough time Needed a job Not "qualified" (actually or didn't think they were qualified)
8. I have listed that you responded on the survey that you identify as [Read off identities], did any of these things affect your participation in research?	Newly added question to help get at identify effects		I have listed that you responded on the survey that you identify as [Read off identities], did any of these things affect your participation in research?
9. Was there anything that hindered your ability to participate in undergraduate research experiences? If so, how large of a barrier was this for you to overcome?	What barriers are present? Could be either SCCT or Science Capital	Do these barriers affect their expectations/participation?	Not enough time Needed a job Not "qualified" (actually or didn't think they were qualified)

Is there anything about your college experience that you feel had an impact on your ability to participate in research? (Ex: Student athlete, transfer student, had to work, etc.)	Specifically seeing if any of these identified influences had a potential impact on participation	Identifying factors	Transfer students' schedules being tight
10. What are your future/career goals?	Outcome Expectations	RQ3	Med-School, grad school, straight to job, unsure
11. Do you think undergraduate research experiences would help you achieve your future/career goals?	Perceived benefits	RQ3	Yes, just don't have the time No, not interested
12. Are there any changes you wish the department or university would make regarding recruitment into research experiences?	Outcome expectations	RQ3	Didn't know some opportunities are paid Wish the department would advertise differently
13. Is there anything else you would like to share with us?	Good to close with anything else you would like to share		

Appendix M

Qualitative Codebook

Sub-theme	Definition	Inclusion/Exclusion	Example
How you think			
General Interest in Science	Codes relating to participation in research due to an interest in science or research.		“I just. I really love science, and I always like was interested in research. I like in high school; I didn't really exactly know what it meant but there was always something I wanted to try.”
Disinterest	Codes relating to students expressing a lack of interest in research.		“… plus, research. I was never interested in research like that, so I don't think that I would be interested in doing something like that.”
Outcome Driven Interest	Codes relating to participation in research due to interest in learning skills, solving problems, and other affective outcomaaes.		“I think it like shows you're interested. It shows you're dedicated, and you want to do research and like work in a lab. And I think it just shows important skills. So absolutely that's definitely a bigger, you know why I want to do it. In the first place.”
Who you know			
Non-Structural Social Capital	Codes relating to social capital interactions outside of the institutional structure.	Includes codes that are not related to the institutional structure, excludes codes that are related to institutional structure.	“I had a friend. She's a computer science major, and she was talking about courses with research, and she told me a little bit about that.”

Qualitative codebook continued			
Sub-theme	Definition	Inclusion/Exclusion	Example
Structural Social Capital	Codes relating to social capital interactions relating to institutional structure; overall positive.	Includes codes that are related to the institutional structure, excludes all other social capital related codes.	"I was talking to my adviser, and I told… like my academic advisor…and I told her I was interested in doing research and grad school in the future. So, she encouraged me to look into research."
Positive communication about research opportunities	Codes related to communication with students about potential research opportunities, generally positive and resulting in students having an increased understanding of available opportunities.	Includes codes that mention positive communication. Excludes codes that specifically mention an individual or negative communication.	"I know some professors will offer, you know, 'I have a research lab', in class."
Negative communication about research opportunities	Codes related to communication with students about potential research opportunities, generally negative and resulting in a barrier to student URE participation.	Includes codes that mention negative communication. Excludes codes that specifically mention an individual or positive communication.	"I've heard just from people in research that it can take up a lot of your time going back and forth from lab to class."
Curricular and Programmatic Opportunities	Codes related to curricular and/or programmatic opportunities contributing to URE participation.	Includes codes related to curriculum or program connection to research excludes codes related to extracurricular activities.	"I'm in the Honors college at [university] so they give special research opportunities to our school students versus other students."

Qualitative codebook continued			
Sub-theme	Definition	Inclusion/Exclusion	Example
Facing known research barriers	Codes related to student navigation of research barriers.		"Um! A little bit of it was schedule, and then I also felt like sometimes niche interests can kind of hinder you, because sometimes the ones you're interested in either don't have any availability or um. They're similar to what you want to do. But they're not quite right, so it could be a little bit sure like tricky to figure out exactly what you're interested in."
What you do			
Individual barriers to doing science	Codes relating to barriers involving individual choice (e.g. social commitments)	Includes codes relating to barriers involving individual choice (e.g. social commitments), excludes codes relating to barriers outside of the student control.	"So mainly was I was involved in a lot of extracurricular activities so like organizations, and I was on boards. So, having to balance that with classes, and the research project was not beneficial for me."
Structural barriers to doing science	Codes relating to barriers involving institutional structure that are outside of the student control	Includes codes relating to institutional structure, excludes codes relating to barriers within student choice or institutional structure supports.	"Just being overwhelmed with, like my classes and classwork, like just not having enough time to do everything."

Qualitative codebook continued			
Sub-theme	Definition	Inclusion/Exclusion	Example
Structural supports to doing science	Codes relating to supports coming from institutional structure outside of the student control	Includes codes relating to institutional structural support, excludes codes relating to institutional structure barriers.	“And but what was really special with that like [professor] was so like lenient with me, he was so like forgiving and kind, and like was okay if I didn’t show up was or was okay. If I didn’t have my work done at certain times, because he knew, and that was kind of something that drove where I went.”
What you dream			
Feelings surrounding research participation	Codes relating to explicit or implicit mentions of feelings derived from research participation, negative or positive.		“…like mentally. I guess there can be some cost there, too. It's like if you're doing a lot of work in little time and being stressed, or just like having to be in an environment. That's not very. I don't know fun all the time.”
Research has changed how I view my future goals	Codes relating to participation in research changing participant future goals.		“Yes, definitely. I came in as a health Science major and I switch to biochemistry. And I switch to I wanna go to grad school instead of going to Med school because I loved doing research a lot more than I liked the idea of med school.”

Qualitative codebook continued			
Sub-theme	Definition	Inclusion/Exclusion	Example
Research is unrelated to career goals	Codes relating to participation in research not contributing or being unrelated to student's career goals.		"I'm in chemistry. And so that kind. The research that we go along with that is, it's probably fairly specific. I would want to be more applied."
Future research-specific aspirations	Codes relating to students' desire to participate in research in the future.		"Yeah, I definitely do think I will, I just. I don't know if i'll be like the kind of like head of the lab kind of thing like that kind of like further, or if it's like a researcher and some kind of company but definitely want to do it in the future."
Research as a steppingstone to the future	Codes relating to student aspiration to use research outcomes to promote their future goals.		"But also, I participated on Ph.D. Students research, my freshman year, and I was credited as an aid in his publication. So, I've been able to put that like on my Linkedin on job resumes, and it shows that I have a bit more of…of like an ability outside of just my degree, because it wasn't in my degree field. It was in chemical engineering, and I'm a geology student. So, it shows that I'm a little bit more extroverted than what paper would say."

Qualitative codebook continued			
Sub-theme	Definition	Inclusion/Exclusion	Example
Expectancy Outcomes			
Career related expectations	Codes relating to expectations of research leading to career related outcomes.		"Um…for like…for the biggest reason it was…you know, wanting to go into research in the future, and seeing if this is something that still interested me, and like how I would do doing it."
Action Expectations	Codes relating to expectation in research participation resulting in specific actions.		"Hopefully will reach some sort of publication by the end of this semester or this academic year."
Affective Expectations	Codes relating to expectations of outcomes relating to affective constructs.		"I felt like I wasn't contributing anything to like at my university, I guess, and that's not even it. It's just more of. I wanted to do some like meaningful work outside of coursework."

REFERENCES

Ackerman, N., Atherton, T., Avalani, A. R., Berven, C. A., Laskar, T., Neunzert, A., Parno, D. S., & Ramsey-Musolf, M. (2018). *LGBT+ Inclusivity in Physics and Astronomy: A Best Practices Guide* (arXiv:1804.08406). arXiv. https://doi.org/10.48550/arXiv.1804.08406

AERA. (2022). *AERA Joins Comments on NSF Survey of LGBTQI+ Indicators, Submits Comments on LGBTQI+ Evidence Agenda*. https://www.aera.net/Newsroom/AERA-Highlights-E-newsletter/AERA-Highlights-October-2022/AERA-Joins-Comments-on-NSF-Survey-of-LGBTQI-Indicators-Submits-Comments-on-LGBTQI-Evidence-Agenda?utm_source=Informz&utm_medium=email&utm_campaign=test

Alexander, J. M., Johnson, K. E., & Kelley, K. (2012). Longitudinal analysis of the relations between opportunities to learn about science and the development of interests related to science. *Science Education*, *96*(5), 763–786. https://doi.org/10.1002/sce.21018

American Chemical Society. (2015). *Undergraduate Professional Education in Chemistry: ACS Guidelines and Evaluation Procedures for Bachelor's Degree Programs*. American Chemical Society.

American Physics Society. (2008). *Statement on Undergraduate Research*. https://www.aps.org/programs/education/upload/undergrad_research_statement.pdf

Archer, L., Dawson, E., DeWitt, J., Seakins, A., & Wong, B. (2015). “Science capital”: A conceptual, methodological, and empirical argument for extending bourdieusian notions of capital beyond the arts. *Journal of Research in Science Teaching*, *52*(7), 922–948. https://doi.org/10.1002/tea.21227

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2012). Science Aspirations, Capital, and Family Habitus: How Families Shape Children’s Engagement and Identification With Science. *American Educational Research Journal*, *49*(5), 881–908. https://doi.org/10.3102/0002831211433290

Artino, A., La Rochelle, J., Dezee, K., & Gehlbech, H. (2014). Developing questionnaires for educational research: AMEE Guide No. 87. *Medical Teacher*, *36*(6). https://www.tandfonline.com/doi/epdf/10.3109/0142159X.2014.889814?needAccess=true&role=button

Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., Lawrie, G., McLinn, C. M., Pelaez, N., Rowland, S., Towns, M., Trautmann, N. M., Varma-Nelson, P., Weston, T. J., & Dolan, E. L. (2014). Assessment of Course-Based Undergraduate Research Experiences: A Meeting Report. *CBE—Life Sciences Education*, *13*(1), 29–40. https://doi.org/10.1187/cbe.14-01-0004

Bak, A. (Director). (2014). *Using Topological Data Analysis on your Big Data*. https://www.youtube.com/watch?v=gtFVdGb9Y8w

Baker, V. L., Pifer, M. J., Lunsford, L. G., Greer, J., & Ihas, D. (2015). Faculty as Mentors in Undergraduate Research, Scholarship, and Creative Work: Motivating and Inhibiting Factors. *Mentoring & Tutoring: Partnership in Learning*, *23*(5), 394–410. https://doi.org/10.1080/13611267.2015.1126164

Ballen, C. J., Thompson, S. K., Blum, J. E., Newstrom, N. P., & Cotner, S. (2018). Discovery and Broad Relevance May Be Insignificant Components of Course-Based Undergraduate Research Experiences (CUREs) for Non-Biology Majors †. *Journal of Microbiology & Biology Education*, *19*(2). https://doi.org/10.1128/jmbe.v19i2.1515

Bangera, G., & Brownell, S. E. (2014). Course-Based Undergraduate Research Experiences Can Make Scientific Research More Inclusive. *CBE—Life Sciences Education*, *13*(4), 602–606. https://doi.org/10.1187/cbe.14-06-0099

Barber, P. H., Shapiro, C., Jacobs, M. S., Avilez, L., Brenner, K. I., Cabral, C., Cebreros, M., Cosentino, E., Cross, C., Gonzalez, M. L., Lumada, K. T., Menjivar, A. T., Narvaez, J., Olmeda, B., Phelan, R., Purdy, D., Salam, S., Serrano, L., Velasco, M. J., … Levis-Fitzgerald, M. (2021). Disparities in Remote Learning Faced by First-Generation and Underrepresented Minority Students during COVID-19: Insights and Opportunities from a Remote Research Experience. *Journal of Microbiology & Biology Education*, *22*(1), ev22i1.2457. https://doi.org/10.1128/jmbe.v22i1.2457

Batty, L., & Reilly, K. (2022). Understanding barriers to participation within undergraduate STEM laboratories: Towards development of an inclusive curriculum. *Journal of Biological Education*, *0*(0), 1–23. https://doi.org/10.1080/00219266.2021.2012227

Biden, J. (2022, June 15). *Executive Order on Advancing Equality for Lesbian, Gay, Bisexual, Transgender, Queer, and Intersex Individuals*. The White House. https://www.whitehouse.gov/briefing-room/presidential-actions/2022/06/15/executive-order-on-advancing-equality-for-lesbian-gay-bisexual-transgender-queer-and-intersex-individuals/

Bingham, B. E. (2021). *Exploring the Impact and Accessibility of Undergraduate Research Experiences in the Natural Sciences* [Master of Science]. University of Guelph.

Bloodhart, B., Balgopal, M. M., Casper, A. M. A., McMeeking, L. B. S., & Fischer, E. V. (2020). Outperforming yet undervalued: Undergraduate women in STEM. *PLOS ONE*, *15*(6), e0234685. https://doi.org/10.1371/journal.pone.0234685

Boriah, S., Chandola, V., & Kumar, V. (2008). Similarity Measures for Categorical Data: A Comparative Evaluation. *Society for Industrial and Applied Mathematics - 8th SIAM International Conference on Data Mining 2008, Proceedings in Applied Mathematics 130*. https://epubs.siam.org/doi/epdf/10.1137/1.9781611972788.22

Bourdieu, P. (1986). The Forms of Capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education*. Greenwood Press.

Boyd, E. A., Lazar, K. B., & Moysey, S. M. (2023). Big data to support geoscience recruitment: Novel adoption of topological data analysis in geoscience education. *GSA Bulletin*.

Boyd, E. A., Lazar, K. B., & Voigt, M. (2023). Humanizing geoscience education research through participant-driven visual representation. *Journal of Geoscience Education*, *0*(0), 1–9. https://doi.org/10.1080/10899995.2023.2193551

Boyd, E., & Lazar, K. B. (2023). *Factors influencing science transfer student participation in undergraduate research*. National Institute for the Study of Transfer Students, Virtual.

Boyer Commission on Educating Undergraduates in the Research University. (1998). *Reinventing Undergraduate Education: A Blueprint for America's Research Universities.* Carnegie Foundation for the Advancement of Teaching.

BrckaLorenz, A., Garvey, J. C., Hurtado, S. S., & Latopolski, K. (2017). High-impact practices and student–faculty interactions for gender-variant students. *Journal of Diversity in Higher Education*, *10*, 350–365. https://doi.org/10.1037/dhe0000065

Busch, C. (2022). Disclosure Decisions: Science and Engineering Instructors as Role Models with Concealable Stigmatized Identities. *ESERA Summer School 2022 EBook of Synopses*. European Science Education Research Association, Utrecht, The Netherlands.

Cacciatore, G. (2021). *Exploring Equity: Citizenship and Nationality*. Harvard Graduate School of Education. https://www.gse.harvard.edu/news/uk/21/02/exploring-equity-citizenship-and-nationality

Cameron, M., Lacy, T. A., Siegel, P., Wu, J., Wilson, A., Johnson, R., Burns, R., & Wine, J. (2021). *First Look at the Impact of the Coronavirus (COVID-19) Pandemic on Undergraduate Student Enrollment, Housing, and Finances*

(Preliminary Data) (2019–20 National Postsecondary Student Aid Study (NPSAS:20), p. 54). U.S. Department of Education. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2021456.

Campus-Wide Award for Undergraduate Research Accomplishments (AURA) | CUR - AURA | Council on Undergraduate Research. (n.d.). Retrieved October 5, 2021, from https://www.cur.org/what/recognition/aura/

Canaria, J. A., Schoffstall, A. M., Weiss, D. J., Henry, R. M., & Braun-Sand, S. B. (2012). A Model for an Introductory Undergraduate Research Experience. *Journal of Chemical Education*, *89*(11), 1371–1377. https://doi.org/10.1021/ed200582v

Carabajal, I. G., Marshall, A. M., & Atchison, C. L. (2017). A synthesis of instructional strategies in geoscience education literature that address barriers to inclusion for students with disabilities. *Journal of Geoscience Education*, *65*(4), 531–541. https://doi.org/10.5408/16-211.1

Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*, *46*(2), 255–308. https://doi.org/10.1090/S0273-0979-09-01249-X

Carnegie Classifications | Definitions. (n.d.). Retrieved September 4, 2020, from https://carnegieclassifications.iu.edu/definitions.php

Carpi, A., Ronan, D. M., Falconer, H. M., & Lents, N. H. (2017). Cultivating minority scientists: Undergraduate research increases self-efficacy and career ambitions for underrepresented students in STEM. *Journal of Research in Science Teaching*, *54*(2), 169–194. https://doi.org/10.1002/tea.21341

Castillo, Y., & Estudillo, A. (2015). Undergraduate Research: An Essential Piece for Underrepresented Students' College Success. *Faculty Publications*. https://scholarworks.sfasu.edu/humanservices_facultypubs/4

Ceyhan, G. D., & Tillotson, J. W. (2020). Early year undergraduate researchers' reflections on the values and perceived costs of their research experience. *International Journal of STEM Education*, *7*(1), 54. https://doi.org/10.1186/s40594-020-00248-x

Chazal, F., & Michel, B. (2016). *Mapper Algorithm with the R-package TDAmapper*. http://bertrand.michel.perso.math.cnrs.fr/Enseignements/TDA/Mapper.html#plotting-the-mapper-graph

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. https://doi.org/10.1111/j.1540-4560.2011.01710.x

Chrousos, G. P., & Mentis, A.-F. A. (2020). Imposter syndrome threatens diversity. *Science*, *367*(6479), 749–750. https://doi.org/10.1126/science.aba8039

CIP user site. (n.d.). Retrieved October 5, 2021, from https://nces.ed.gov/ipeds/cipcode/browse.aspx?y=56

Clancy, K. B. H., Nelson, R. G., Rutherford, J. N., & Hinde, K. (2014). Survey of Academic Field Experiences (SAFE): Trainees Report Harassment and Assault. *PLOS ONE*, *9*(7), e102172. https://doi.org/10.1371/journal.pone.0102172

Collins, P. H. (2019). *Intersectionality as Critical Social Theory*. Duke University Press.

Cooper, K., Brownell, S., & Gormally, C. (2019). Coming Out to the Class: Identifying Factors that Influence College Biology Instructor Decisions About Revealing Their LGBQ Identities in Class. *Journal of Women and Minorities in Science and Engineering*, *25*. https://doi.org/10.1615/JWomenMinorScienEng.2019026085

Cooper, K., Cala, J. M., & Brownell, S. E. (2021). Cultural capital in undergraduate research: An exploration of how biology students operationalize knowledge to access research experiences at a large, public research-intensive institution. *International Journal of STEM Education*, *8*(1), 6. https://doi.org/10.1186/s40594-020-00265-w

Cooper, K., Gin, L. E., Akeeh, B., Clark, C. E., Hunter, J. S., Roderick, T. B., Elliott, D. B., Gutierrez, L. A., Mello, R. M., Pfeiffer, L. D., Scott, R. A., Arellano, D., Ramirez, D., Valdez, E. M., Vargas, C., Velarde, K., Zheng, Y., & Brownell, S. E. (2019). Factors that predict life sciences student persistence in undergraduate research experiences. *PLOS ONE*, *14*(8), e0220186. https://doi.org/10.1371/journal.pone.0220186

Cooper, K., Gin, L. E., Barnes, M. E., & Brownell, S. E. (2020). An Exploratory Study of Students with Depression in Undergraduate Research Experiences. *CBE—Life Sciences Education*, *19*(2), ar19. https://doi.org/10.1187/cbe.19-11-0217

Cooper, K., Gin, L. E., & Brownell, S. E. (2020). Depression as a concealable stigmatized identity: What influences whether students conceal or reveal their depression in undergraduate research experiences? *International Journal of STEM Education*, *7*(1), 27. https://doi.org/10.1186/s40594-020-00216-5

Creamer, E. (2019). Evaluating Quality in Mixed Methods Research Publications. In *Evaluating Quality*.

Crenshaw, K. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum*, *1989*(1).

Creswell, J. W., & Plano-Clark, V. L. (2007). Choosing a Mixed Methods Design. In *Designing and Conducting Mixed Methods Research* (pp. 62–79). SAGE Publications.

Demery, A.-J. C., & Pipkin, M. A. (2021). Safe fieldwork strategies for at-risk individuals, their supervisors and institutions. *Nature Ecology & Evolution*, *5*(1), 5–9. https://doi.org/10.1038/s41559-020-01328-5

DeWitt, J., Archer, L., & Mau, A. (2016). Dimensions of science capital: Exploring its potential for understanding students' science participation. *International Journal of Science Education*, *38*(16), 2431–2449. https://doi.org/10.1080/09500693.2016.1248520

Diamond, L. M., & Alley, J. (2022). Rethinking minority stress: A social safety perspective on the health effects of stigma in sexually-diverse and gender-diverse populations. *Neuroscience & Biobehavioral Reviews*, *138*, 104720. https://doi.org/10.1016/j.neubiorev.2022.104720

Doyle, J. (2017). *Describing and Mapping the Interactions between Student Affective Factors Related to Persistence in Science, Physics, and Engineering* [Ph.D., Florida International University]. https://doi.org/10.25148/etd.FIDC001978

Doyle, J., & Potvin, G. (2015). In search of distinct graduate admission strategies in physics: An exploratory study using topological data analysis. *The Physics Education Research Conference Proceedings*. http://scholar.google.com/citations?view_op=view_citation&hl=en&user=20EF5t4AAAAJ&citation_for_view=20EF5t4AAAAJ:Y0pCki6q_DkC

Dugan, J. (2021). Beware of Equity Traps and Tropes. *ASCD*, *78*(6). https://www.ascd.org/el/articles/beware-of-equity-traps-and-tropes

Eagan, M. K., Hurtado, S., Chang, M. J., Garcia, G. A., Herrera, F. A., & Garibay, J. C. (2013). Making a Difference in Science Education: The Impact of Undergraduate Research Programs. *American Educational Research Journal*, *50*(4), 683–713. https://doi.org/10.3102/0002831213482038

Erickson, O. A., Cole, R. B., Isaacs, J. M., Alvarez-Clare, S., Arnold, J., Augustus-Wallace, A., Ayoob, J. C., Berkowitz, A., Branchaw, J., Burgio, K. R., Cannon, C. H., Ceballos, R. M., Cohen, C. S., Coller, H., Disney, J., Doze, V. A., Eggers, M. J., Farina, S., Ferguson, E. L., … Dolan, E. L. (2022). "How Do We Do This at a Distance?!" A Descriptive Study of Remote Undergraduate Research Programs during COVID-19. *CBE—Life Sciences Education*, *21*(1), ar1. https://doi.org/10.1187/cbe.21-05-0125

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications.

Fogg-Rogers, L., & Hobbs, L. (2019). Catch 22—Improving visibility of women in science and engineering for both recruitment and retention. *Journal of Science Communication*, *18*(4), C05. https://doi.org/10.22323/2.18040305

Formann, A. K., Ehlers, T., & Scheiblechner, H. (1980). Die Anwendung der “Latent-Class-Analyse” auf Probleme der diagnostischen Klassifikation am Beispiel der Marburger Verhaltensliste. [The application of “latent class analysis” to problems of the diagnostic classification with the Marburg Behavior Inventory as an example.]. *Zeitschrift Für Differentielle Und Diagnostische Psychologie*, *1*(4), 319–330.

Frantz, K. J., Demetrikopoulos, M. K., Britner, S. L., Carruth, L. L., Williams, B. A., Pecore, J. L., DeHaan, R. L., & Goode, C. T. (2017). A Comparison of Internal Dispositions and Career Trajectories after Collaborative versus Apprenticed Research Experiences for Undergraduates. *CBE—Life Sciences Education*, *16*(1), ar1. https://doi.org/10.1187/cbe.16-06-0206

Freeman, J. B. (2018). LGBTQ scientists are still left out. *Nature*, *559*(7712), 27–28. https://doi.org/10.1038/d41586-018-05587-y

Freeman, J. B. (2020). Measuring and Resolving LGBTQ Disparities in STEM. *Policy Insights from the Behavioral and Brain Sciences*, *7*(2), 141–148. https://doi.org/10.1177/2372732220943232

Gentile, J., Brenner, K., & Stephens, A. (Eds.). (2017). *Undergraduate Research Experiences for STEM Students: Successes, Challenges, and Opportunities*. National Academies Press. https://doi.org/10.17226/24622

Giles, S., Jackson, C., & Stephen, N. (2020). Barriers to fieldwork in undergraduate geoscience degrees. *Nature Reviews Earth & Environment*, *1*(2), Article 2. https://doi.org/10.1038/s43017-020-0022-5

Gin, L. E., Pais, D., Cooper, K. M., & Brownell, S. E. (2022). Students with Disabilities in Life Science Undergraduate Research Experiences: Challenges and Opportunities. *CBE—Life Sciences Education*, *21*(2), ar32. https://doi.org/10.1187/cbe.21-07-0196

Godwin, A., Thielmeyer, A. R. H., Verdín, D., Benedict, B. S., Baker, R. A., & Doyle, J. (2019). Using Topological Data Analysis in Social Science Research: Unpacking Decisions and Opportunities for a New Method. *ASEE*.

Gonzales, R. G. (2016). *Lives in Limbo: Undocumented and Coming of Age in America* (1st ed.). University of California Press. http://www.jstor.org/stable/10.1525/j.ctv1xxt36

Greathouse, M., BrckaLorenz, A., Hoban, M., Ronald Huesman Jr., Rankin, S., & Stolzenberg, E. B. (2018). *Queer-Spectrum and Trans-Spectrum Student Experiences in American Higher Education: The Analyses of National Survey Findings*. Rutgers University. https://tcc-j2made.s3.amazonaws.com/uploads/2018/09/White-Paper-Final.pdf

Grineski, S. E., Morales, D. X., Collins, T. W., Nadybal, S., & Trego, S. (2022). A US National Study of Barriers to Science Training Experienced by Undergraduate Students during COVID-19. *International Journal of Environmental Research and Public Health*, *19*(11), Article 11. https://doi.org/10.3390/ijerph19116534

Habig, B., Gupta, P., & Adams, J. D. (2021). Disrupting deficit narratives in informal science education: Applying community cultural wealth theory to youth learning and engagement. *Cultural Studies of Science Education, 16*(2), 509–548. https://doi.org/10.1007/s11422-020-10014-8

Haeger, H., White, C., Martinez, S., & Velasquez, S. (2021). Creating More Inclusive Research Environments for Undergraduates. *Journal of the Scholarship of Teaching and Learning*, *21*(1), Article 1. https://doi.org/10.14434/josotl.v21i1.30101

Hall, L. M., & Belch, H. A. (2000). Setting the Context: Reconsidering the Principles of Full Participation and Meaningful Access for Students with Disabilities. In *New Directions for Student Services: Serving Students with Disabilities*. John Wiley & Sons.

Hamrick, K. (2022). *Women, Minorities, and Persons with Disabilities in Science and Engineering*. National Science Foundation. https://ncses.nsf.gov/pubs/nsf21321/data-tables

Harackiewicz, J. M., Durik, A. M., Barron, K. E., Linnenbrink-Garcia, L., & Tauer, J. M. (2008). The role of achievement goals in the development of interest: Reciprocal relations between achievement goals, interest, and performance. *Journal of Educational Psychology*, *100*, 105–122. https://doi.org/10.1037/0022-0663.100.1.105

Harsh, J. A., Maltese, A. V., & Tai, R. H. (2011). Undergraduate Research Experiences From a Longitudinal Perspective. *Journal of College Science Teaching; Washington*, *41*(1), 84–91.

Hewlett, J. A. (2018). Broadening Participation in Undergraduate Research Experiences (UREs): The Expanding Role of the Community College. *CBE—Life Sciences Education*, *17*(3), es9. https://doi.org/10.1187/cbe.17-11-0238

Hughes, B. E. (2018). Coming out in STEM: Factors affecting retention of sexual minority STEM students. *Science Advances*, *4*(3), eaao6373. https://doi.org/10.1126/sciadv.aao6373

Iacobucci, D., Popovich, D. L., Moon, S., & Román, S. (2022). How to calculate, use, and report variance explained effect size indices and not die trying. *Journal of Consumer Psychology*, *n/a*(n/a). https://doi.org/10.1002/jcpy.1292

IBM. (2021). *SPSS Statistics*. https://www.ibm.com/products/spss-statistics?p1=Search&p4=p50715561164&p5=e&cm_mmc=Search_Google-_-1S_1S-_-WW_NA-_-spss_e&cm_mmca7=71700000060951760&cm_mmca8=kwd-98310549&cm_mmca9=EAIaIQobChMIoOOa5PfZ6AIVAe21Ch0zcgmZEAAYASAAEgJGtPD_BwE&cm_mmca10=407020396680&cm_mmca11=e&gclid=EAIaIQobChMIoOOa5PfZ6AIVAe21Ch0zcgmZEAAYASAAEgJGtPD_BwE&gclsrc=aw.ds

Jones, M. G., Ennes, M., Weedfall, D., Chesnutt, K., & Cayton, E. (2020). The Development and Validation of a Measure of Science Capital, Habitus, and

Future Science Interests. *Research in Science Education*. https://doi.org/10.1007/s11165-020-09916-y

Kallschmidt, A. M., & Eaton, A. A. (2019). Are lower social class origins stigmatized at work? A qualitative study of social class concealment and disclosure among White men employees who experienced upward mobility. *Journal of Vocational Behavior*, *113*, 115–128. https://doi.org/10.1016/j.jvb.2018.08.010

Katkin, W. (2003). The Boyer Commission Report and its Impact on Undergraduate Research. *New Directions for Teaching and Learning*, *2003*(93), 19–38. https://doi.org/10.1002/tl.86

Khan, M. M. R., Arif, R. B., Siddique, Md. A. B., & Oishe, M. R. (2018). Study and Observation of the Variation of Accuracies of KNN, SVM, LMNN, ENN Algorithms on Eleven Different Datasets from UCI Machine Learning Repository. *2018 4th International Conference on Electrical Engineering and Information & Communication Technology (ICEEiCT)*, 124–129. https://doi.org/10.1109/CEEICT.2018.8628041

Kirkpatrick, C., Schuchardt, A., Baltz, D., & Cotner, S. (2019). Computer-Based and Bench-Based Undergraduate Research Experiences Produce Similar Attitudinal Outcomes. *CBE—Life Sciences Education*, *18*(1), ar10. https://doi.org/10.1187/cbe.18-07-0112

Krim, J. S., Coté, L. E., Schwartz, R. S., Stone, E. M., Cleeves, J. J., Barry, K. J., Burgess, W., Buxner, S. R., Gerton, J. M., Horvath, L., Keller, J. M., Lee, S. C., Locke, S. M., & Rebar, B. M. (2019). Models and Impacts of Science Research

Experiences: A Review of the Literature of CUREs, UREs, and TREs. *CBE—Life Sciences Education*, *18*(4), ar65. https://doi.org/10.1187/cbe.19-03-0069

Kuchynka, S. L., Salomon, K., Bosson, J. K., El-Hout, M., Kiebel, E., Cooperman, C., & Toomey, R. (2018). Hostile and Benevolent Sexism and College Women's STEM Outcomes. *Psychology of Women Quarterly*, *42*(1), 72–87. https://doi.org/10.1177/0361684317741889

Kuh, G. D. (2008). *High-Impact Educational Practices: What They Are, Who Has Access to Them, and Why They Matter*. Association of American College & Universities.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*(1), 159–174.

Lent, R., Brown, S., & Hackett, G. (1994). Toward a Unifying Social Cognitive Theory of Career and Academic Interest, Choice, and Performance. *Journal of Vocational Behavior*, *45*, 79–122.

Lewis, K. L., Stout, J. G., Pollock, S. J., Finkelstein, N. D., & Ito, T. A. (2016). Fitting in or opting out: A review of key social-psychological factors influencing a sense of belonging for women in physics. *Physical Review Physics Education Research*, *12*(2), 020110. https://doi.org/10.1103/PhysRevPhysEducRes.12.020110

Lillywhite, A., & Wolbring, G. (2019). Undergraduate Disabled Students as Knowledge Producers including Researchers: A Missed Topic in Academic Literature. *Education Sciences*, *9*(4), Article 4. https://doi.org/10.3390/educsci9040259

Limeri, L. B., Asif, M. Z., & Dolan, E. L. (2019). Volunteered or Voluntold? The Motivations and Perceived Outcomes of Graduate and Postdoctoral Mentors of

Undergraduate Researchers. *CBE—Life Sciences Education*, *18*(2), ar13. https://doi.org/10.1187/cbe.18-10-0219

Lopatto, D. (2009). *Science in Solution: The Impact of Undergraduate Research on Student Learning* [Computer file]. Research Corporation for Science Advancement.

Lopatto, D., Hauser, C., Jones, C. J., Paetkau, D., Chandrasekaran, V., Dunbar, D., MacKinnon, C., Stamm, J., Alvarez, C., Barnard, D., Bedard, J. E. J., Bednarski, A. E., Bhalla, S., Braverman, J. M., Burg, M., Chung, H.-M., DeJong, R. J., DiAngelo, J. R., Du, C., … Elgin, S. C. R. (2014). A Central Support System Can Facilitate Implementation and Sustainability of a Classroom-Based Undergraduate Research Experience (CURE) in Genomics. *CBE—Life Sciences Education*, *13*(4), 711–723. https://doi.org/10.1187/cbe.13-10-0200

Marín-Spiotta, E., Barnes, R. T., Berhe, A. A., Hastings, M. G., Mattheis, A., Schneider, B., & Williams, B. M. (2020). Hostile climates are barriers to diversifying the geosciences. *Advances in Geosciences*, *53*, 117–127. https://doi.org/10.5194/adgeo-53-117-2020

McNair, T. B., & Veras, J. (2017). Committing to equity and inclusive excellence: Intentionality and accountability. *Peer Review*, *19*(2).

McPherson, E. (2014). Informal learning in SME majors for African American female undergraduates. *Global Education Review*, *1*(4), Article 4. https://ger.mercy.edu/index.php/ger/article/view/30

Moote, J., Archer, L., DeWitt, J., & MacLeod, E. (2020). Science capital or STEM capital? Exploring relationships between science capital and technology, engineering, and maths aspirations and attitudes among young people aged 17/18. *Journal of Research in Science Teaching*, *57*(8), 1228–1249. https://doi.org/10.1002/tea.21628

Morales, D. X., Grineski, S. E., & Collins, T. W. (2017). Faculty Motivation to Mentor Students Through Undergraduate Research Programs: A Study of Enabling and Constraining Factors. *Research in Higher Education*, *58*(5), 520–544. https://doi.org/10.1007/s11162-016-9435-x

Munch, E. (2017). A User's Guide to Topological Data Analysis. *Journal of Learning Analytics*, *4*(2). https://doi.org/10.18608/jla.2017.42.6

Murugan, J., & Robertson, D. (2019). *An Introduction to Topological Data Analysis for Physicists: From LGM to FRBs* (arXiv:1904.11044). arXiv. http://arxiv.org/abs/1904.11044

Myers, K. R., Tham, W. Y., Yin, Y., Cohodes, N., Thursby, J. G., Thursby, M. C., Schiffer, P., Walsh, J. T., Lakhani, K. R., & Wang, D. (2020). Unequal effects of the COVID-19 pandemic on scientists. *Nature Human Behaviour*, 1–4. https://doi.org/10.1038/s41562-020-0921-y

National Center for Education Statistics. (2022a). *Integrated Postsecondary Education Data System*. https://nces.ed.gov/ipeds/datacenter/InstitutionByName.aspx

National Center for Education Statistics. (2022b). *Use of Supports Among Students With Disabilities and Special Needs in College*.

NSF's 10 Big Ideas—Special Report | NSF - National Science Foundation. (n.d.). Retrieved April 11, 2023, from https://www.nsf.gov/news/special_reports/big_ideas/

O'Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods*, *19*, 1609406919899220. https://doi.org/10.1177/1609406919899220

Onwuegbuzie, A. J., Johnson, R. B., & Collins, K. M. T. (2011). Assessing legitimation in mixed research: A new framework. *Quality & Quantity*, *45*(6), 1253–1271. https://doi.org/10.1007/s11135-009-9289-9

Pandey, A., & Jain, A. (2017). Comparative Analysis of KNN Algorithm using Various Normalization Techniques. *International Journal of Computer Network and Information Security*, *9*(11), 36–42. https://doi.org/10.5815/ijcnis.2017.11.04

Pierszalowski, S., Bouwma-Gearhart, J., & Marlow, L. (2021). A Systematic Review of Barriers to Accessing Undergraduate Research for STEM Students: Problematizing Under-Researched Factors for Students of Color. *Social Sciences*, *10*(9), Article 9. https://doi.org/10.3390/socsci10090328

Quinn, D. M., Williams, M. K., Quintana, F., Gaskins, J. L., Overstreet, N. M., Pishori, A., Earnshaw, V. A., Perez, G., & Chaudoir, S. R. (2014). Examining Effects of Anticipated Stigma, Centrality, Salience, Internalization, and Outness on Psychological Distress for People with Concealable Stigmatized Identities. *PLOS ONE*, *9*(5), e96977. https://doi.org/10.1371/journal.pone.0096977

R Core Team. (2017). *R*. R Foundation for Statistical Computing. https://www.R-project.org/.

Ramirez, E. (2011). “No One Taught Me the Steps”: Latinos’ Experiences Applying to Graduate School. *Journal of Latinos and Education*, *10*(3), 204–222. https://doi.org/10.1080/15348431.2011.581105

Ramirez, J. L., & Paz Galupo, M. (2019). Multiple minority stress: The role of proximal and distal stress on mental health outcomes among lesbian, gay, and bisexual people of color. *Journal of Gay & Lesbian Mental Health*, *23*(2), 145–167. https://doi.org/10.1080/19359705.2019.1568946

Ries, K., & Gray, S. D. (2018). Fostering Undergraduate Research with a Nontraditional Student Population. *Journal of Chemical Education*, *95*(9), 1443–1447. https://doi.org/10.1021/acs.jchemed.8b00284

Robinson, J. L. (2018). *Visible and concealed marginalized identity and microaggressions: Impact on well-being* [Ph.D., Texas Women’s University]. https://twu-ir.tdl.org/handle/11274/10734

Rubin, D. B., & Little, R. J. A. (2019). *Statistical Analysis with Missing Data*. John Wiley & Sons.

Russell, S. H., Hancock, M. P., & McCullough, J. (2007). Benefits of Undergraduate Research Experiences. *Science*, *316*(5824), 548–549.

Saldaña, J. (2016). *The Coding Manual for Qualitative Researchers* (3rd ed.). SAGE.

Sandquist, E., Cervato, C., & Ogilvie, C. (2019). Positive Affective and Behavioral Gains of First-Year Students in Course-Based Research across Disciplines. *Scholarship* *and Practice of Undergraduate Research*, *2*(4), 45–57. https://doi.org/10.18833/spur/2/4/9

Santana, L. M., & Singh, C. (2021). Negative impacts of an unwelcoming physics environment on undergraduate women. *2021 Physics Education Research Conference Proceedings*, 377–383. https://doi.org/doi.org/10.1119/perc.2021.pr.Santana

Santuzzi, A. M., Waltz, P. R., Finkelstein, L. M., & Rupp, D. E. (2014). Invisible Disabilities: Unique Challenges for Employees and Organizations. *Industrial and Organizational Psychology*, *7*(2), 204–219. https://doi.org/10.1111/iops.12134

Scheitle, C. P., & Dabbs, E. (2021). Religiosity and identity interference among graduate students in the sciences. *Social Science Research*, *93*, 102503. https://doi.org/10.1016/j.ssresearch.2020.102503

Serrano, D. H., Hernández-Serrano, J., & Gómez, D. S. (2020). Simplicial degree in complex networks. Applications of topological data analysis to network science. *Chaos, Solitons & Fractals*, *137*, 109839. https://doi.org/10.1016/j.chaos.2020.109839

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. https://doi.org/10.1002/sce.10131

Singh, G., Memoli, F., & Carlsson, G. (2007). *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*. The Eurographics Association. https://doi.org/10.2312/SPBG/SPBG07/091-100

St. John, K., Riggs, E., & Mogk, D. (2016). Sexual Harassment in the Sciences: A Call to Geoscience Faculty and Researchers to Respond. *Journal of Geoscience Education*, *64*(4), 255–257. https://doi.org/10.5408/1089-9995-64.4.255

Stevens, I. E. (2016). *Topological Data Analysis and the MAA National Study of College Calculus* [M.A.]. University of Georgia.

Szteinberg, G. A. (2012). *Long-term effects of course-embedded undergraduate research: The CASPiE longitudinal study* [Ph.D., Purdue University]. http://search.proquest.com/docview/1328162252/abstract/CECD820277F44843PQ/1

The National Academies of Science Engineering and Medicine (NASEM). (2011). *Expanding Underrepresented Minority Participation: America's Science and Technology Talent at the Crossroads*. National Academies Press. https://doi.org/10.17226/12984

The National Academies of Science Engineering and Medicine (NASEM). (2019). *The Science of Effective Mentorship in STEMM*. The National Academies Press. https://doi.org/10.17226/25568

Thompson, J. J., & Jensen-Ryan, D. (2018). Becoming a "Science Person": Faculty Recognition and the Development of Cultural Capital in the Context of

Undergraduate Biology Research. *CBE—Life Sciences Education*, *17*(4), ar62. https://doi.org/10.1187/cbe.17-11-0229

Traxler, A. L. (2020). Who Belongs in Science: Numbers Are Not Enough. In *Addressing Gender Bias in Science & Technology* (Vol. 1354, pp. 73–90). American Chemical Society. https://doi.org/10.1021/bk-2020-1354.ch005

Trogden, B. G., Kennedy, C., & Biyani, N. K. (2022). Mapping and Making Meaning from Undergraduate Student Engagement in High-Impact Educational Practices. *Innovative Higher Education*. https://doi.org/10.1007/s10755-022-09608-7

University of Utah Center for Natural and Technological Hazards. (2020). *Initial Impacts of COVID-19 on Undergraduate Research*. University of Utah.

U.S. Department of Education. (n.d.). *National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), [2018], [Enrollment]*. Retrieved July 26, 2021, from https://nces.ed.gov/ipeds/use-the-data

Vincent-Ruz, P., Grabowski, J., & Schunn, C. D. (2018). The Impact of Early Participation in Undergraduate Research Experiences on Multiple Measures of Premed Path Success. *Council on Undergraduate Research Quarterly*, *1*(3), 13–18. https://doi.org/10.18833/spur/1/3/12

Walther, J., Sochacka, N. W., & Kellam, N. N. (2013). Quality in Interpretive Engineering Education Research: Reflections on an Example Study. *Journal of Engineering Education*, *102*(4), 626–659. https://doi.org/10.1002/jee.20029

Wasserman, L. (2018). Topological Data Analysis. *Annual Review of Statistics and Its Application*, *5*.

Whitehead, M. A. (2019). “Where are My People At?”: A Community Cultural Wealth Analysis of How Lesbian, Gay, and Bisexual Community College Students of Color Access Community and Support. *Community College Journal of Research and Practice*, *43*(10–11), 730–742. https://doi.org/10.1080/10668926.2019.1600611

Yoder, J. B., & Mattheis, A. (2016). Queer in STEM: Workplace Experiences Reported in a National Survey of LGBTQA Individuals in Science, Technology, Engineering, and Mathematics Careers. *Journal of Homosexuality*, *63*(1), 1–27. https://doi.org/10.1080/00918369.2015.1078632

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

Zilberman, A., & Ice, L. (2021). Why computer occupations are behind strong STEM employment growth in the 2019–29 decade: Beyond the Numbers: U.S. Bureau of Labor Statistics. *U.S. Bureau of Labor Statistics: Beyond the Numbers*, *10*(1). https://www.bls.gov/opub/btn/volume-10/why-computer-occupations-are-behind-strong-stem-employment-growth.htm