RICE UNIVERSITY

Qualitative Data Analysis and Biodata Measure Development of Rice Undergraduates' STEM Formative Experiences

by

BRITTANY BRADFORD

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Master of Arts

APPROVED, THESIS COMMITTEE

Margaret E. Beier, Chair
Associate Professor, Psychological Sciences

Eduardo Salas
Professor and Allyn R. & Gladys M. Cline Chair, Department Chair, Psychological Sciences

Michelle R. Hebl
Martha and Henry Malcolm Lovett Professor, Psychological Sciences and Management

HOUSTON, TEXAS

December 2018
ABSTRACT

Qualitative Data Analysis and Biodata Measure Development of Rice Undergraduates' STEM Formative Experiences

by

Brittany Bradford

Underrepresented minorities (URMs), females, and first-generation college students declaring a science, technology, engineering, or mathematics (STEM) major face unique challenges as they transition into college; moreover, they ultimately fail to graduate with STEM majors at rates commensurate with other students. Research has shown that children's formative experiences may impact their STEM self-efficacy, interests, and identity for the rest of their academic and professional careers. This study used a thematic qualitative data analysis of 35 semi-structured interviews with Rice STEM students regarding their formative experiences. Six theoretical dimensions were developed: math perceptions, science perceptions, classroom experiences, STEM identity, exposure to STEM, and parental expectations. Participants reported low exposure to science in elementary school, strong parental influence, preferences for hands-on and problem solving tasks, and teacher support. Next steps for this research involve developing and validating a biodata measure based on these dimensions and determining its predictive validity for STEM identity and self-efficacy.
ACKNOWLEDGEMENTS

I am grateful to my advisor, Dr. Margaret Beier, as well as my committee members, Dr. Eduardo Salas and Dr. Mikki Hebl, for their feedback and guidance on this and other research projects. My cohort has been a source of constant encouragement and inspiration; thank you to Abby Corrington, Julie Dinh, Denise Reyes, and Amanda Woods. My lab members keep me sane and provide endless support (thank you, Meghan Davenport, Jackie Gilberto, Michelle Kim, and Jackie Torres). I am also grateful for all the other amazing people I have met in the Psychological Sciences department. Finally, to my parents, my best friends, and my sister: thank you for supporting me every step of the way on my journey to graduate school.
Contents

Acknowledgements ........................................................................................................ iii

Contents .......................................................................................................................... iv

List of Figures .................................................................................................................. iv

List of Tables .................................................................................................................... vi

Chapter 1: Introduction ................................................................................................. 1

  Academic Underperformance of URM Students ......................................................... 3
  STEM Retention ............................................................................................................ 5
  Formative STEM Experiences ....................................................................................... 8
  Development of STEM Identity ..................................................................................... 14
  Other STEM Outcomes ................................................................................................. 17
  Biodata ........................................................................................................................ 20
  Biodata Measures .......................................................................................................... 22

Chapter 2: The Present Study ....................................................................................... 24

  Overview of Study ........................................................................................................ 24
  Qualitative Interviews .................................................................................................. 24
  Research Questions ....................................................................................................... 26

Chapter 3: Method ........................................................................................................ 27

  Participants .................................................................................................................... 27
List of Figures

Figure 1: Simplified figure of SCCT framework ................................................................. 9

Figure 2: The six interrelated developmental processes of the SCCT framework .......... 11
List of Tables

Table 1: Categories of Thematic Codes from Student Interviews.................................34

Table 2: Elementary and Middle School STEM Codes..............................................35
Chapter 1: Introduction

The higher education system in the United States is plagued with multiple systemic issues that inadvertently hinder talented students from underrepresented minority groups (URMs) and female college students from successfully graduating from selective colleges and universities with science, technology, engineering, and mathematics (STEM) majors (Bettinger, Boatman, & Long, 2013). Rice University has been no exception to the nationwide trend of high STEM attrition among students in these demographics (only 38% of Black and 57% of Mexican-American Rice students who matriculate as a STEM major actually graduate from Rice with a STEM major, compared to 67% of STEM-matriculating Rice students overall; McSpedon, Saterback, & Wolf, 2016). In 2012, the university implemented a summer academic program jointly funded by a National Science Foundation S-STEM grant and the university itself (the Rice Emerging Scholar Program, or RESP) for incoming STEM students considered "at-risk" for STEM underperformance and attrition (McSpedon et al., 2016).

Research suggests that formative student experiences (operationalized in this paper as experiences that occurred prior to high school) may predict students' identity as STEM students in college (Lindahl, 2007). STEM identity can be conceptualized, among multiple other methods, as the extent to which students believe they are competent in, perform well in, and are recognized by others as proficient in one or more STEM fields (Carlone & Johnson, 2007). Students’ self-efficacy, or belief that they are capable of performing a task to meet their own performance objectives (Bandura, 1994), is also particularly valuable to study in the context of underprepared students. As students transition from less competitive high schools into academically demanding STEM
coursework, it would be valuable to uncover whether certain formative experiences buffer or exacerbate students' sense of self-efficacy and STEM identity under circumstances in which these self-perceptions may be challenged.

The research project described in this paper involved completing 35 semi-structured interviews with Rice STEM students, qualitatively analyzing transcripts of these interviews, and then developing a thematic qualitative data structure model that will be used in the future to design a biodata measure of students’ STEM formative experiences.

Both Rice University and future research on STEM students’ success will benefit from the findings of this study. The mechanisms by which students’ interests in STEM topics and careers change over the course of childhood are not well-understood (Archer et al., 2010), yet I have found almost no research that examines specific formative experiences of students in STEM fields in particular or of their experience as URMs or first-generation college students, and none involving subsequent measure development. Consequently, this study will provide qualitative insights into STEM formative experiences that the field is currently lacking by developing a data structure model that will be used in a future study to generate and validate items on a biodata measure. In the future, I plan to validate the measure on Rice STEM students, as well as a broader sample of STEM students. For Rice students, my goal is to predict which students will succeed as STEM majors at Rice. For the broader sample, I plan to determine the scale’s predictive validity for STEM identity and self-efficacy, which will contribute to the larger research literature aimed at increasing underrepresented groups’ presence in STEM careers in the United States.
Academic Underperformance of URM Students

The national retention rate of college students in STEM majors is below 50% overall, and is even lower for underrepresented minority (URM) students (Wilson et al., 2012). Similarly, the percentage of women of any racial or ethnic group declaring a STEM major is less than half that of White males (Riegle-Crumb & King, 2010). The precursors to this shortfall appear long before students make STEM major decisions in college: URM high school graduation rates are consistently lower than those of Asian and White Americans. As of 2011, 91% of Asian American students and 88% of White students of high school age graduated from high school, whereas only 78% of Black and 71% of Hispanic American students did (Kim, 2011). Examining differences in academic outcomes from a younger age, URMs are especially underprepared in math, a discrepancy that can be seen as early as age nine (Lent, Lopez, & Bieschke, 1991).

Further, not only is there a national shortage of qualified math and science teachers in the U.S., but the most poorly-prepared teachers disproportionately teach in the highest poverty and highest URM-percentage school districts (May & Chubin, 2003). As a result, students in these schools undoubtedly experience different and perhaps less academically enriching environments in these classrooms from the day they enter elementary school compared to students in other school districts. Students raised in rural and large urban areas (where URMs are overrepresented demographically) also show lower academic performance starting in preschool than children raised in small urban and suburban areas (Miller & Votruba-Drzal, 2013).

In contrast to the academic trend of URM underpreparedness compared to other students, female students of all backgrounds tend to perceive math and science
underperformance compared to male students despite a lack of evidence. For example, female high school students report lower levels of proficiency in math than male students despite negligible to nonexistent gender disparities in class performance (female students actually outperform male students in most academic subjects throughout their primary and secondary school careers, including most math and science classes; Riegle-Crumb & King, 2010). In their overview of the research on female students in STEM, Rittmayer and Beier (2009) discuss self-efficacy, influenced by mastery experiences of successful practice; vicarious experiences of observation of others’ performance; social persuasion (the influence of outside judgments and feedback); and physiological (affective and physical) reactions to STEM-related experiences as important directions for future research on females in STEM to explore.

A recent meta-analysis by Sheu and colleagues (2018) found a two-factor solution for the sources of self-efficacy of STEM students, with mastery experiences, persuasion, and physiological state contributing to a single “direct personal experiences” factor, and a second factor resulting from vicarious experiences. Clearly, there is a need to explore not only the direct goal of increasing STEM major retention among underrepresented college students, but also to examine how early life experiences influence students' paths toward declaring a STEM major in spite of any additional internal or external hurdles the students may have faced when they were younger.

The underrepresentation of URMs in STEM majors is most strongly correlated with high school underperformance, rather than ethnic or racial identity. Research suggests that the two best predictors of college completion are: 1) entering college immediately after graduating from high school and 2) taking a high school curriculum
that requires skill in reading at grade level and completing at least one math course beyond basic algebra (Chen, Wu, & Tasoff, 2010). Socioeconomic status (SES) also predicts high school dropout: in 2014, low SES students (those in the bottom quartile of US household income) were almost twice as likely to drop out as the middle 50% of students by SES (11.6% dropout rate compared to 6.15% for mid-SES students) and over four times more likely than students in the top SES quartile (2.8% dropout rate; McFarlan, Cui, & Stark, 2018). A study by Riegle-Crumb and King (2010) found that disproportionate gender and race differences in physical sciences and engineering majors are markedly reduced when the student’s level of high school preparedness is controlled for statistically. Further, URMs are disproportionately not enrolling in college right after high school, come from less competitive high schools, and are more likely to be from low SES backgrounds (Brock, 2010; Kim, 2011). Additionally, for female students, underpreparedness for college STEM coursework may contribute to STEM underperformance and attrition. Data from the National Center for Education Statistics showed that the highest math and science classes females of all backgrounds took in high school were significantly less academically advanced than classes taken by White and Asian males (Riegle-Crumb & King, 2010). In context of this correlational evidence, attempting to counteract any of the disadvantages that URMs and females face prior to college might improve their college STEM retention outcomes.

**STEM Retention**

The majority of college major switching occurs from STEM majors into non-STEM majors rather than the reverse pattern (Seymour & Hewitt, 1997). One of the consequences of this trend over the long term is that the U.S. and many other developed
nations are facing a shortfall of qualified college graduates within multiple STEM-related subfields, many of which have job growth rates and salaries that are higher than the national average (Maltese & Tai, 2010). Female and URM students are of particular concern in terms of STEM persistence in college. For example, URMs have lower college graduation rates overall, and both women and URMs have lower representation in STEM fields in particular (Murphy, Gaughan, Hume, & Moore, 2010). Both groups tend to encounter at least two additional barriers to retention in STEM majors: lack of support from educators themselves and lack of university support (Murphy et al., 2010). The underrepresentation of URMs and females in STEM fields is often conceptualized as a "leaky pipeline," wherein these students are "lost" somewhere along the path to successful STEM major graduation (Murphy et al., 2010).

Quantitative analyses of STEM retention at the post-secondary and high school to post-secondary transition are the subject of ongoing work by STEM researchers. Studies have ranged from STEM coursework interventions (see Torres, Saterback, and Beier (2016) for research on the effect of a project-based learning course for engineering students on GPA and Beier et al., (2018) for the effect of project-based learning courses on STEM college outcomes) to longitudinal studies of predictors of college STEM outcomes (see Ackerman, Kanfer, and Beier (2013) for predictors of STEM major attrition by a range of individual differences). Analyses of STEM grades for students at risk for school underperformance, including whether they participated in a STEM summer bridge program, have also been conducted (see Bradford et al. (2018)). However, literature on STEM experiences before high school is relatively sparse, for both quantitative and qualitative studies.
One of the most well-known qualitative studies of STEM college students was by Seymour and Hewitt (1997), who conducted interviews with both URM and White students who persisted in or changed from a STEM major. They primarily captured students’ experiences in college, with some discussion of high school experiences. The authors found a variety of themes among URM switchers. One important cause of STEM attrition was URM students' experienced discrepancy in culturally-driven academic expectations as they entered college. For example, STEM professors may expect only a portion of their class to be intelligent or motivated enough to understand the material; conversely, students of certain backgrounds may be entering with the mindset that they should be able to earn high grades with enough preparation and faculty support (Seymour & Hewitt, 1997). URM students were also more likely to report that they held an incorrect perception before matriculating of how much preparation or effort was required to succeed in STEM majors, as well as a lack of access to broader STEM education, meaning that many of these students entered college with a narrow STEM interest and found that their broadly-focused entry-level classes did not teach the application of the topics to practical problems (Seymour & Hewitt, 1997). URM and first-generation students are also more likely to have insufficient financial resources and greater family obligations, such as the need to work a part-time job year-round, which add additional constraints to their ability to focus their time and effort on their STEM classes (Petty, 2014). Similarly, Próspero and Vohra-Gupta (2007) found that first-generation college students worked more hours than non-first-generation students and were more likely to drop out by the end of their second year. Inadequate high school preparation was also frequently cited by URM students as a reason for switching out of STEM, and these
students entered college overconfident because they had succeeded in less academically rigorous STEM classes in high school (Seymour & Hewitt, 1997), which could have impacted these students’ self-efficacy and STEM identity after entering college.

It is important to note that female and URM STEM students do not have homogeneous experiences in college. For example, Black students, especially Black women, struggle with having few role models, perceived inferiority in their ability, and exclusion from other peer groups (Seymour & Hewitt, 1997). All STEM students benefit from peer support; however, Native American and Hispanic students especially benefit from the affirmation and emotional support group membership brings because it is similar to their own cultural values of family support and kin networks (Seymour & Hewitt, 1997). Conversely, Black and Asian American students tend to not seek out peer groups, albeit for different reasons. For example, certain Black peer groups have shown a trend of offering support or commiseration for failure but not celebrating good performance; Asian American peer groups may ostracize those who are performing poorly (Seymour & Hewitt, 1997). In conclusion, college students of different ethnic backgrounds and gender identities have different experiences in college, a finding that could conceivably be traced to their diverging academic and life experiences long before college matriculation.

**Formative STEM Experiences**

Drawing from the child development literature, childhood experiences that have long-term academic ramifications may be both specific events that students experienced or a more general context of their background, upbringing, or other broader environmental or social factors that impacted their childhood (Miller & Votruba-Drzal, 2013). These
formative experiences can be viewed from a social cognitive career theory perspective (SCCT), which is a framework describing how students develop and build on academic and career interests, make choices regarding these interests, and eventually succeed or fail as a result of these choices (Lent, Brown, & Hackett, 1994). See Figure 1 for a condensed representation of the SCCT model.

![Figure 1: Simplified figure of SCCT framework, based on Lent, Brown, and Hackett (1994)](image)

Much of the first three stages of the SCCT framework – acquiring self-efficacy, developing career-related interests, and building career-related aspirations – take place during a child's elementary and middle school years, though there may be some additional development that happens in their high school years and beyond (Lent et al., 1999). Children's self-efficacy predicts a wide range of behavioral and cognitive outcomes, including resilience in the face of difficulties, motivation, causal attributions of successes or failures, and strength of commitment to aspirations (Bandura, Barbaranelli,
Caprara, & Pastorelli, 2001). Lent et al. (1994) note that efforts to improve students' sense of self-efficacy can impact their perceptions, goals, and other cognitive processes, indicating that important early life experiences can have a lasting impact on one's longer-term academic and career decisions. For example, there is evidence that parents' self-efficacy in believing they are capable of impacting and encouraging their children's academic strengths influences their children's own academic self-efficacy, which may in turn ultimately impact the development of their children's occupational interests (Bandura et al., 2001). Within the SCCT framework, one might expect that students experience major events that impact these first three stages on their path from childhood to eventual career pursuit well before entering college. Further, any changes students experience in self-efficacy can be domain-specific but do not necessarily need to be; a change in self-efficacy in a general skill area, such as teamwork ability or goal-setting skills, can "spill over" into increased self-efficacy in specific domains, such as math and science school classwork (Lent, Hackett, & Brown, 1999). SCCT research continues to support the importance of self-efficacy, including studies on STEM students in particular. For example, Lent and colleagues (2008) tested students at predominantly White and predominantly Black universities in entry-level engineering classes and found support for self-efficacy predicting students’ interests and goals five months later. In another sample of engineering students, self-efficacy predicted academic satisfaction and intended engineering persistence two semesters later (Lent et al., 2015). See Figure 2 for a model of the six processes of the SCCT model.
Early life experiences may be explored through qualitative interviewing, a form of data collection resulting from interactions with individuals that prompts the interviewee to provide rich, detailed responses that the researcher will later analyze to either test *a priori* hypotheses or to synthesize exploratory findings (Warren & Karner, 2005). In the SCCT framework, formative experiences (termed "sources of self-efficacy" in Figure 1) are neither the beginning nor the end of the model: experiences are influenced by individual-level factors such as demographic characteristics and personality traits along with the context of the experience, which are in turn theorized to impact behavior indirectly via first impacting psychological mechanisms including self-efficacy expectations and outcome expectations, which ultimately predict students' interests and goals.
There are few qualitative studies on STEM college students’ experiences in general, and fewer still on the experiences of URM students, first generation college students, and students entering college at risk for being academically underprepared in STEM fields. One exception is by Zani and Nogueira (2006), who conducted a qualitative analysis of critical incidents that nursing students experienced in interactions with faculty members and grouped the critical incidents into categories such as “teacher hinders student’s creativity” and “teacher encourages student during the learning process.” The authors were able to develop meaningful recommendations on suggested teaching strategies from this qualitative data. A study by Stokes, Levine, and Flessa (2014) conducted qualitative interviews with 29 White and Hispanic students in geosciences at the University of Arizona in order to determine positive and negative incidents these students experienced that influenced their decision to pursue a degree in geosciences, their self-confidence, and their positive perceptions of the major. Though this study did not explore formative experiences, it did find different types of familial incidents between Hispanic and White students, which highlights underlying cultural differences and provides evidence that factors external to STEM students may nevertheless impact students’ individual STEM outcomes.

Though I could find no studies that examined specific formative experiences of students in STEM areas specifically, there is ample evidence regarding how children and teenagers conceptualize various STEM fields and STEM experiences, as well as research on the relationship between certain formative experiences and children's overall academic achievement. These studies support the predictions of SCCT regarding the downstream impact of experiences prior to college matriculation, especially in regard to self-efficacy.
For example, O’Brien, Martinez-Pons, and Kopala (1999) conducted a correlational study of eleventh-grade students and found that mathematics self-efficacy predicted students' career interest in STEM fields, and that mathematics self-efficacy was predicted by students' academic achievement, ethnic identity, and SES. URM students, those from lower SES backgrounds, and females were all more likely to have lower mathematics self-efficacy than other groups (O’Brien et al., 1999). These findings imply that experiences the students had before the eleventh grade influenced their self-efficacy, which may have long-term ramifications for students' interest in math and science.

Examining academic experiences from an even younger age, there is evidence that most students report being interested in science until about the age of ten, then interest declines among many children, especially girls, though the mechanism driving this shift in attitudes is unknown (Lindahl, 2007). Further, research suggests that students' perceptions of science are largely established by the time students are fourteen (Lindahl, 2007). In a similar finding, a large-scale study by The Royal Society (2006) of current STEM practitioners (termed "SET", for science, engineering, and technology) found that approximately one-third of respondents began considering a science, engineering, or technology career before the age of 11, and another third by the age of fourteen. Joyce and Farenga (1999) found that gifted elementary school students (above the 90th percentile on both math and science achievement tests) had decided whether or not they liked science by the time they were nine. These studies provide further support for the claim that childhood experiences and perceptions have long-term effects on individual STEM outcomes.
Development of STEM Identity

At its core, identity can be conceptualized as both one’s subjective experience of reality as well as the awareness that others (as a collective whole) recognize this sense of awareness within that individual (Erikson, 1994). Science identity has been shown to predict science career commitment, in addition to self-efficacy (Chemers et al., 2010). Broadening this construct to include all STEM fields, STEM identity is an integral construct to consider when attempting to understand why students do or do not self-identify as STEM researchers. An academic understanding of identity lies at the crossroads of sociology and psychology, and the joint Personality and Social Structure Perspective (PSSP) model is a widely used framework to conceptualize and study identity at three distinct levels (Cote & Levine, 2014). The three levels are 1) social identity, which is impacted by cultural expectations and social norms; 2) personal identity, which is impacted by primary relationships and interactions; and 3) ego identity, which is a unified, continuous sense of identity similar to personality (Cote & Levine, 2014). Social identities are fundamentally shaped by an individual's demographic characteristics, including gender, race, and social class and represent how people relate to these social categories (Frable, 1997). They become a particularly salient concern for individuals when overarching cultural beliefs and ideologies conflict with their personal or group experiences, creating a struggle to build meaning and purpose within this contradiction (Tajfel, 1978). The impact of this juxtaposition between a society's larger experienced reality and an individual's own experienced reality is a particularly important consideration when evaluating identity within diverse, underrepresented groups in STEM majors and careers. Underrepresented STEM students may have built identities that on all
three levels of the PSSP model are markedly different from other Rice students’ identities (in 2016 the Rice undergraduate student population was 61% White or Asian American; Rice University, 2016) as a result of having experiences and influences in their lives that other students did not. The three levels of identity could have influenced these students' self-efficacy and ultimately their STEM interests in a way distinct from other students.

Similar to the development of self-efficacy and academic and career interests, identity formation begins long before students apply to college as a STEM major. Based on the previously-discussed evidence of the academic impact that pre-high school experiences can have on children and that certain academic and STEM behaviors seem to be “set” by a certain age, it seems reasonable to expect that experiences and beliefs students have in middle and elementary school will have a greater impact on their STEM identity than experiences in high school. As an example of pre-adolescents being a critical age to research, Archer and colleagues (2010) conducted focus groups on science interest with students aged ten and eleven from an identity lens, postulating that declining interest in science after this age results from a divergence among students' perceptions of science, the common methods of teaching it, and how perceptions of science interact with students' developing self-identities. In terms of gender differences, the authors found that the children conceptualized "real" science in fundamentally masculine terms ("dangerous" and "risky," for instance), indicating that girls may experience an incongruence in identity as something that must be "endured" if they wish to pursue an interest in science (Archer et al., 2007). From a PSSP perspective, this may be a problem of an ego identity conflicting with a social identity.
However, there is evidence that STEM identity may be malleable among children. For example, Capobianco, French, and Diefes-Dux (2012) developed and validated a measure of engineering identity on pre-adolescent children and concluded that engineering identity in children is not fixed or pre-determined but develops as a result of students' experiences in school in general and with engineering-related tasks in particular.

In that same vein, however, STEM identity can also be impacted negatively. Of particular concern, teaching methods and school curricula that under-recognize the value that students from diverse backgrounds bring into academic settings may have negative long-term consequences for STEM outcomes. The "funds of knowledge" concept is that all students possess ingrained knowledge, skills, and wisdom that they bring with them into any new environment (Moll, Amanti, Neff, & Gonzalez, 1992). These "funds" may not be readily assessed, but they shape students' perceptions and worldview fundamentally (Moll et al., 1992). There is evidence that students from non-majority backgrounds experience a disconnect between a STEM identity and their own identity starting from a young age, as their funds of knowledge are not fully realized in the classroom (Barton & Tan, 2009). For instance, in elementary and middle school science classes, these students must navigate not only the material they are learning for class, but an entirely different way of conceptualizing science than they have learned from their own background (Aikenhead & Jegede, 1999). These experiences, if they are perceived by students as reflective of their own inherent abilities rather than as an additional challenge due to their background in a non-majority culture, may impact students’ development of self-efficacy in the topic, which aligns with the first stage of the SCCT framework of acquiring self-efficacy. To the extent that these students can reconcile their current identities with the
dominant ways of understanding, valuing, and practicing science, these students may be able to incorporate an identity as a science "practitioner" fully into a cohesive perception of themselves (Barton & Tan, 2009), which aligns with the SCCT framework of the second and third stages of developing and building upon STEM interests.

**Other STEM Outcomes**

Several studies of childhood academic experiences have provided certain general expectations for the types of experiences linking childhood STEM identity to other STEM outcomes at the university level that participants may report. One category of relevant childhood experiences includes peer interactions. Véronneau, Vitaro, Brendgen, Dishion, and Tremblay (2010) conducted a longitudinal study of Canadian students from grades 2 through 7 and found that higher academic achievement predicted later increases in peer acceptance, decreases in peer rejection, and increases in peers’ achievement from grades 4 and above. Further, peer rejection incidents predicted later decreases in academic achievement (Véronneau et al., 2010). Students appear to self-select into peer groups of similar levels of academic achievement by the time they leave elementary school, and peer rejection experiences can have negative implications for later academic achievement, independent of high peer acceptance experiences and peers’ academic achievement levels (Véronneau et al., 2010).

Parental influence also appears to play a role in creating life experiences that may predict children's long-term STEM outcomes. For example, a study of women in non-traditional careers (defined statistically based on 1987 US Department of Commerce data) compared to traditional careers found significant differences in these women’s childhood experiences (Coats & Overman, 1992). The analyzed career paths included
many STEM fields, such as medicine, dentistry, and science and math research, as well as executive business positions and litigation. Women who participated in competitive sports as children and those who received parental encouragement (particularly from their fathers) that promoted achievement and encouragement for their daughters to stand up for themselves were more likely to work in one of these non-traditional careers as adults (Coats & Overman, 1992). It is possible that these findings may not be applicable to male URM students or those of low SES backgrounds, but the experiences of childhood play and the themes of parental messages that students received as children are still worth exploring to see whether any of these formative experiences impacted students' choices to become a STEM major, their STEM identity, and their self-efficacy.

Other themes of broader teacher and community support, in addition to parental and familial support, may be relevant. Russell and Atwater (2005) conducted a qualitative study of 11 Black biology college seniors and uncovered several common themes, such as strong family support and teacher encouragement, which could possibly reflect formative experiences these students had. Brown (2002) conducted a similar qualitative study of Hispanic engineering students and found core themes that included family support, supportive teachers, having small classes prior to college, and coming from small communities, all of which could have shaped these students' early STEM experiences. Through qualitative interviews with geosciences students, Stokes et al. (2014) found variations in themes of family influence by race, as Hispanic students reported approximately twice as many incidents related to familial influence as White students, as well as more negative familial incidents. This finding indicates that the effect of negative experiences in particular with family members should be explored further.
Regarding the impact of teachers, further evidence that teachers influence student STEM outcomes comes from Zani and Nogueira (2006), who conducted a qualitative analysis of nursing students’ experiences in interactions with faculty members and grouped the incidents into categories such as “teacher hinders student’s creativity” and “teacher encourages student during the learning process” (p. 744). The authors were able to develop meaningful recommendations on suggested teaching strategies from this qualitative data.

Finally, children's exposure to STEM fields may also predict STEM outcomes. For example, Capobianco and colleagues' (2012) study on engineering identity on pre-adolescent schoolchildren found that deliberately educating children on what engineers do influenced their STEM identity (as operationalized by quantitative self-report measures), which may imply that education on the existence and job tasks of certain STEM occupations at the elementary and middle school level may impact future STEM identity and related academic and career decisions. Jarrett and Burnley (2007) found that early experiences, such as playing with construction toys like Legos and participating in out-of-class science experiences such as museum visits predicted geology undergraduate students’ science interest. Clearly, the impact of experiences of interest and exposure to one or more STEM fields may also influence students' STEM outcomes.

Finally, based on the SCCT framework, one might expect some students to describe experiences that are not necessarily related to their STEM field but changed their overall sense of self-efficacy in significant ways, such as through experiences of mastering a topic or task, receiving good grades in general, or receiving positive feedback from peers or authority figures. Cumulatively, these studies provide a strong foundation
for future research designed to understand why and how childhood experiences may ultimately influence STEM identity and self-efficacy.

**Biodata**

Various measures may be able to explore the link between students' early life experiences and their STEM identity and self-efficacy. In particular, autobiographical data, or biodata, may have predictive validity. Biodata consists of one's past behaviors and experiences and has been used to predict outcomes ranging from job performance to job retention to college GPA (Owens & Schoenfeldt, 1979). Biodata measurement has evolved over the years from collecting basic objective data such as number of siblings or education level to capturing more nuanced data such as preferences, hobbies, and interests (Schmitt & Golubovich, 2013). One strength of biodata is that measures tend to have high differential validity, meaning they have low or insignificant correlations with each other and with other predictive measures (Owens & Schoenfeldt, 1979).

Biodata has been used extensively in research on work populations, most often for employee selection purposes (Mumford, Costanza, & Connelly, 1996). However, several studies have shown promising predictive power for student academic outcomes as well (Schmitt et al., 2009). One early example is from Owens and Schoenfeldt (1979), who assessed high school attitudes of undergraduates. The authors clustered participants into subgroups based on response patterns and correlated group membership with college GPA and academic probation and dismissals. In other early studies, biodata on factors such as faculty ratings and extracurricular activities were used to cluster students into subgroups to determine relationships between subgroup membership and academic outcomes (Schmitt et al., 2009). In an example of more recent uses of biodata that have
evolved since this early research, Oswald et al. (2004) found in a study of 644 undergraduates that high school biodata provided incremental validity in predicting first-year college GPA, absenteeism, and self and peer-rated behavior beyond SAT/ACT scores and personality measures. In a study on college students, Schmitt et al. (2009) captured biodata that encompassed students' backgrounds, their interests and hobbies, and their typical behavior in a variety of situations.

Biodata may also be able to predict outcomes beyond objective academic performance. For example, though it was a work context, Mumford, Connelly, and Clifton (1990) found that subgroup biodata differences predicted motivational outcomes as well as objective performance outcomes. Other studies have found that biodata can predict behavior (e.g. turnover) and attitudes (e.g. job satisfaction; Schmitt & Golubovich, 2013), meaning it could be appropriate for a less performance-oriented construct such as STEM identity. A study by Jarrett (1999) regarding science attitudes found that elementary school science teachers’ opinions of their own childhood teachers, whether they played with specific toys as a child, and whether they had science-related hobbies predicted how participants currently felt about science, and participants’ science teachers in elementary school predicted (with small to medium effect sizes) their current science interest and their confidence in their ability to teach science (Jarrett, 1999).

Biodata has shown promising evidence in reducing adverse impact, or underrepresentation of certain groups in a selection context, which is particularly relevant for the students in this study, most of whom are URMs and/or female. For example, the use of biodata in college admissions tends to produce less adverse impact than relying solely on cognitive ability measures such as the SAT or ACT and provides incremental
predictive validity (Schmitt et al., 2009). Further, Levashina, Morgeson, and Campion (2012) found that requiring respondents to produce written responses to biodata items (an approach designed to reduce faking) increased the biodata measure’s predictive validity while maintaining the benefit of not causing adverse impact. Overall, biodata shows promise for future research in academic populations, including students from non-majority groups.

**Biodata Measures**

Biodata scale development is an approach to data collection that is designed to be as systematic and objective as possible (Mael, 1991). However, for several decades after its introduction, biodata development often failed to meet traditional standards for proving psychometric validity (Dean, Russell, & Muchinsky, 1999), and studies varied in whether they used empirically or rationally-designed scales. Empirically-designed scales rely on items’ proposed associations with the criteria of interest, and items are ranked and scored according to their proposed association with a reference group (Mumford & Owens, 1987); conversely, items generated rationally are generally based on the researcher’s knowledge of the field and understanding of what should rationally relate to the criteria of interest (Russell, 1994). Results on employee samples have been mixed in terms of whether rationally-generated scales predict job performance better than empirically-designed scales. The field of biodata has begun to evolve again, and many studies have shifted from purely retrospective analyses of biodata correlates to a reference group, to the intentional crafting of measures designed to capture certain constructs determined *a priori* (Schmitt & Golubovich, 2013). However, even in relevant studies (such as the biodata measurement of science teachers’ formative experiences in
science; Jarrett, 1999), biodata questions were based on topics covered in teacher training sessions and were not theoretically developed.

Ultimately, it is critical to understand both the experiences that URMs and females in STEM majors face as well as predictors of students’ STEM identity and self-efficacy, which may include experiences that occurred years before students entered college. Research on non-majority STEM students who have shown the capacity to succeed academically despite facing challenges other students have never encountered is necessary to further theory on STEM interest and identity development.
Chapter 2: The Present Study

Overview of Study

This research project involved 1) exploring Rice University STEM students’ formative experiences in STEM settings and their more general academic and life experiences before high school and 2) developing a qualitative data structure model to capture themes and experiences identified in the qualitative data and serve as a framework to guide item development on a future biodata questionnaire.

To achieve these objectives, I used semi-structured qualitative interviews, focusing on capturing students’ elementary and middle school life and academic experiences. I interviewed STEM students from a sample of those identified by Rice as those least prepared for Rice-level STEM coursework (see Beier, Saterbak, McSpedon, and Wolf (2017) for more detail on the selection process Rice uses to identify these students and McSpedon, Saterbak, and Wolf (2016) for details on an optional STEM summer bridge program Rice offers for these least prepared incoming STEM students). Next, I performed a qualitative thematic data analysis of the transcripts in order to elicit key themes and experiences students reported, and finally, I developed a data structure model from the thematic analysis. In a future study, I plan to use the data structure model to generate biodata items and develop a measure of STEM formative experiences, and then validate this measure to determine whether it predicts STEM identity and self-efficacy.

Qualitative Interviews

As previously discussed, qualitative research on STEM students is lacking, and it is almost nonexistent on the formative experiences of URM STEM students. Qualitative
data is particularly useful in capturing data that is difficult to collect objectively and/or is too complex for a single framework to fully describe (Creswell, 2007), which is appropriate for the current state of STEM formative experience research. Qualitative research has been increasingly conceptualized in the broader framework of participants’ political, social, and cultural contexts (Creswell, 2007), an emphasis that is especially pertinent to this study due to the participants’ differing racial, cultural, gender, and social identities. Qualitative research enables exploration of social issues and human experiences in greater depth than is possible through purely quantitative analysis (Creswell, 2007). For this study, qualitative analysis is particularly useful in the context of seeking the reasoning and context behind students’ experiences.

In an underexplored research area like STEM formative experiences, qualitative research can ensure that exploration of concepts precedes the development of constructs (Gioia, Corley, & Hamilton, 2013), and that it avoids imposing a priori assumptions (DiCicco-Bloom & Crabtree, 2006). Qualitative interviewing produces the most useful data when all interviewees have life experiences in the interview topics, enabling the researcher to make effective comparisons among interviewees (DiCicco-Bloom & Crabtree, 2006). As most of the students in this study come from one of a handful of largely similar backgrounds (female, first-generation, and/or URM), I expected their qualitative data to be effective for data analysis purposes for this reason. The qualitative data from this study can provide insight into the early experiences that have shaped students’ STEM decisions and perspectives profoundly, whether consciously or unconsciously.


**Research Questions**

I proposed three research questions for this study. First, I planned to explore but made no specific hypotheses about the patterns that students’ reported formative experiences would follow.

*Research Question 1:* What themes of major life and academic experiences do students report?

Second, during the data analysis process I expected to find themes of responses that could be clearly linked to certain well-established psychological constructs. Again, however, I made no *a priori* hypotheses about which constructs I expect to find, to minimize the possibility of biasing my qualitative analysis.

*Research Question 2:* What constructs will be associated with these formative experiences?

Finally, I made no specific hypothesis about the resulting data structure from the qualitative interviews, other than I anticipated that it would include both formative experiences and psychological constructs.

*Research Question 3:* What hierarchy of formative experiences and psychological constructs will best represent the data obtained from the qualitative interviews?
Chapter 3: Method

Participants

Participants \((n = 35)\) were Rice undergraduate students, all of whom qualified to participate in the Rice Emerging Scholars Program (RESP), which is Rice University’s STEM summer bridge program for incoming STEM students who come from less prepared academic backgrounds. RESP-qualifying students are those in the bottom 20% of the SAT or ACT scores and mathematics assessment scores of all matriculating students (McSpedon et al., 2016). The majority of RESP students are URMs and/or first-generation students who have graduated from under-resourced high schools that may not have adequately prepared them for college coursework, particularly STEM coursework, at the level of Rice’s course rigor (McSpedon et al., 2016). Students were randomly selected to be invited to interview. Overall, 62 students were contacted by email, resulting in a 56.4% response rate.

As expected, participants were highly academically successful in high school. They had an average high school rank in the top 3.02% of their graduating class \((SD = 6.06\%)\) and comprised eight valedictorians, an average ACT score of 32.10 (which is an approximately 98th percentile score; \(SD = 1.84\)) and had taken a mean of 2.43 STEM AP exams \((SD = 1.27)\) in high school.

The mean age of participants was 19.01 years, and 23 participants were female (66%), and 12 were male. Participants were racially diverse: 11 Black (31%), seven White (20%), six Hispanic (17%), four Asian (11%), one Hawaiian/Pacific Islander (3%), and five mixed-race (14%) students. Sixteen participants were freshmen and 19 were sophomores. Thirteen students were RESP participants and 12 were from the control
group of students who either declined RESP or who were not able to be invited due to limited program resources.

**Measures**

In order to capture participants' biodata for biodata questionnaire development, I developed a protocol designed to assess multiple types of qualitative data. To capture the most comprehensive qualitative data from participants, it is necessary to use open-ended and direct questions to elicit detailed narratives from the interviewees (DiCicco-Bloom & Crabtree, 2006). I included questions designed to elicit specific STEM-related memories (e.g. “What is your earliest memory of being interested in any STEM topic?”, “What were your experiences in math in elementary school like?”) as well as items that reflected the recurring themes in the research on formative experiences discussed earlier in the literature review, such as parental influence (e.g. “What role, if any, did your family members play in your decision to pursue a STEM major?”), teacher support (e.g. “What role, if any, did your teachers or other authority figures play in your decision to pursue a STEM major?”), and exposure to STEM (e.g. “During elementary or middle school did you participate in any other STEM activities besides those required by your school?”).

The interview protocol addressed, in roughly chronological order, students’ critical pre-college life, academic, and personal experiences, first in their elementary and middle school years, and then their high school years that may have impacted the student's choice to enter Rice as a STEM major. I structured my interview questions in a way that asked students to provide context and background information on their experiences instead of recounting the experience in isolation, per Creswell's (2007)...
recommendation, and probed for more context when necessary during these interviews. See Appendix A for the interview protocol.

I pilot tested the interview on four undergraduate students in the Adult Skills and Knowledge (ASK) lab, including a former RESP participant. Three of these students entered Rice as STEM majors, and all had STEM interests for at least part of their lives before college. I used responses to the interviews as well as feedback from these students to refine my questions for the final version used on participants and described in this paper.

**Procedure**

**Qualitative interviews.** Interviewees were emailed inviting them to participate in an interview that would take between 30 minutes to an hour, with compensation of twenty dollars. Interviews were conducted in the ASK lab, in a private room with the participant, using two audio recorders. Participants provided their written consent to participate in the study, including acknowledgement of the fact that their responses would be audio recorded, transcribed, and used anonymously for this and future projects. Participants were advised that they could pass any question they did not want to answer, though none chose to do so. As per Flanagan's (1954) recommendations on maximizing the useful information produced from interviews, before beginning the interview, the protocol explained to the interviewee the sponsor of the research, the general aim of the study, what group of interest the interviewee was in that merited them being invited to the interview (being a STEM student at Rice), and assurances that all responses would remain anonymous.
Qualitative data coding. Interviews were transcribed by undergraduate research assistants (URAs) in the ASK Lab. All interviews were completed before the coding process began. The coding process involved two URAs and me coding transcripts for relevant themes separately, with each URA being responsible for coding half of the transcripts. The URAs did not know the purpose of the study or its research questions. I also coded all transcripts, and then met one-on-one with each URA for consensus coding, during which we discussed and resolved any discrepancies in the themes we found. I used ATLAS.ti qualitative software to input all codes after the URAs and I had met, and after all codes were added, I began identifying broader categories and developing subcategories of codes using the software.

I used a thematic data analysis approach to guide my process of identifying themes in the data. Thematic data analysis is an approach to qualitative data analysis that involves identifying themes within a data set, organizing, and reporting findings (Braun & Clarke, 2006). It is useful for capturing similarities and differences between different participants (Braun & Clarke, 2006). One of its advantages is that it is an effective approach to use for developing a structure for data sets (which was aligned with my intention to develop a data structure model from these data, a process I describe on page 48; King, 2004). Another advantage of thematic data analysis is that it is useful as a “translator,” bridging understanding and helping unite theory between qualitative and quantitative researchers (Boyatzis, 1998), which also corresponded with my plan to build a quantitative measure from qualitative data. Further, my data analysis was inductive (rather than deductive), meaning that the data guided my subsequent analysis and data structure model. Inductive methods refer to letting the data guide theme development and
allowing data from different interview questions to be combined into themes if appropriate, rather than working “downward” with deductive analysis, which involves starting with a theory and looking for themes to support *a priori* proposed constructs (Boyatzis, 1998).

I used the most recent guidelines on the qualitative data analysis process described by Gioia and colleagues (2013) in order to develop a thematic data structure that would reflect the broader theory behind the themes found in the interview process and which would ultimately guide my biodata measure development. This approach is a well-established process designed to apply scientific rigor in data analysis without sacrificing the ability to develop new theories from qualitative data and has been continually refined since its first introduction in 1991 (Gioia et al., 2013). One of its core tenets involves accepting interviewees as “knowledge agents.” In terms of this research, it means accepting that students know their goals and can consciously frame past experiences, how those experiences impacted them, and their thoughts and intentions. Because all participants are high-achieving students, I believed that students met this assumption. The analysis involves proceeding from direct experiences of informants (first-order) to second-order themes which encompass broader concepts and theoretical constructs, and finally third-order aggregate dimensions, which broadly reflect higher-order themes found in the analysis and the current literature. Selecting this approach was also intended to provide a structured process to guide future biodata item development.

After I coded and reviewed the data, I developed six third-order aggregate dimensions, comprising two second-order themes and four first-order experiences within each second-order theme. I completed this process independently. Because the factor
analysis and scale validation I plan to conduct in the future will allow an objective quantitative exploration of whether and how the proposed model of second and third-order dimensions supports the data, I considered the structure I developed to be appropriate for the purposes of this study. The model will almost certainly be refined after future scale validation and factor analysis, but at this stage the model provides a theoretically sound foundation to inform subsequent quantitative analyses.
Chapter 4: Results

Qualitative Data Analysis

During the qualitative coding process, I divided the codes into several major
categories. Certain responses could be reasonably coded as Likert-scale items (e.g. “How
did you feel about your science class’s quality in elementary school?”), so I separated
those items to use in analyses in a more quantitative manner. Other codes (e.g. when
asked to describe the characteristics of their high school friends, almost all students
responded that their friends were either almost all STEM students or a mix of STEM and
non-STEM students). When all or almost all interviewees responded to questions in a
way that fell into mutually exclusive categories I separated these codes as well.

After separating these codes, 271 codes that represented responses that were
thematic in nature and were suitable for qualitative analysis remained. See Table 1 for the
categories and subcategories of themes, and Table 2 for the specific codes for elementary
and middle school STEM experiences and opinions. This is a relatively large number of
codes, which was a result of the fact that over the course of interviews, which were on
average about 30 minutes in length, responses covered a wide range of topics. These
interviews were unlike many other qualitative interviews designed to cover a narrower
topic or predetermined construct. For example, interview questions in this study spanned
almost the entire length of the interviewee’s life, covering academic, personal, and work
domains.
Table 1: Categories of Thematic Codes from Student Interviews

<table>
<thead>
<tr>
<th>Elementary School Experiences</th>
<th>High School</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience (General)</td>
<td>6 HS Experience</td>
</tr>
<tr>
<td>Influential Experience in STEM</td>
<td>3 HS Friends Experience</td>
</tr>
<tr>
<td>Math Experience</td>
<td>14 HS Teacher Influence</td>
</tr>
<tr>
<td>Science Experience</td>
<td>11</td>
</tr>
<tr>
<td>STEM Experience</td>
<td>22 Path to STEM</td>
</tr>
<tr>
<td></td>
<td>56 Barrier</td>
</tr>
<tr>
<td>Middle School STEM Experiences</td>
<td>18 Career Expectations</td>
</tr>
<tr>
<td></td>
<td>18 College Help</td>
</tr>
<tr>
<td>Family Influence</td>
<td>18 Helped with STEM</td>
</tr>
<tr>
<td>Family Experience</td>
<td>15 STEM Identity</td>
</tr>
<tr>
<td>Parent Characteristics</td>
<td>2 STEM Interest</td>
</tr>
<tr>
<td>Parent Expectations</td>
<td>12 STEM Motivation</td>
</tr>
<tr>
<td>Parent STEM Influence</td>
<td>9 Student Characteristics</td>
</tr>
<tr>
<td></td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Total 271</td>
</tr>
</tbody>
</table>

*Note.* Number is the count of subcodes within each theme.
Table 2: Elementary and Middle School STEM Codes

<table>
<thead>
<tr>
<th>Elementary School Math Experience</th>
<th>Middle School STEM Experiences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Didn't Think Math Problems Were Useful</td>
<td>STEM Classes Were Not Challenging Enough</td>
</tr>
<tr>
<td>Disliked Having to Solve Problems a Certain Way</td>
<td>Took Advanced Math and/or Science</td>
</tr>
<tr>
<td>Disliked Timed Math, Found It Stressful</td>
<td>Exposed to STEM Field(s)</td>
</tr>
<tr>
<td>Frustrated/Bored with ES Math Being Too Easy</td>
<td>Influential STEM Course</td>
</tr>
<tr>
<td>Liked Competitive Math in ES</td>
<td>STEM Extracurriculars</td>
</tr>
<tr>
<td>Liked Solving Difficult Problems</td>
<td>Volunteering in STEM Field/Activity</td>
</tr>
<tr>
<td>Necessary but Boring</td>
<td>Good/Encouraging Teachers</td>
</tr>
<tr>
<td>Only Liked When They Were Good at It Was Difficult but Getting Better Made Me Enjoy It More</td>
<td>Influence from Specific Teacher</td>
</tr>
<tr>
<td>Felt Behind in Math in ES and MS</td>
<td>Teacher Explained About STEM Opportunities</td>
</tr>
<tr>
<td>Helped Others</td>
<td>Active in STEM Clubs/Extracurricular Activities</td>
</tr>
<tr>
<td>Helped Siblings in Math</td>
<td>Enjoyed Hands-On Math and/or Science</td>
</tr>
<tr>
<td>Liked Doing Extra Math Problems</td>
<td>Enjoyed STEM Hobbies (Outside of School)</td>
</tr>
<tr>
<td>Liked Doing Well in Competitions</td>
<td>Enjoyed STEM Projects</td>
</tr>
<tr>
<td>Liked the Attention of Being a STEM Student Math Success</td>
<td>Started Planning for College</td>
</tr>
<tr>
<td></td>
<td>STEM Competitions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Elementary School Science Experience</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Did Science Projects on Their Own</td>
<td></td>
</tr>
<tr>
<td>Enjoyed Labs/Experiments in Class</td>
<td></td>
</tr>
<tr>
<td>Not Exposed to Science in ES</td>
<td></td>
</tr>
<tr>
<td>Teacher Gave Them Science Books to Read</td>
<td></td>
</tr>
<tr>
<td>Thought Science Fair Was Pointless</td>
<td></td>
</tr>
<tr>
<td>Science Fair Was Okay/Neutral</td>
<td></td>
</tr>
<tr>
<td>Took Science Fair Seriously/Enjoyed</td>
<td></td>
</tr>
<tr>
<td>Did Not Enjoy Science Until It Got More Advanced</td>
<td></td>
</tr>
<tr>
<td>Did Not Like Science Projects</td>
<td></td>
</tr>
<tr>
<td>Science Is Boring/Pointless</td>
<td></td>
</tr>
<tr>
<td>Science Was Not Advanced Enough</td>
<td></td>
</tr>
</tbody>
</table>

Qualitative Data Structure Model

The second and third-order themes of the data structure framework are discussed on the following pages. See Appendix B for the full data structure model.
**Math perceptions in elementary school.** Interviewees reported differing levels of motivation when working on new or competitive math assignments, which I aggregated into a “Learning/performance orientation toward math” second-order theme. Mastery versus performance orientation refers to whether one’s primary goal is to increase competence or knowledge in a domain or to meet external standards for competence and avoid being perceived as failing at a task (Ames & Archer, 1988). Below is an example of a student with a learning orientation for new math topics:

“*Math was more difficult in elementary school... [but] I kind of like embraced the challenges when I was little. It wasn’t something that was very discouraging, like it was just something I knew was hard.*”

The second theme I developed was “Level of understanding of math concepts.” This theme encompassed students’ depth of understanding of math concepts and their conceptions of the utility of math. Below is an example of a student’s opinion on the usefulness of learning math.

“*I didn’t like math a lot just because a lot of it isn’t super practical... I never really understood the point of some of it or, like, why we had to do it a certain way.*”

**Science perceptions in elementary school.** Interviewees reported diverse ranges of exposure to science in elementary school, ranging from occasional experiments to daily exposure to a science curriculum, which I developed as a “Quality of science education” theme. Below is an example of a student who reported little exposure to science in elementary school.
“I wasn't really exposed to it [science in elementary school] very much... I was talking to some friends and they were like...yeah, we used to like do science experiments when we were little. And I was like, yeah, I just don't remember doing anything with science until like late middle school or early high school.”

They also reported science interests that began at different periods in their lives, from pre-school and younger to elementary school or middle school, which I developed as an “Intrinsic interest in science” theme. Below is an example of a student who preferred STEM-related media from an early age:

“[My STEM interest started] before I started going to school, I guess when I was like five years old. It was because of TV... I was always watching the news with my family, and my mother and I would always try to see broadcasts of space and physics-associated things.”

**Classroom experiences.** Interviewees reported influential classroom experiences, including interpersonal experiences with both classmates and teachers, described as “Interpersonal classroom experiences.” Below is an example of behavior from a teacher encouraging STEM in elementary school.

“[I had] one teacher...that really heavily pushed science. We would do two science lessons a day, two math lessons a day. He was very much a STEM-oriented teacher, so I felt like he did a very good job. He really wanted us to deeply understand what we were doing, and he pushed us to go beyond the textbook.”

Interviewees also reported different levels of quality of their science and math education in elementary school, as well as whether more advanced options were
available. One student reported their perception of working on an advanced STEM curriculum:

“I was in... a GT program at my school, it was like gifted and talented, so it like selected out a few students and so we would do like specialized math in there, like a little bit accelerated math... I always thought that it was interesting, and the lessons were always really fun.”

**STEM identity.** Interviewees reported different experiences that impacted their STEM identity, including themes of “Fixed or growth mindset” and “vocational interest type.” A fixed mindset is one in which an individual believes his or her intelligence and abilities are fixed; a growth mindset allows for the belief in one’s ability to grow in their intelligence and abilities with enough persistence (Dweck & Leggett, 1988). One student reported evolving from a fixed to a growth mindset in the following quote:

“My brother was the really smart one in the family, so I didn’t know if I could do it [STEM], and I started to doubt myself... But over time it was kind of like I wanted to prove people otherwise, because I’m smart enough to do it.”

Many students showed a consistent vocational interest in preferring problem solving (Investigative) and/or hands-on (Realistic) activities. One student reported their preference for hands-on activities, which corresponds to the Realistic vocational interest type, which is a preference for manual manipulation and a concrete approach to problem solving (Holland, 1997).

“Me and my brother used to like play around with Legos... I was like, oh, this is interesting, like, this is better than my dolls. And we would build stuff and that was cool and that was my first experience. And in school we had this rocket
building unit in fifth grade, and I thought that was interesting, and it was cool because we shot out little rockets... That was kind of how I was introduced to it [my STEM interest], and I got really into chemistry later on and it just kind of combined.”

The Investigative vocational interest type corresponds with a preference for working with data, solving problems, understanding information, and showing curiosity (Holland, 1997). One student discussed independently researching science topics as a child.

“I was really interested in like space and NASA and stuff, like I always did my research in that growing up. In elementary school I read encyclopedias about space and stuff astronomy all sorts of things; I was really into that.”

Perceptions of parents’ expectations. Interviewees reported having a strong influence from their parents in their STEM interest development. I coded one second-order theme as “Active engagement in STEM with parent,” which referred to students who received help with STEM homework or participated in STEM hobbies with their parents. One student talked about how their parents encouraged a science interest:

“We, like, my parents, would play games, and they didn't want me to be afraid of bugs or anything, so we would like, if I ever found a bug instead of squishing it or throwing it away, you would put it under the microscope and you would like look at it... We played with microscopes a lot.”

The second theme was “Parental encouragement of STEM interests,” which referred to how much support and encouragement interviewees received from their parents in pursuing a STEM career or interest in particular.
“I used to like reading a lot, but my dad was like, you should do STEM, like engineering, because engineering is money, and stuff like that. He actually wanted me to do electrical engineering [in particular].”

**Exposure to STEM.** Interviewees reported varying levels of exposure to STEM career options and awareness of professionals who worked in STEM careers. The first theme, “Exposure to STEM topics/people,” referred to having close relationships to adults in their lives who worked in STEM. One student reported on an older sibling:

“My brother is ten years older than me, and when I was in 5th grade...he was already in college, and he majored in electrical engineering. And so just from being around him, that’s when my interest was sparked.”

The second theme, “STEM career awareness,” referred to gaining knowledge or receiving encouragement on specific STEM careers. One student reported their first motivation to become an engineer:

“I wasn’t very social, so I would go eat lunch with [my teacher], and I’d talk about life and about my interests, and he told me at one point, like, oh you’ll be a great engineer. So, I was like...now I need to be an engineer!”

In general, many students, perhaps due to the range in their family composition, socioeconomic status growing up, and where they grew up, provided responses that were almost the exact opposite of other students’ responses. For example, students ranged widely on how they felt about competitive math assignments in elementary school. One student reported finding these assignments stressful:
“I remember like we used to do like things like multiplication races and I would like never ever finish like really early, like everybody else would, and it was just... harder to like pick up the concepts and stuff.”

However, other students found competitive assignments motivating:

“We... [had] the First in Math online program that was optional, and I started participating in it because the teacher would give rewards, because it was kind of like a rewards program, like the more math problems you do... you get more points. And I think like every 1,000 points you got, the professor would treat you to a McDonald’s meal... I think that’s how I got started [being interested in math].”

Having a wide range of responses in the qualitative interviews is helpful in establishing that a wide range of responses to certain items can be reasonably expected in the biodata measure I plan to develop in the future.

The second purpose of this study was to identify and implement a qualitative data analysis that would provide a structure to inform future biodata item creation. Because there is no generally accepted process for systematically creating biodata items (Schmitt & Golubovich, 2013), and content and construct validity have been unreserach in biodata validation (Allworth & Hesketh, 2000), I developed the approach described in this paper independently, with the goal of following a process that was informed by theory, unlike most biodata scales. The data structure model I selected is a broadly accepted qualitative data process used to derive constructs, theoretical interests, and more narrow topics of interest and will serve as a guide to item development. This was in line with Mumford and Owens' (1987) recommendation, who noted that although it rarely
happens in practice, content validity should be established by specific hypotheses concerning domains of interest, and then biodata items generated. The approach I used in this paper, in addition to providing meaningful qualitative data on STEM students’ formative experiences, also serves as a strong theoretical bridge to future biodata item development.
Chapter 5: Discussion

This study used qualitative interviews with Rice STEM undergraduate students to collect details on their formative STEM experiences, after which I qualitatively analyzed and structured the data in a framework suitable for guiding future quantitative biodata item measurement and validation.

During the interview process, I found that most students were able to provide distinct themes that were useful for qualitative data analysis. They could readily and clearly convey influential childhood STEM experiences and could provide without prompting contextual data relevant to the experience as well as their own perceptions of the significance of the event or experience. Because no student reported formal exposure to engineering or computer science as an academic class during elementary or middle school, all the STEM themes I explored in this project referring to classroom experiences are science and math topics.

Addressing Research Question 1, a discussion of the major themes follows. As expected, during these interviews I found themes of parental support and influence, extended family influence, teacher support, peer interactions, and exposure to STEM fields. First, students tended to report receiving clear communication from their parent(s) growing up about their academic expectations. Not all parents communicated a preference for their child entering STEM in particular in college, but almost every interviewee reported that their parents clearly communicated that attending college in general was expected of them from at least elementary school age and sometimes earlier than that. Although some students did report some influence from their extended family, usually in either the identity of being the “smart” one in the family or through pressure to
become a STEM major (or, in this sample, often the pressure was to become a doctor in particular), almost all the interviewees reported that their parents buffered them from any strong or negative pressure from their extended family.

Second, most students reported strong, positive relationships with at least one teacher in elementary school or middle school. The teacher was not always a science or math teacher and did not even necessarily provide extensive one-on-one support (though many students did report these experiences); what appeared to be most important was the availability of positive, supportive teachers who seemed willing to work with and encourage students who needed further attention. Several students reported being grateful to one or more teachers who identified them as talented and offered them advanced assignments more suited to their ability level.

The impact of peer interactions in elementary school and middle school occurred primarily as either prosocial helping behaviors, in which students proactively helped their classmates or friends with science or math assignments, or as receiving an identity as a STEM student or as especially intelligent from their peers, sometimes as early as elementary school. Teachers sometimes also contributed to students’ perceptions of themselves as being high in STEM ability or intelligence by making comments to the student or in front of the student. Many students in this study explicitly noted that these external perceptions influenced their own identities and subsequent behavior (usually through increasing academic effort in order to support the expectations these students perceived from others).

Every interviewee articulated their first clear STEM interest as occurring before eighth grade, although the exposure came from a wide variety of sources. Some students
reported an influential science book or TV show, others reported learning about a new topic in science class or being told they would be a good fit for a certain STEM field (often medicine or engineering), and for others attending a career day or college major day during elementary or middle school sparked their initial STEM interest. Several students reported an older sibling majoring in STEM as being influential to their own STEM plans, in some cases by learning of the existence of a specific STEM topic for the first time, and in other cases by gaining confidence that if their older sibling was successful in college and STEM, then the interviewee could also succeed in STEM.

Conversely, several themes arose during the qualitative analysis that I had not anticipated from my prior review of the literature. The primary difference was that math and science experiences in elementary and middle school did not follow parallel paths, and students did not report the same general experiences in both classes. One of the primary distinctions between students’ math and science classroom experiences was that math was problem solving-oriented and took place every day, but for most of these students, science at the elementary school and middle school level was primarily fact-based, with little to no opportunities for problem solving. Also, many students reported either no or very little science education at the elementary school level, which may be a result of the fact that many of the students in this study attended underresourced public elementary schools that may not have had the resources to offer extensive science materials. Consequently, if students reported liking (or even having an opinion on) science in elementary school, the opinion often came from outside or self-directed exposure to one or more science topics.
Participants also differed in their opinions of learning math and science, and these opinions did not necessarily correlate. Some students explicitly stated that as a child they believed learning math would be useful for their future goals or enjoyed it either for its own sake (and enjoyed solving difficult problems) or for the external benefits of doing well in math (such as receiving praise from teachers or classmates), but other students found math unimportant for their future goals, boring, or difficult and therefore unenjoyable. Some students also did not enjoy math because they resented being forced to solve problems the way the curriculum taught. Finally, many students reported math class being a competitive experience, frequently with speed-based assignments, other students being aware of each other’s performance, and external rewards for high performance.

Conversely, science class was almost entirely lecture-based, though multiple students reported science experiments as being one of their favorite experiences in elementary and middle school. Many students who reported enjoying these experiences stated explicitly that they liked seeing the application of theory into practice. However, a few students were unsatisfied with these science experiments, perceiving them as not advanced enough to gain a deeper understanding of the topic or to learn something useful or practical. Some students identified a clear and enduring science interest in elementary school, but many students did not receive exposure to a science topic in enough depth to influence their decision to major in STEM in college until middle school or even high school.

Regarding engineering opinions, because no student reported formal exposure to engineering education, the student usually retroactively identified their interest after they
had learned the purpose of engineering and conceptualized their interests as falling into this category, or by being explicitly told by a teacher or parent that they were a good fit for becoming an engineer. Finally, some students identified their STEM interests as being related to becoming a physician rather than a specific STEM subject interest. Unsurprisingly, these students often had STEM interests, but students were usually oriented toward fulfilling the steps necessary to obtain a medical degree, and their STEM interests were secondary to this goal (in fact, several students in this study who switched out of STEM majors still planned to attend medical school after graduating from Rice).

After coding all the themes found in the qualitative analysis, I addressed Research Question 2 by using the qualitative data structure method developed by Gioia and colleagues (2013) to organize the many themes in this underexplored research area and to gain insight into the relevant theories that apply to early life experiences. See Appendix B for the thematic data structure model. Finally, addressing Research Question 3, I used this structure to guide how I will approach designing a biodata questionnaire on STEM formative experiences. I describe the proposed biodata development and subsequent validation process in the next section.

**Future Directions**

My next steps for this line of research are to 1) create a biodata questionnaire on STEM formative experiences and 2) validate the questionnaire on multiple samples, including Rice STEM students from more traditional backgrounds as well as non-Rice STEM students from historically underrepresented backgrounds.
Planned Biodata Measure Development Process

First, I plan to use the first-order categories established in the qualitative analysis of this study to guide item development. This approach will allow me to build items that capture different ways participants responded to or encountered an experience that would be associated with the construct of interest, which is critical to theoretically sound item development (Mumford et al., 1996). As recommended by Owens and Schoenfeldt (1979), the questionnaire will capture both “input variables” (what was done to the participant) and “prior behaviors” (behaviors that were in the participant’s control). For the purposes of item creation, I plan to consider “input variables” as consisting of actions of parents, teachers, or other students, and “prior behaviors” as consisting of a wide variety of actions, such as homework helping behaviors, class preferences, problem solving approaches, and participation in certain math and science-related activities. Whenever possible, the items I develop will be based on themes and experiences that participants themselves provided (e.g. frequency of participation in science fairs, whether students attended STEM-related summer camps).

As suggested by Mumford and Owens (1987), I expect many or all item responses I generate to fit on a continuum (such as responses for how frequently something occurred in the past). Also, per the authors’ recommendations, I plan to describe discrete situations when possible and ask for information that would be theoretically verifiable, which is designed to reduce the likelihood of faking. If the item addresses a topic that the respondent would reasonably not know or not remember, I will provide an “escape” option, per the guidance of Owens, Glennon, and Albright (1962), and I will describe
items and responses in either neutral or positive terms, rather than negative terms, as advised by Mumford and Owens (1987).

**Planned Scale Validation Process**

The second stage of the planned biodata development process is to validate the measure I develop. It is possible that themes or individual items will not all predict or will correlate differently with the dependent variables of interest. Another goal of this measure is to establish incremental validity over certain scales of particular relevance to the SCCT model, including self-efficacy and STEM identity, and ultimately determine the biodata’s predictive validity for STEM retention (successful graduation from college with a STEM degree) as the distal outcome of interest. Further, I plan to determine whether these measures will be applicable to a wider population beyond only RESP-qualifying Rice students. For example, studies of academic resilience have shown that some predictors of academic success, such as social support, provide proportionately more benefit for low SES and/or URM students than for other students (Luthar, Cicchetti, & Becker, 2000). I might find that this trend also applies to STEM outcomes, meaning that some factors will not correlate with self-efficacy, STEM identity, and/or STEM retention for higher SES or non-URM student samples.

There are multiple ways to score biodata measures, including empirical scoring, rational scoring, and the factor-analytic (or internal) approach (Schmitt & Golubovich, 2013). I plan to design a measure that could be used with the third approach, which is the most recent development in the literature. Factor-analytic scoring has been used to group respondents with different life histories into subgroups, and it can be both developed from theory and the results then used to further theory (Owens & Schoenfeldt, 1979).
whereas empirical and rational scoring are more atheoretical (Schmitt & Golubovich, 2013). Because there is little research on the academic achievement of STEM students from less prepared backgrounds, I would prioritize the ability to generate items that reflect the underlying constructs in the data structure model, with the goal of ultimately using the quantitative measure to refine and further the theoretical constructs of interest in STEM identity development and retention.

I will group individual responses into subgroups using factor analysis and identify whether subgroups have different predictive validities for various outcomes. The two outcomes I anticipate exploring in the scale validation process are STEM identity and self-efficacy. STEM identity is not a static construct and can change over time and across contexts (Carlone & Johnson, 2007), and consequently it is reasonable to consider the possibility it can evolve not only due to individual, self-driven mechanisms but also due to outside influences. It may be that some students are able to reconcile their self-identity with that of a STEM student despite feeling that their demographic characteristics do not fit with their perception of a typical STEM student. The other outcome I plan to explore is self-efficacy, which, similar to STEM identity, also changes over time (Bandura et al., 2001) and also could conceivably be altered by external life events. Students from less prepared backgrounds and with different life experiences could have different levels of self-efficacy than students from other backgrounds.

Ultimately, I propose that certain reported biographical experiences will be predictive of STEM identity and self-efficacy, though I do not make any specific predictions on significant correlations or their directions. Further, as the SCCT framework directly conceptualizes certain life experiences as predicting self-efficacy, and
life experiences as more distal predictors of career interests and goals, I hypothesize that that self-efficacy will predict STEM identity beyond its direct correlation with biodata.

Limitations

Although both the qualitative data analysis in this study and the planned biodata measure development process employ a more theoretical foundation than do the vast majority of studies in these domains, qualitative data analysis is by its nature inexact. In particular, it is possible that both the URAs and I missed salient themes in interview transcripts, though we made every effort to err on the side of more codes over fewer and thoroughly read every transcript multiple times.

Regarding the measure I plan to develop, the field of biodata research has only in around the last decade experienced an increased interest in exploring theory in scale development and validation. The process I describe in this paper is more theoretically sound than many similar studies because it is based on a well-established data structure model, but to my knowledge it has never been applied to biodata scale development. It is possible that other biodata researchers will develop their own systematic scale development process that is different from the technique I plan to use.

Finally, findings from this qualitative data analysis and future scale development and scale validation may not be generalizable beyond STEM students of high ability from under-resourced backgrounds. For this study, these students are the population of interest, but it is possible that the scale, once tested, may not be valid for examining outcomes of STEM students more generally. Future scale validation will determine the limits, if any, on the scale’s predictive validity.
Chapter 6: Conclusion

This study provided insight into the theory behind childhood STEM interest development and formative experiences, which is needed in a field that has generally underexplored these experiences in a systematic way and has often failed to provide a theoretical rationale for data coding and analyses. This lack of theoretical grounding is an even stronger concern in the biodata literature, which often uses rationally generated biodata items and subject matter experts to infer what seems to make sense to include in an inventory. By using interviewees who did not know their experiences would be used to develop a scale, I was able to access qualitative data unbiased by interviewee’s attempts to fit their responses into simple or predetermined categories and explore nuance and context through a data-driven approach to analyzing qualitative data. The scale I plan to develop from the data structure model will have the advantage of being based on interviewees’ life histories, furthering more theory-rich research in the domains of qualitative data analysis and biodata measure development of students’ formative experiences in STEM. The future biodata questionnaire will ultimately be validated on samples of STEM students from various backgrounds to determine which experiences are predictive of STEM identity and self-efficacy, and whether validity differs by students’ backgrounds and abilities. In conclusion, this research project, including its planned future studies, will bring qualitative insights into quantitative measure development and increase the field’s understanding of how and why STEM students succeed and graduate as STEM majors.
References


https://doi.org/10.1002/tea.21465


https://doi.org/10.1002/tea.20237


https://doi.org/10.1177/014662168701100101


Appendix A: Interview Protocol

Formative Experiences

1) What is your earliest memory of being interested in any STEM topic?
   a. Follow-up: Were you encouraged, discouraged, or neither in this interest?

2) What were your experiences in math in elementary school like?
   a. Follow-up: How did you feel about your classes, your teachers, and the topic in general?

3) What were your experiences in science in elementary school like?
   a. Follow-up: For example, how did you feel about your classes, your teachers, and the topic in general?

4) What role, if any, did your family members play in your decision to pursue college and/or a STEM major?
   a. Follow-up if necessary: For example, these could be parents, guardians, grandparents, aunts or uncles, older or younger siblings, or cousins.

5) What role, if any, did your friends or peers play in your decision to pursue college and/or a STEM major?
   a. Follow-up if necessary: For example, these could be friends, teammates, classmates, cousins close to your age, others in religious or language classes.

6) What role, if any, did your teachers or other authority figures play in your decision to pursue college and/or a STEM major?
   a. Follow-up if necessary: For example, these could be a teacher, whether or not they taught you, a principal, a religious leader, or a counselor.

7) During elementary or middle school did you participate in any other STEM activities besides those required by your school?
   a. If yes: Which activity? Which years?

Entering STEM

8) Can you think of any barriers you encountered in entering college, declaring a STEM major, or both?
a. If yes: Can you give me examples of barriers that you encountered?
b. Probe if necessary for more information on what made them barriers.

9) Can you think of what helped you in your goal to enter college as a STEM major?
a. Probe if necessary for more information on how they helped.
b. Probe for more facilitators until no more are offered. Can you give any more examples?

10) Can you think of who helped you in your goal to enter college as a STEM major?
a. Probe if necessary for more information on how they helped.
b. Probe for more facilitators until no more are offered. Can you give any more examples?

11) What influenced you to declare a STEM major?
a. If they struggle to respond, state, "A person, experience, or event, for example."
   b. Probe for more influences.

12) How prepared were you for STEM classes at Rice, based on your high school experiences?

Current Interests

13) What are your biggest STEM interests? These can be as specific as a single subfield or as broad as all STEM disciplines.
a. Follow-up: Have they changed over the course of your life? If so, how?

14) How do you feel about your STEM classes at Rice?
a. Probe if necessary for all aspects not addressed: For example, how much do you enjoy them? How time-consuming are they? How difficult do you find them? How do you feel about your expectations for the grade you will receive in them?

15) How do you feel about your STEM classes at Rice compared to your non-STEM courses?
a. Probe if necessary for all aspects not addressed: For example, are there significant differences in how much you enjoy them? In the classroom experience? In your learning experience?
Future Expectations

16) How likely do you think it is that you will change from the specific STEM major you declared when entering Rice to another STEM major?
   a. Into what major, or potential majors? Why?

17) How likely do you think it is that you will change to a non-STEM major entirely?
   a. Into what major, or potential majors? Why?
Appendix B: Thematic Data Structure Model

First Order Categories
- Response to new math topics
- Mindset for competitive assignments
- Perceptions of utility of math
- Level of creativity in solving math problems
- Independent hobbies in science
- Preferences for science TV shows or books
- In-class science experiments
- Depth of science curriculum
- Personalized encouragement from teacher(s)
- Helping behaviors in class
- Advanced curriculum in STEM
- Enjoyment of STEM classes

Second Order Themes
- Learning/performance orientation toward math
- Level of understanding of math concepts
- Intrinsic interest in science
- Quality of science education
- Interpersonal classroom experiences
- School academic experience in STEM

Theoretical Dimensions
- Math perceptions in elementary school
- Science perceptions in elementary school
- STEM classroom experiences
- School academic experience in STEM
- Interpersonal classroom experiences
- Quality of science education
- Intrinsic interest in science
- Depth of science curriculum
- Independent hobbies in science
- Preferences for science TV shows or books
- In-class science experiments
- Personalized encouragement from teacher(s)
- Helping behaviors in class
- Advanced curriculum in STEM
- Enjoyment of STEM classes
- Level of creativity in solving math problems
- Perceptions of utility of math
- Mindset for competitive assignments
- Response to new math topics