# Does STEM Success Start Young? Exploring Higher Ed Students' Early Academic Experiences in Science and Math at Scale

Brittany C. Bradford
OpenStax, Rice University
Houston, Texas USA
bb25@rice.edu

## ABSTRACT

The United States is experiencing a shortage of STEM workers [24], with many students leaving the pipeline before attaining a career in STEM [18]. STEM education researchers have identified factors at the high school and college level that contribute to attrition, but earlier life events remain underexplored [14]. In this work-in-progress paper, we examine childhood experiences through the lens of qualitative analyses and discuss our ongoing development of an overall understanding of the relevant life and academic "themes" that shape students' lives before entering secondary school. We are currently testing our seven-theme model on the OpenStax Kinetic large-scale research infrastructure using quantitative surveys of participants' biographical data. Our findings from this study will inform future refinement of the survey and its themes, with a particular emphasis on understanding the influence of contextual demographic and psychosocial factors. Over the longer-term, we hope to support research that identifies critical early STEM experiences and offers insight into where certain students might benefit the most from additional STEM support or experiences.

## CCS CONCEPTS

• **Education** • **General and reference** → *Surveys and overviews*
• **Social and professional topics** → *User characteristics*

## KEYWORDS

Research methods at scale; STEM; biographical data; elementary school student experiences; qualitative data

**ACM Reference format:**

Brittany C. Bradford. 2023. Does STEM Success Start Young? Exploring Higher Ed Students' Early Academic Experiences in Science and Math at Scale. In *Proceedings of the Tenth ACM Conference on Learning @ Scale (L@S '23), July 20 1–22, Copenhagen, Denmark.* ACM, New York, NY, USA, *5 pages.* https://doi.org/10.1145/3573051.3596189



*L@S '23*, July 20–22, 2023, Copenhagen, Denmark

ACM ISBN 979-8-4007-0025-5/23/07 $15.00
https://doi.org/10.1145/3573051.3596189

## 1 INTRODUCTION

The United States, like many developed nations, is experiencing a shortfall of STEM workers in dozens of fields [24]. This school-to-career "pipeline" has been described as "leaky," with many more students leaving the pipeline than entering [18]. Students' contextual factors (especially their gender and race/ethnicity) also influence how likely they are to persist in STEM to the point of an eventual career, meaning that certain students are much more likely to leave STEM paths than other students [23].

Researchers have postulated many reasons for this STEM attrition, from lack of skilled science and math high school teachers [17], to exclusionary and overwhelming college STEM classes [22], to a lack of a sense of a greater purpose in pursuing certain STEM paths [1]. All these and other factors undoubtedly play a significant role in causing some students to leave STEM, but research has often overlooked predictors before students are in secondary school.

Many academic theories emphasize the critical role early life experiences, both in the classroom and at home, play in predicting important academic and career outcomes as students progress to high school, higher education, and their eventual careers [13,14]. However, only recently has the field begun to apply these research findings into studying STEM life trajectories (e.g., [7,15]), and to our knowledge, none has done so via exploring specific early life experiences.

In this work-in-progress paper, we discuss the pivotal role of certain general and STEM-specific childhood experiences of students using both qualitative and quantitative research methods, as well as our plans to expand our research efforts to leverage STEM biographical data at scale using the OpenStax Kinetic research infrastructure.

## 2 PRIOR DEVELOPMENT OF THEMES OF STUDENTS' EARLY STEM EXPERIENCES

Over several prior studies ([3,4]), we used qualitative and quantitative research methods to understand college students' early academic and life experiences, both generally and with STEM specifically.

These early experiences are captured with biographical data, or *biodata*, which comprises measures of a person's background and/or prior experiences [19]. Biodata measures provide a way to capture historical experiences and correlate these experiences with a broad range of variables, from behavior [21] to motivation [16] to interest [12].

## 2.1 Qualitative Development of Theme Structure

Our first prior study was a qualitative study that leveraged quotes in one-on-one, in-person interviews with 35 undergraduate STEM students to develop qualitative themes to use in the design of a biodata measure. After thematic qualitative analysis [5], we identified five core themes that students shared in their lives up to middle school: 1) exposure to STEM activities and careers, 2) receiving encouragement for STEM activities and achievement, 3) receiving help in school and on school assignments, 4) participating in hands-on STEM experiences, and 5) participating in math or science competitions.

Example quotes from each of these preliminary themes are shown below.

**Exposure to STEM**

*"My parents… sent me to a summer camp in STEM at Stanford… It was [on] mathematical game theory… That was before seventh grade… [That's]… how I started thinking about maybe [majoring in] math."*

*"My oldest brother… is doing mechanical engineering at Notre Dame… He helped me in my decision… to go into STEM."*

**Encouragement for STEM**

*"When I was younger [my parents] encouraged me to be involved in science clubs and science projects so I just kind of had exposure to that area a lot. … That was the only thing I knew… [My parents] were like, 'Oh, you should do [science].' And I was like, 'Oh, okay, that makes sense.'"*

*"I used to like reading a lot, but my dad was like, 'You should do… engineering,' because engineering is money…. He actually wanted me to do electrical engineering… [My parents] played a pretty big role in [becoming an engineering major]."*

**Receiving Help in School Settings**

*"My mom helped me on [math] homework. She sat me down and told me these are the steps. She taught me how to do long division."*

*"In sixth grade… my math teacher… helped me out when I was struggling… He would definitely step in and offer extra help that I don't think he offered other students."*

**Hands-On STEM Experiences**

*"We would build stuff… in school. We had this rocket building unit in fifth grade and I thought that was interesting…because we shot out little rockets… That was how I was introduced to [STEM], and I got really into chemistry later on."*

*"When I was little [my family] always did a lot of stuff outside… My brother and I enjoyed… catching bugs and having little plastic peanut butter jar terrariums… [We would] keep them for a while and feed mosquitoes to spiders."*

**STEM Competitions**

*"I had really good math experiences. We did a pre-competition type thing in third grade and I always… liked it…. Then in fourth grade we had this CML [math] competition. It was… the students in our class taking tests every six weeks, and the person who takes the best scores at the end… I got a trophy and… the best scores at the end so I was really happy."*

*"I started doing academic UIL competitions in sixth grade. I started competing in the science competition… and I became pretty interested initially in biology just because… at the time that was the most interesting subject to me."*

Working from these core qualitative themes, we then progressed to quantitative measure development. We wrote items for each of these five themes, with each theme comprising 10-14 items. Sample items for each theme are shown below:

- *"How often did you talk to siblings or cousins who were studying STEM in high school or college?"* (STEM Exposure)
- *"To what extent did your parents encourage you to pursue a science, math, or engineering interest?"* (STEM Encouragement)
- *"How likely were you to participate in science competitions that were voluntary?"* (STEM Competitions)
- *"How often did your parents help you with your homework?"* (Receiving Help in School Settings)
- *"How often did you choose to work on hands-on STEM activities when you had the choice to do other activities?"* (Hands-On STEM Experiences)

## 2.2 Pilot Testing Theme Structure

We then pilot tested all survey items on 12 undergraduate students, who evaluated each item's perceived fit into the five themes. The students were asked to evaluate how well each item seemed to fit into the theme we had assigned it, as well as whether it seemed to fit into more than one theme or into no theme.

After reviewing participant feedback about the proposed theme structure, we expanded the framework to six themes and reworded several items to more precisely fall into its assigned theme. See the following figure for the evolution of the theme structure into version 1 of the survey.

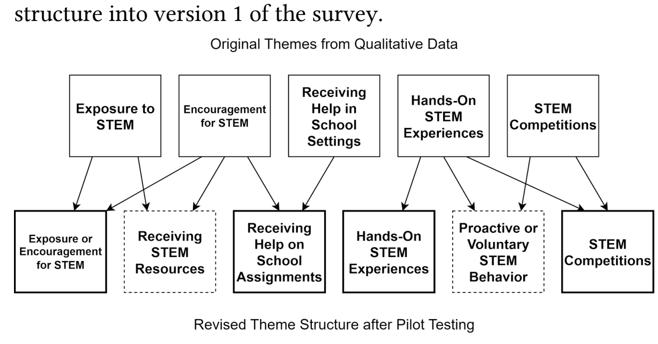


**Figure 1. Development of the second version of STEM biodata themes.**

## 2.3 First Study of Biodata Measure's Theme Structure

Our next study was quantitative and entailed administering the survey to approximately 150 undergraduate students who entered the university as STEM majors. We initially used a confirmatory factor analysis, a statistical approach designed to evaluate the fit of a hypothesized factor structure [6], to test whether our six-theme structure fit our survey responses well. However, we found that our structure was not a good fit for the data, so we instead used exploratory factor analysis, which examines all intercorrelations between all variables to generate solutions that reduce the individual survey responses to a certain number of factors [11]. This approach allowed us to explore different numbers of factors that might be a better fit for our survey responses.

The best fit for our data was a seven-theme model. All themes made intuitive sense once we reviewed the content of the individual items. The changes primarily consisted of distinguishing general early academic support from STEM-specific support from parents; specifying teacher versus parental support; and separating science and math experiences into two different themes. After reorganizing the items, we were satisfied with this restructuring of the data to use in our next study (see Figure 2 below).

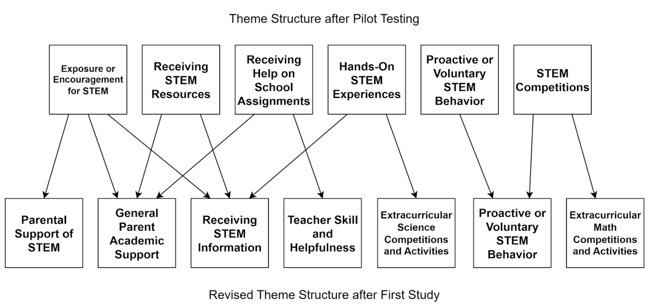


**Figure 2. Development of the third version of STEM biodata themes.**

## 2.4 Second Study of Biodata Measure's Theme Structure

For our second quantitative study, we administered the new version of the survey to a separate sample of approximately 175 freshman and sophomore undergraduate students in any major. We once again conducted a confirmatory factor analysis to determine the fit of our new theme structure to participants' responses.

We found that five of the seven themes fit the data well, but two–Receiving STEM Information and Proactive or Voluntary STEM Behavior–showed significant cross-loadings (meaning that the data did not show a strong distinction between these two underlying themes [2]). We elected to retain the seven-theme structure and rewrote items in these two themes to more clearly distinguish between the themes. We then finalized the fourth version of the measure. See the figure below for an overview of all prior work we have discussed in this section.

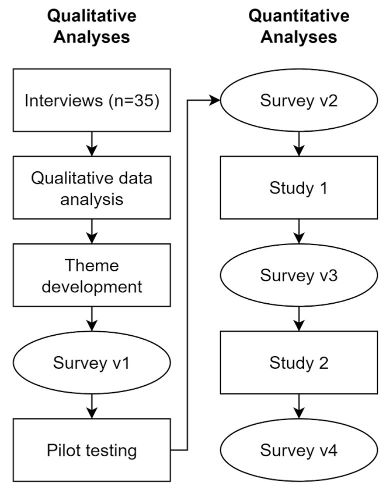


**Figure 3. Overview of our prior STEM biodata research milestones**

## 3 CURRENT RESEARCH AGENDA

We are currently administering the latest version of the survey to participants on the OpenStax Kinetic research platform, which is open to all online adult learners. Our goal is to collect approximately 300 participant responses; we currently have about 110 participants and anticipate that we will have all data collected for this validation study by Fall 2023.

Once we have reached this sample size, we will analyze how well the data fit our survey responses using confirmatory factor analysis. If our survey data meet established statistical criteria for an acceptable fit, we will proceed to other validation analyses, as well as examining the measure's correlations with other measures in our Kinetic library of learner characteristics (see the next subsection for more information).

If the data do not fit the theme structure well, we will continue to look for ways to refine the survey. In particular, if Receiving STEM Information and Proactive/Voluntary STEM Behavior continue to cross-load in this study, we will likely end up combining these themes into a single theme. We will then administer the refined version of the survey (version 5) on a new sample of participants on OpenStax Kinetic. See the figure below for an overview of our current research agenda.

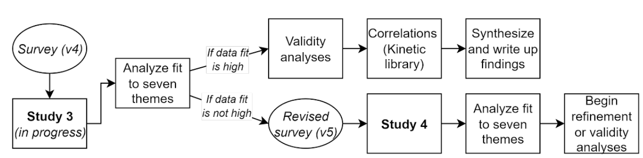


**Figure 4. Overview of current and future stages of our research process**

## 3.1 Leveraging Data at Scale to Explore Learner Contexts

One of the advantages of administering this and other surveys at scale is that large-scale online studies offer researchers access to a far broader sample than typical psychological studies [20], which tend to be based on undergraduate student samples that are not representative of the US population [9]. For example, scaled data collection on Kinetic will allow us to survey participants who did not attend formal post-secondary education, who have graduated and are no longer undergraduates, and/or those who are older than the traditional undergraduate age. Finally, our initial qualitative analyses focused on students with strong demonstrated STEM interests to maximize the relevant content for item generation, but broadening our sample at scale will enable us to study participants with a more diverse range of early STEM and life experiences to understand group differences in biodata themes.

Through the Kinetic research infrastructure, we can also extensively explore the influence of learners' other contexts. We have access to both traditional demographic information (e.g., gender, race/ethnicity, education level), as well as psychosocial and learning characteristics (e.g., vocational interests, STEM interest). For example, certain Holland Vocational Interest, or RIASEC, types [10] might be more likely to score highly on certain biodata themes, such as the Receiving STEM Information theme. By obtaining a larger sample size, we will be able to explore group level differences in learners' contexts with reasonable statistical certainty. See the following figure for characteristics and constructs that might help us understand how learners' contexts impact their early life experiences.

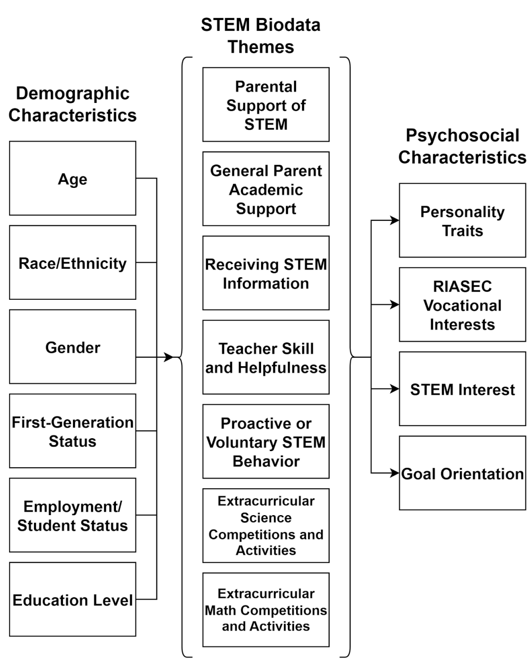


**Figure 5. Planned relationships to explore with OpenStax Kinetic's library of learner characteristics**

# 4 FUTURE DIRECTIONS

Once we have finalized the biodata survey's theme structure by validating it on a diverse range of participants at scale, we will then examine the measure's convergent and discriminant validity [8] against similar STEM measures on another large-scale sample. We hope to further education research by also exploring the measure's predictive validity in identifying influential early life experiences that predict STEM-specific outcomes, including both academic outcomes (e.g., STEM major persistence, STEM GPA) and psychosocial outcomes (e.g., STEM interest, STEM career intentions).

Longer term, we hope our findings can be used to inform the development of experimental studies and interventions to identify, develop, and/or sustain a STEM interest in elementary and secondary students. Further, leveraging a large sample size will enable us to examine group differences in childhood experiences that smaller studies might have overlooked, and how these different biodata themes impact outcomes for diverse learners. Consistent group differences might imply that different interventions might be more successful for certain types of students, and we will continue to explore learners' contexts in all steps of our research agenda.

# 5 CONCLUSION

We have developed and begun validating a biodata survey measure of students' early academic and STEM-specific experiences. Our latest validation study is currently in progress on the large-scale OpenStax Kinetic research infrastructure, and the results from this study will guide our next steps in refining the measure until it can be released to all STEM education researchers. We hope the biodata measure and our findings will ultimately further research that identifies and supports students in STEM, particularly those who are the most likely to leave the STEM pipeline.


# ACKNOWLEDGEMENTS

I would like to thank Dr. Margaret Beier and Dr. Frederick Oswald in the Department of Psychological Sciences at Rice University for their feedback on the aforementioned prior STEM biodata research. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305N210064 to Rice University, NSF Award 1842378, as well as philanthropic foundations such as Schmidt Futures and Valhalla Foundation. The opinions expressed are those of the author and do not represent views of the Institute or the U.S. Department of Education.


# REFERENCES


[1] Jill M. Allen, Gregg A. Muragishi, Jessi L. Smith, Dustin B. Thoman, and Elizabeth R. Brown. 2015. To grab and to hold: Cultivating communal goals to overcome cultural and structural barriers in first-generation college students' science interest. *Transl. Issues Psychol. Sci.* 1, 4 (2015), 331–341. DOI:https://doi.org/10/ghbrgv

[2] Tihomir Asparouhov, Bengt Muthén, and Alexandre J. S. Morin. 2015. Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et al. *J. Manag.* 41, 6 (September 2015), 1561–1577. DOI:https://doi.org/10.1177/0149206315591075

[3] Brittany C. Bradford. 2018. Qualitative data analysis and biodata measure development of Rice undergraduates' STEM formative experiences. Unpublished master's thesis. Rice University, Houston, TX. Retrieved from https://scholarship.rice.edu/handle/1911/105920

[4] Brittany C. Bradford. 2020. Examining STEM formative experiences and college STEM outcomes from a Social Cognitive Career Theory perspective. Unpublished doctoral dissertation. Rice University, Houston, TX. Retrieved April 5, 2023 from https://scholarship.rice.edu/handle/1911/109195

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 2 (2006), 77–101. DOI:https://doi.org/10/fswdcx

[6] T. A. Brown. 2012. Confirmatory Factor Analysis. In *Handbook of Structural Equation Modeling*, Rick H. Hoyle (ed.). Guilford Publications.

[7] Angela Byars-Winston, Yannine Estrada, Christina Howard, Dalelia Davis, and Juan Zalapa. 2010. Influence of social cognitive and ethnic variables on academic goals of underrepresented students in science and engineering: A multiple-groups analysis. *J. Couns. Psychol.* 57, 2 (2010), 205–218. DOI:https://doi.org/10/cj6pct

[8] Donald T Campbell and Donald W Fiske. 1959. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 56, 2 (1959), 81–105. DOI:https://doi.org/10/ds75wg

[9] Paul HP Hanel and Katia C Vione. 2016. Do student samples provide an accurate estimate of the general public? *PloS One* 11, 12 (2016), e0168354. DOI:https://doi.org/10.1371/journal.pone.0168354

[10] John L Holland. 1997. *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Psychological Assessment Resources, Odessa, FL.

[11] Daire Hooper. 2012. Exploratory Factor Analysis. In *Approaches to Quantitative Research – Theory and its Practical Application: A Guide to Dissertation Students*, H. Chen (ed.). Oak Tree Press, Cork, Ireland. Retrieved from https://arrow.tudublin.ie/buschmanbk/8

[12] Olga S. Jarrett. 1999. Science interest and confidence among preservice elementary teachers. *J. Elem. Sci. Educ.* 11, 1 (March 1999), 49–59. DOI:https://doi.org/10/fq53sk

[13] Robert W. Lent and Steven D. Brown. 2006. On conceptualizing and assessing social cognitive constructs in career research: A measurement guide. *J. Career Assess.* 14, 1 (February 2006), 12–35. DOI:https://doi.org/10/d7pdmk

[14] Robert W. Lent, Steven D. Brown, and Gail Hackett. 1994. Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *J. Vocat. Behav.* 45, 1 (August 1994), 79–122. DOI:https://doi.org/10/bpdwrw

[15] Robert W Lent, Matthew J Miller, Paige E Smith, Bevlee A Watford, Kayi Hui, and Robert H Lim. 2015. Social cognitive model of adjustment to engineering majors: Longitudinal test across gender and race/ethnicity. *J. Vocat. Behav.* 86, (2015), 77–85. DOI:https://doi.org/10/gdcggp

[16] M D Mumford, MS Connelly, and TC Clifton. 1990. The meaning of life history measures: Implications for scaling strategies.

[17] National Academies of Sciences, Engineering, and Medicine. 2016. *Barriers and opportunities for 2-year and 4-year STEM degrees: Systemic change to support students' diverse pathways*. National Academies Press, Washington, D.C.

[18] National Academy of Engineering and National Research Council. 2012. *Community colleges in the evolving STEM education landscape: Summary of a summit*. National Academies Press, Washington, DC.

[19] William A. Owens and Lyle F. Schoenfeldt. 1979. Toward a classification of persons. *J. Appl. Psychol.* 64, 5 (1979), 569. DOI:https://doi.org/10/bp5kx3

[20] Katharina Reinecke and Krzysztof Z Gajos. 2015. LabintheWild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, 1364–1378.

[21] Neal Schmitt and J Golubovich. 2013. Biographical information. In *APA Handbook of Testing and Assessment in Psychology*. 437–455.

[22] Elaine Seymour and Nancy M Hewitt. 1997. Talking about leaving: Why undergraduates leave the sciences. *Boulder CO Westview* (1997).

[23] Marissa E. Thompson. 2021. Who's getting pulled in weed-out courses for STEM majors? *Brookings*. Retrieved April 4, 2023 from https://www.brookings.edu/blog/brown-center-chalkboard/2021/07/26/whos-getting-pulled-in-weed-out-courses-for-stem-majors/

[24] Yi Xue and Richard C Larson. 2015. STEM crisis or STEM surplus? Yes and yes. *Mon. Labor Rev. US Bur. Labor Stat.* 2015, (May 2015).