Boston College
Lynch School of Education

Department of
Measurement, Evaluation, Statistics, and Assessment

SCIENCE FASCINATION:
INVESTIGATING CHANGE OVER TIME IN MIDDLE SCHOOL STUDENTS' FASCINATION IN SCIENCE USING A LEARNING ACTIVATION FRAMEWORK

Dissertation
by

RYAN R. AUSTER

submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

May 2022

Abstract

Science Fascination: Investigating change over time in middle school students' fascination in science using a Learning Activation framework

Ryan R. Auster, Author
Michael Russell, Ph.D., Dissertation Chair

This paper describes the construct of fascination in science, a non-cognitive trait combining interest, curiosity, and mastery skills, and the particular relevance of fascination in science for students during middle school. Grounded in the theory of *Science Learning Activation* and employing data from the longitudinal Activated Learning Enables Success study of 2014 (ALES:14), cohorts of sixth and eighth graders were measured on fascination five times over two school years, allowing for an investigation of change over time. Multilevel models were constructed for each grade-level cohort in an effort to determine patterns of change, while also testing for relationships with several student-level characteristics and class-level instructional variables. Results suggest discontinuous patterns of change in fascination, with declining fascination scores in grade 6 boosted over the summer break and declining fascination scores in grade 8 rising the following school year. While the impact of instructional variables was negligible, relationships with several individual covariates were observed, primarily indicating the importance of family support for science. Future research should focus on context-specific elements of in-school activities, along with additional out-of-school factors that may influence fascination.

To Miles and Jack

Acknowledgements

As the final addition to this dissertation, I have thought about writing this page for years, the culminating piece to something I feared might never come to fruition. It is simultaneously the least important and most meaningful section of my work. I am fortunate to have had a tremendous number of individuals propel me over the last decade, and though I can't possibly express my gratitude to you all, if you are reading this and you have been in my life, please know how thankful I am for how you've shaped me.

First and foremost, I would like to thank my wife, Erin, for her continued love and support, and an unwavering belief in my ability to complete this work despite all evidence to the contrary. Not only was she forced to endure years of comments and questions about my continuing pursuit of this degree, but also cancelled weekend plans and delayed vacations, all because "I should be home writing." Without a doubt, there is no one else who would have stuck with me and loved me despite my obvious flaws, and I will remain committed to becoming a better husband and father for the rest of my life.

I also want to thank my parents, not only for their emotional support over the years I have spent working on this (well, over all the years of my life), but most recently for their help caring for my children as I made the final push toward the finish line. There is simply no way I could've found the time or energy to write or revise the last 100 drafts without their adamant presence in my house.

My committee members—Drs. Michael Russell, Laura O'Dwyer, and Katherine McNeill—all deserve special recognition for my success as well. Were it not for their years, and years, and *years* of patience, encouragement, and insistence that I complete what I started, I would have abandoned this long ago. In particular, Dr. Russell's guidance throughout what I must assume has been one of the most infamously long programs of study saved me from remaining an ABD. I would also like to thank former Dean Elizabeth Sparks for her tremendous kindness and understanding when I was struggling to make progress over the years.

Thanks also to my colleagues at the Museum of Science, and in particular, Dr. Christine Reich for her mentorship. She provided time and space for my continued progress on this, helping to make it a priority after I had lost considerable momentum.

To my wonderful friends spread out across the country, thank you. Mac and Joel, your ability to have a conversation with nothing but puns has always made me want to be smarter, and join you as PhDs. Chris, your path to graduation and mine miraculously aligning made the work easier, knowing I always had someone to commiserate with. Alisha and Rookie, though you don't know it, your constant reference to me as "almost-doctor" never failed to motivate me to remove the "almost" part and make good on that title. Dan, John, and Ben, our adventures—past and future—provide much needed invigoration and inspiration, even if we are no longer as young and carefree as we once were.

Finally, to my sons Miles and Jack, thank you for reminding me every day for whom I was writing. When I thought this was for me, I felt directionless; when you arrived, I found intention. You will always be my greatest source of pride.

Contents

List of Figures

List of Tables

Chapter 1: Purpose of the Dissertation Research

Scientific literacy—the ability to use scientific skills, concepts, and reasoning to assist in everyday decision-making—is essential for everyone in the 21st century due to the increased presence of science and technology in our world. However, data suggest that most students in the United States are neither learning critical components of science at a level considered acceptable (e.g., U.S. Department of Education, 2018) nor pursuing science beyond what is required in high school (e.g., National Science Board, 2018). This may seem surprising given the naturally inquisitive nature many children have at a young age as they seek to explore and understand the scientific world around them (Driver, 1985), as science is fundamentally about asking questions to better understand how and why things work. Over time, the impact becomes pronounced and two-fold: adults with less scientific knowledge at their disposal fail to integrate science into everyday decision-making related to both personal and public policy-related matters (von Winterfeldt, 2013), and subsequently are prevented from working in many science-related jobs later in life (e.g., Lacey & Wright, 2009).

Unfortunately, most school-based science experiences do more to repress rather than foster students' opportunities for pursuing self-expression of the curious world around them (Engel, 2011). A primary cause of this is an educational failure to promote a fascination in science, particularly in middle school. A psychological construct closely related to interest, fascination can influence an individual's ability to think about or pay attention to something or motivate them to reengage in the future. As noted in the National Research Council's *A Framework for K-12 Science Education* (2012), "The actual doing of science or engineering can also pique students' curiosity, capture their interest, and motivate their continued study" (p. 42).

While enhancing interest or fascination in science may lead to increased learning and is certainly theorized to cause increased scientific literacy, it also is the spark that ignites into wonder, obsession, and imagination. These words are more than variables in the equation of education; they are at the very heart of what it means for a child to discover the world around him or her. Those with these traits are most often the passionate individuals who make societal advancements. As Albert Einstein said, "Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress giving birth to evolution. It is, strictly speaking, a real factor in scientific research" (as cited in Hadzigeorgiou, 2016, p. 1). Fascination is not just crucial in helping children learn, it is crucial to helping them imagine.

This dissertation will use data from the Activated Learning Enables Success study of 2014 (ALES:14) to examine the patterns of change and factors associated with change in middle school student fascination in science. Although similar to interest and often used interchangeably, fascination adds depth to the desire to learn more, and refers to an individual's emotional and cognitive attachment with topics and tasks. Following two cohorts of middle-school students—one from grade 6 to 7 and another from grade 8 to 9—longitudinal analyses will test for changes in levels of fascination in science and student- and classroom-level factors relating to those changes. Specifically, three-level growth models will be constructed to investigate the relationship between changes in science fascination over time and student-level factors, including gender, racial minority status, resources available at home, and family support for school, as well as classroom-level characteristics including the amount of hands-on activities provided, the balance between student-centric and textbook-based teaching, and technology use in the classroom. The results of this study will allow researchers and teachers to better understand how to support increased fascination in science and minimize declining science fascination in students over time.

Description of the Problem

Science is an inescapable aspect of life in the 21st century. Advanced communication devices keep individuals around the globe in contact at all hours of the day; new vaccines are developed to combat disease more rapidly than ever before; technologies are invented to help grow food and provide clean water for the expanding global population; biologists continue to discover thousands of new species of plants and animals every year. Despite science's apparent relevance, students in the U.S. are choosing not to pursue scientific courses of study in their educational pathways (Osborne, Simon, & Collins, 2003).

One seemingly positive impact of the ubiquity of science is the labor market expansion in areas related to science, technology, engineering, and mathematics

(STEM[1]). The U.S. Bureau of Labor Statistics reports that STEM occupations have experienced above average growth in recent years, with employment increasing by 10.5% from 2009 to 2015, compared to 5.2% growth in non-STEM occupations, a trend that is projected to continue into the next decade (Fayer, Lacey, & Watson, 2017). These same data show that 93% of STEM jobs have wages above the national average, as STEM workers earn a median annual salary of almost $76,000 – more than twice the median wage ($35,080) for all workers (Vilorio, 2014). Having skilled individuals capable of filling these many jobs who will earn more is essential to our economic stability.

It may seem unsurprising that these lucrative STEM occupations often have high educational requirements. In fact, over 99% of these jobs typically require some type of postsecondary education for entry compared to 36% of overall employment; more than 75% require a bachelor's degree or higher (Fayer, Lacey, & Watson, 2017). Unfortunately, American students are not obtaining degrees in science and engineering at the same rate as students in other countries. According to the National Center for Science and Engineering Studies, the United States accounted for only 10% of the global output of bachelor's degrees awarded in 2014 in the areas of science and engineering, compared with 10% from the top eight producing nations in the European Union, 22% from China, and 25% from India

[1] While science, technology, engineering, and mathematics were lumped together in educational policy initiatives as early as the mid-1980s, the term "STEM" wasn't popularized until the mid-2000s (Loewus, 2015). Science is to STEM as a square is to a rectangle: science is always contained in STEM, but not all STEM-related discussions should be assumed to pertain to science. For this reason, I mention STEM as it has become increasingly relevant in discussions of science-related research, but will focus exclusively on fascination in *science* as the topic of this dissertation.

(National Science Board, 2018). Though the U.S. awarded the highest number of doctorate degrees for any individual country in science and engineering in 2014, more than one-third (37%) of these were conferred to temporary visa holders (i.e., international students studying abroad). Currently, the United States is not providing the skilled workforce necessary to fill the increasing number of STEM-related positions available (Linn, Lewis, Tsuchida, & Songer, 2000).

Science education is not strictly for those seeking jobs in STEM careers. Another impact of the ubiquity of science is the critical importance of scientific literacy to all citizens. The success of a democratic society depends on participation by the people who are prepared to deal intelligently with social issues relating to science, such as sustainable energy alternatives, genetically modified foods, and vaccinations (DeBoer, 2000; Kahan, 2013). As the presence of science and technology in our lives persists, there is a need to support citizens in their decision-making about these issues (National Academy of Engineering Committee on Technological Literacy, 2002), and in particular, how they integrate evidence into their reasoning of socio-scientific issues (Kelly, 2007; Zeidler, Sadler, Simmons, & Howes, 2005). Accordingly, the National Research Council (NRC) has listed increasing STEM literacy for all students, including those who will *not* pursue STEM-related careers, as one of its three critical goals, specifically because of the need to understand science to make personal decisions and engage in civic discourse (NRC, 2011).

Developing a scientifically literate citizenry and burgeoning workforce in STEM-related fields means preparing today's students with certain skills and a scientific mindset. As defined by the Organisation for Economic Co-operation and Development (OECD), literacy has several components, including a general understanding of important scientific concepts, frameworks, and methods, as well as the strengths and limitations of science. Furthermore, they emphasize the ability to apply this understanding in situations in which decisions need to be made. "Scientific literacy is the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity" (National Science Board, 2012, pp. 132-33).

Although developing these skills and mindsets is extremely important for school-aged youth, considering scientific literacy a life-skill is practical because no one can know in advance all that they need to know to make decisions about new complex issues that continue to arise. As Jon Miller (2010) notes:

> *How many adults can claim that they studied stem cells or nanotechnology when they were students? In the decades ahead, the number and nature of new scientific issues reaching the public policy agenda will not be limited to subjects that might have been studied in school but will reflect the dynamic of modern science and technology*. (n.p.)

Resultantly, the ability to learn about new scientific issues and develop informed opinions about their applicability to one's life is (or must become) a critical outcome of science education. Whether considering the task of producing the next generation of the science and engineering workforce or creating a science-literate and engaged citizenry, supporting science education at the national level continues to be a priority for school systems and policymakers alike (U.S Department of Education, 2016).

Unfortunately, recent studies suggest that science achievement among school-aged youth in the United States is low relative to students of similar ages in other countries, and low relative to national proficiency standards. Results from the Trends in International Mathematics and Science Study (TIMSS) 2019 show a gap in science achievement between U.S. students and those in other countries. At the fourth grade, seven education systems outperformed the U.S., while in the eighth grade, this number increases to ten. Perhaps more strikingly, only 15% of American 4^{th} and 8^{th} graders performed at or above the *Advanced* international science benchmarks at their respective grade levels; more than 20% of 4^{th}-grade students fail to perform at even the *Intermediate* benchmark, and this percentage increases to 30% by the 8^{th} grade (Mullis, Martin, Foy, Kelly, & Fishbein, 2020). National Assessment of Educational Progress (NAEP) results in the United States also suggest mediocre achievement in science. While the percentage of students in the eighth-grade at or above the *Proficient* level increased from 32% in 2011 to 34% in 2015, only 25% of 12^{th}-grade students met or exceeded the *Proficient* benchmark in 2015 (U.S. Department of Education, 2018). The pattern seen in both studies suggests that not only are U.S. students not performing well relative to their peers in other countries and national proficiency standards, but also that performance declines as students progress through required schooling.

While analyzing patterns of student achievement is important, it does not capture the complete picture of science learning experiences for elementary-aged children, including their inspiration and preparation to study science. Indeed, attitudes toward science have consistently been shown to be relevant to advancing achievement in science, whether considering feelings toward science (e.g., Mullis et al., 2020), self-efficacy (e.g., Bandura, 1977; Zimmerman, 2000), or interest (e.g., Harackiewicz, Barron, Tauer, & Elliot 2002). Frequently, such motivational constructs are lumped together into one and referred to as the affective side of learning science (Potvin & Hasni, 2014).

Complicating the issue of sub-par academic achievement in science has been a demonstrated decline in learners' desire and ambition to pursue science when they have the opportunity to choose their course of study (Osborne, Simon, & Collins, 2003). It is widely acknowledged that most students hold positive views toward science in the early grades (Mantizcopoulos, Patrick, & Samarapungavan, 2008; Mantizcopoulos, Samarapungavan, & Patrick, 2009), but as they progress through elementary and middle school, many students become disenfranchised and lose interest over time (Gottfried, Fleming, & Gottfried, 2001; Osborne, Simon, & Collins, 2003; Simpson & Oliver, 1990). "As a result, too many American students conclude early in their education that STEM subjects are boring, too difficult, or unwelcoming, leaving them ill-prepared to meet the challenges that will face their generation, their country, and the world" (President's Council of Advisors on Science and Technology (PCAST), 2010, p. viii). In fact, what seems to be occurring is that these declining attitudes correspond with a decrease in science fascination, preventing students from choosing future science experiences and thereby negatively influencing science learning. This is particularly troublesome, as STEM experiences during early adolescence between the ages of 10 and 14 have been shown to be key in predicting involvement in additional science education and science career choice (Bonnette, Crowley, & Schunn, 2019; Maltese & Tai, 2011; Tai & Maltese, 2009).

Learning Activation

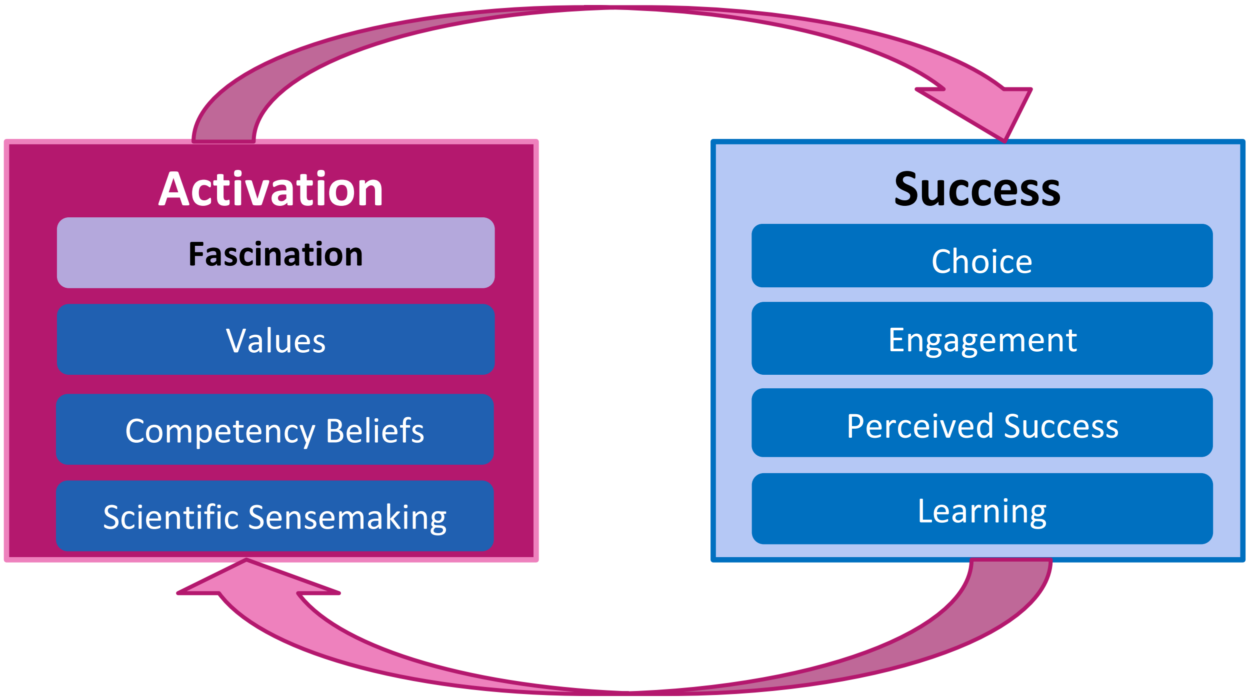
This work is grounded in prior research using the theory of *Science Learning Activation*: "a state composed of dispositions, skills, and knowledge that enables success in proximal science learning experiences" (Dorph, Shields, Tiffany-Morales, Hartry, & McCaffrey, 2011). Science activation comprises four critical dimensions:

- *Competency beliefs* – the extent to which a person believes that they are good at science;
- *Fascination* with natural and physical phenomena – a person's emotional and cognitive attachment with science topics and task;
- *Scientific sensemaking* – the degree to which a person engages with science learning as a sensemaking activity; sub-dimensions include: questions, experiment, evidence, explanation, and nature of science; and

- *Values* science – the degree to which a person values science, including the knowledge learned in science, the ways of reasoning used in science, and the role that science plays in families and communities.

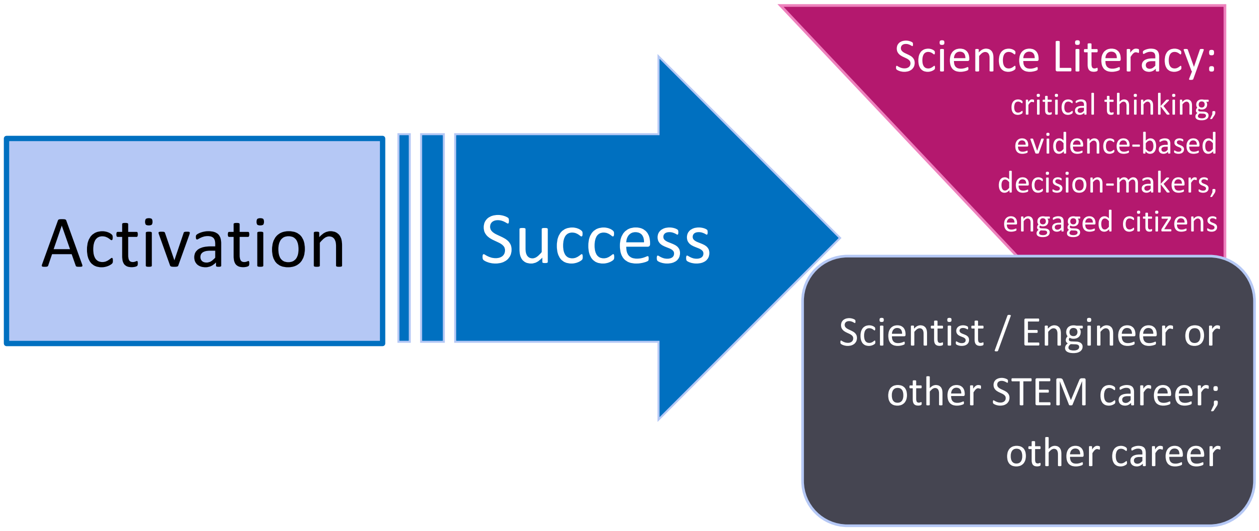
Activating a student in terms of his or her science learning will position the individual to be successful by encouraging that person to choose to participate in science learning opportunities (choice); be positively engaged, both affectively and cognitively, in such opportunities (engagement); believe himself or herself to be successful throughout the experience (perceived success); and meet the intended learning goals (learning). The four dimensions of activation then enable success, which leads to additional opportunities for science activation and further success in an iterative fashion. Figure 1.1 below shows the theorized cyclical relationship between Science Learning Activation and the outcomes of success (Dorph, 2016).

Figure 1.1. Science Learning Activation



While these successes pertain to proximal outcomes, the theory posits that they in turn can lead to distal successes in the form of STEM-related career choices such as becoming a scientist or engineer, along with enhanced lifetime science literacy. Figure 1.2 below emphasizes the intended long-term outcomes of Science Learning Activation: as repeated science activation leads to success over time, aspects of science literacy are developed, and pathways to STEM careers are illuminated (Dorph, 2016).

Figure 1.2. Distal outcomes of Science Learning Activation



Purpose of the Study

Visible in Figure 1.1 above, *fascination* is the focal dimension in the proposed dissertation. According to the Oxford Dictionary, fascination (n.d.) takes on an autological definition:

1. The power to fascinate someone; the quality of being fascinating.
2. The state of being fascinated.

The suggested section on usage then goes on to link the definition of fascination with interest directly: "A person has a fascination with something they are very interested in, whereas something interesting holds fascination for a person." While interest is an educational construct with considerable research behind it, fascination is not, likely due to its multidimensional nature:

> *This dimension includes aspects of what many researchers have referred to as curiosity, interest in science both in and out of school, and mastery goals for science content. It also includes affective elements such as emotions related to science, scientific inquiry, and knowledge. Cited research to date in each of these areas suggests that each of these constructs may be compelling motivators to choice towards, engagement during, and attainment in science learning. Therefore, Fascination should be an important driver towards these aspects of success.*
>
> (Chung, Cannady, Schunn, Dorph, & Bathgate, 2016b, p. 1)

Fascination includes aspects of curiosity and wonderment, positive affect, and obsession. While it is closely related to interest, this dissertation will use fascination in science as a unidimensional trait (the technical details of which will be discussed in Chapter 3) while acknowledging significant overlap with the psychological construct of interest. Both terms (interest and fascination) will be discussed in more detail in Chapter 2.

Despite increased focus in recent years on motivational pillars in education and the acknowledgement that advancing student interest matters for persistence in science, research has yet to identify longitudinal patterns of change in science fascination and the factors that influence these patterns for middle school students. The proposed dissertation seeks to identify both patterns of change in student science fascination and factors associated with this change. Using the Activated Learning Enables Success study of 2014 (ALES:14), fascination scale scores were measured five times over two academic years from more than 3,700 middle school students, along with other science activation measures, background characteristics of the students, and relevant instructional characteristics of the classrooms. The resultant longitudinal data set enables several multilevel models to be constructed: one for the cohort of $6^{th}/7^{th}$ graders, and one for the cohort of $8^{th}/9^{th}$ graders. Changes in science fascination over time will be modeled to examine three topics:

- age cohort differences in change over time,
- non-linear patterns in change over time, and
- factors associated with change in science fascination over time.

Research Questions

This dissertation seeks to extend previous research done as part of the Activated Learning Enables Success study of 2014 by explicitly investigating changes in student fascination in science over time. Specifically, the current study will address the following research questions:

RQ1: What is the average change in science fascination of middle school students over time? And how much does this change vary, on average?

RQ2: To what extent is predicted student science fascination associated with the instructional characteristics of the classroom, including student-centric teaching, hands-on methods, and classroom technology use? And are these effects sustained over time?

RQ3: To what extent are middle schools equally successful in sustaining science fascination for students whose gender, race, family support for science, or economic backgrounds differ?

Each of the above questions will be investigated separately for students progressing from grade 6 to 7 and grade 8 to 9 to determine if there are age-related differences.

Importance of the Study

The first research question investigating average change in science fascination of middle school students over time is particularly important as a growing body of work—admittedly focused on science interest, rather than fascination—demonstrates that students lose interest in school-based science, right around the time they enter high school (Barmby, Kind, & Jones, 2008; Christidou, 2011; Krapp & Prenzel, 2011). However, it is not universally true that all students become disinterested, nor have these studies produced consistent findings in terms of the corollaries of disinterest (Potvin & Hasni, 2014). Moreover, the majority of these studies looking at changes in science interest utilize only two time points. The ALES:14 data set, on the other hand, offers five measures of science fascination over two years, enabling statistical methods to test not only for change over time, but non-linear patterns of growth as well.

The second research question investigating classroom instructional aspects is also critical in the larger discussion of changing student fascination in science, particularly from a practical aspect. "Focusing on the potential for situational interest inherent in the material and mode of presentation may help teachers promote learning for all students regardless of their idiosyncratic interests" (Hidi & Harackiewicz, 2000, p. 157). Again, recognizing that interest is closely related to fascination, and acknowledging that the ALES:14 study lacks data on the science material taught, this quote illustrates the influence instructional activities have on changing student interest. An understanding of exactly how these modes of classroom presentation relate to patterns of change may help teachers support growth in science fascination with their students.

The third research question investigating student-level covariates influencing growth in science fascination is relevant and necessary to further identify the personal characteristics and demographic influences on science fascination. Research continues to show that gender (e.g., Baram-Tsabari & Yarden, 2011) and racial/ethnic background (e.g., Catsambis, 1995) are related to science attitudes, achievement, and career interests. Parental involvement has also been shown to indirectly affect science attitudes, as parents typically have control over out-of-school science activities such as museum visits and science camps (e.g., George & Kaplan, 1998). Therefore, in order to adequately model changes in student science fascination, these variables must be included so that the relationships can be taken into account and better understood.

Successful completion of this work will allow researchers, educators, and policymakers to better understand how changes in student science fascination occur over the middle school years, if this change differs between early middle schoolers and beginning high schoolers, and when, if at all, opportunities exist to capitalize on existing fascination to help students expand this fascination in science even further. This research design allows us to track changes in fascination within individuals over the course of multiple school years and model the observed changes using relevant predictive variables. Further, this analysis can explore the variation in changes of fascination scores with non-time varying variables that are traditionally associated with science achievement, like gender, socio-economic status-as measured by resources availability, and racial/ethnic identity. Overall, these models will provide an estimate of the mean growth for fascination and variation around that growth, an assessment of the reliability of the variables in the model used to predict changes in fascination, and an estimate of the correlation between the initial value on fascination and the growth observed over time (Raudenbush & Bryk, 2002).

Chapter 2: Review of the Literature

The focus on the importance of science education in the United States is more than a century old. Beginning with the mechanization of farming at the turn of the 20th century, the nation was compelled to learn a new set of skills in an effort to embrace the technological revolution, and scientific literacy became an important issue for all citizens (Gatewood & Obourn, 1963). However, the launching of the Russian space satellite Sputnik in 1957 likely did more to spark science education reform in the U.S. than any other event in history, leading to the creation of the National Defense Education Act of 1958 (Harris & Miller, 2005). Described as an "educational emergency bill" by Congress, it led directly to the formation of the National Aeronautics and Space Administration (NASA) that same year, an organization that continues to play a role in pushing science and technology in America today. A quarter of a century later, The National Commission on Excellence in Education released the infamous report, "A Nation at Risk: The Imperative for Educational Reform." Specifically calling out science and technology, it cautioned that the failure to support education in the transition from the industrial age to the information age was akin to an act of war, the results of which could be disastrous for national security and the U.S. economy (Gardner, 1983; Harris & Miller, 2005). What followed was a series of policies by various White House administrations, each attempting to establish education (and science in particular) as a national priority, and each with extremely high expectations of students, teachers, and schools. These policies shifted their focus away from educational inputs, such as per-pupil spending on materials, to educational outcomes, such as proficiency scores on standardized assessments (U.S. Department of Education, 2003).

Most recently, science education reform in the United States has focused on improving and aligning standards-based curricula across states, leading to the development and implementation of the Next Generation Science Standards (NGSS), following the adoption of common core standards by many states in other curricular areas such as mathematics and English & language arts. As summarized by the National Research Council's (2012) *A Framework for K-12 Science Education*:

> *The overarching goal of our framework for K-12 science education is to ensure that by the end of 12th grade, all students have some appreciation of the beauty and wonder of science; possess sufficient knowledge of science and engineering to engage in public discussions on related issues; are careful consumers of scientific and technological information related to their everyday lives; are able to continue to learn about science outside school; and have the skills to enter careers of their choice, including (but not limited to) careers in science, engineering, and technology.* (p. 1)

At face value, the objectives laid out in the design of the NGSS framework are complimentary to the idea of scientific literacy articulated in the previous chapter, with several poetic and well-intended goals noted above. However, the NGSS has come under criticism by some who argue that, contrary to claims made within the standards document, there is still too much emphasis on passive learning (i.e., receiving scientific facts) rather than knowledge construction (i.e., doing/exploring science); in other words, there is a noticeable lack of “epistemic agency” (Miller, Manz, Russ, Stroupe, & Berland, 2018). One content analysis found that participatory practices accounted for less than 7% of the NGSS standards, although participatory knowledge (non-participatory practice) made up an additional 35%, leaving close to 60% of these new standards categorized as non-participatory in both knowledge and practice (Hoeg & Bencze, 2016).

The NGSS have been commended for emphasizing issues of equity and diversity, including the visibility placed on such issues by devoting an entire chapter of the NRC Framework, titled “Equity and Diversity in Science and Engineering Education” (NRC, 2012). Despite this, an essential question has been raised with regard to standards and their impact on student achievement: what good have they done? If, in fact, standards-based reform was working, not only would science achievement be on the rise, but achievement gaps between ethnic groups would theoretically be shrinking—neither of which has been easily observed. Complicating this matter tremendously is the continued emphasis placed on high-stakes assessment and the recommendation to implement a new testing system to measure student performance within the NGSS, despite the recognition that such assessments are burdensome for both teachers and students, detract from valuable instructional time, and perpetuate achievement gaps (Rodriguez, 2015).

Research on science achievement is abundant, and many studies show persistent gaps between students associated with demographic factors such as race, English language proficiency, and socioeconomic levels (Morgan, Farkas, Hillemeier,

& Maczuga, 2016). Low levels of science achievement among school-age children leads to an adult populace less able to understand critical public policy issues requiring greater scientific literacy and reasoning (e.g., climate change), as well as lower employment opportunities and prosperity (National Academy of Sciences, National Academy of Engineering, & Institute of Medicine [NASNAEIM], 2010; 2011). However, why these science achievement gaps occur in the first place is poorly understood, and most studies have analyzed achievement at a particular time point rather than longitudinally (e.g., Liu & Whitford, 2011). Byrnes and Miller (2007) identified only 12 longitudinal studies of science achievement published since 1992, and of these, the maximum number of predictors used was eight, with an average of five. Thus, “there is no way to tell the difference between important, authentic predictors and relatively minor or even spurious predictors” of science achievement among students in the United States (p. 600).

Though academic achievement in science is certainly important to the discussion of science education, the relationship between the affective side of learning science (e.g., fascination) and academic achievement is messy. In their 2014 review, Potvin and Hasni summarize a number of studies finding a negative relationship between interest, motivations, and attitudes towards science and technology and school-based science performance. Among them, Osborne & Dillon note that the “higher the average student achievement, the less positive is their attitude towards science” (as cited in Potvin & Hasni, 2014, p. 86). Complexly, the most recent TIMSS data support the opposite, a positive relationship between feelings toward science and average science achievement. Specifically, students reporting they "like [science] very much" had higher average science scores than their grade-level peers reporting they "like somewhat" or "do not like" science, a pattern consistent at both grades 4 and 8. Consistent with other studies, positive feelings toward science was found to decrease with age, with 52% of 4th grade students reporting they "like [science] very much" dwindling to 35% by 8th grade (Mullis et al., 2020).

In line with the theory of Science Learning Activation—that success in science (through choice, engagement, and perceived success) is enabled by activating a student's science learning (through fascination, values, competency beliefs, and sensemaking)—the literature review that follows focuses on the affective side of student science learning, rather than academic achievement. Moreover, this work seeks to isolate the dimension of *fascination* as a construct of particular importance. The Science Learning Activation Lab describes fascination as being composed of "curiosity, interest in science both in and out of school, and master goals for science content" (Chung et al., 2016b, p. 1). Thus, in an effort to make sense of the breadth of prior work in these areas, three prominent sections are presented: mastery goals, curiosity, and interest. There is more literature to review in the area of interest, notably various theories of interest development; therefore, I spend considerably more time in that section relative to others. As will be discussed in Chapter 3 when alluding to the specifications of the fascination scale, however, all three of these subdimensions are given relatively equal weight when measuring students' fascination in science. (See Appendix A1.)

Mastery Goals

Sometimes referred to as task-involvement goals or learning goals, the construct of mastery goals places attention on the intrinsic value of learning. Students who are mastery goal-oriented focus on understanding content as the motivation to learn, with a sense of self-efficacy driven by the notion that continued effort will lead to success or content mastery (Ames & Archer, 1988). This stands in contrast with the construct of performance goals (also referred to as ego-involvement goals) which places attention extrinsically on achievement or recognition that others can see, such as the favorable judgement of individual competence based on academic performance. Distinguishing mastery and performance goal behavior further is self-conceptualization: individuals who pursue mastery goals tend to conceive of intelligence as malleable and work to develop that quality, whereas those who pursue performance goals often conceive of intelligence as fixed and work to document their successes (Dweck & Leggett, 1988). Research linking mastery and performance goal behavior suggests that a mastery goal approach is likely to elicit a motivational pattern that will maintain achievement behavior, whereas a performance goal approach is likely to elicit a motivational pattern that will seek to avoid failure (Ames, 1992).

Mastery goals foster orientation toward the development of new skills, an intense understanding of content, the improvement of competence, and as the name implies, achieving a sense of mastery. All of this is reinforced in a demonstration of the willingness to engage in the process of learning: mastery goals have been shown to increase the amount of time spent on learning tasks (e.g., Butler, 1987), encourage persistence when challenged (e.g., Elliott & Dweck, 1988), and enhance the quality of engagement in the learning process (Ames, 1992). Pursuit of mastery goals, then, can be seen as a sincere rather than superficial approach to learning, with students better able to bridge gaps in knowledge over time as they are less consumed with the possibility of failure (Bonnette, Crowley & Schunn, 2019).

Curiosity

As a basic human instinct, curiosity has been discussed and debated for centuries, beginning with the ancient Greeks who treated it as a virtue to be nurtured, and later during the Middle Ages when curiosity was indicted as a vice (Lowenstein, 1994). As early as the mid-18th century, philosophers like Edmund Burke (1757) were noting the captive nature curiosity has over us:

> *We see children perpetually running from place to place, to hunt out something new: they catch with great eagerness, and with very little choice, at whatever comes before them; their attention is engaged by everything, because everything has, in that stage of life, the charm or novelty to recommend it. But as those things, which engage us merely by their novelty, cannot attach us for any length of time, curiosity is the most superficial of all affections; it changes its object perpetually.* (n.p.)

In the discussion of education, these sentiments still linger over 100 years later. “Obviously, a student who is curious about something is interested in it, but curiosity is more than just interest.” (Gardner, 1987).

Broadly defined as a desire for acquiring new knowledge and new sensory experience, curiosity can be separated into two categories, labeled perceptual and epistemic. Whereas perceptual curiosity leads to an increased perception of stimuli, epistemic curiosity is characterized as a “drive to know” (Berlyne, 1954, as cited in Litman & Spielberger, 2003, p. 75). Both perceptual and epistemic curiosity can be seen in a child’s inquisitive nature about the scientific world around them, seeking at first new sights, sounds, and smells, and returning to these natural wonders repeatedly until a firm understanding is established. Curiosity, and its satisfaction, may help transition students from early exposure in science to sustained, well-developed interest (Bonnette, Crowley & Schunn, 2019).

Interest

In the early 20th century, John Dewey addressed the American Association for the Advancement of Science and noted “students have not flocked to the study of science in the numbers predicted” (1910, p. 122), citing that teaching as an accumulation of knowledge rather than a method of thinking or attitude of mind was to blame. Several years later, Dewey would assert that interest is the driving factor behind most learning behaviors (1913). Although he was not the first to suggest this, as Johan Friedrich Herbart theorized interest to be both a desirable condition of learning as well as an important educational outcome, Dewey certainly popularized the notion of interest as a critical condition of learning (Krapp & Prenzel, 2011; Ryan & Deci, 2000; Wigfield & Cambria, 2010). Functioning through a "catch and hold" mechanism, Dewey (2013) believed interest would first seize an individual's attention through cognitive stimulation (the catch) and then be maintained through finding deeper meaning (the hold). He described interest as active, objective, personal, emotional, and dynamic, because together these components create an individual's form of self-expression. Importantly, interest could not be imposed on an individual but must be fostered through genuine learning activities and then capitalized on, resulting in the pursuit of a valuable, worthwhile activity (Covington, 2000a, 2000b).

In this sense, interest may be viewed as a driving force in successful learning and achievement, but that may depend on the definition of interest that is applied. As defined by the Oxford English Dictionary, interest (n.d.) can mean any of the following:

1. The feeling of wanting to know or learn about something or someone.
2. The quality of exciting curiosity or holding attention.
3. An activity or subject which one enjoys doing or studying.

In looking at these definitions, the cognitive and affective components are readily identifiable, but determining the role of each in stimulating learning or prolonged engagement with an activity, task, or object can pose a complex problem. As well, the neurological component, denoted "seeking" by Panksepp (2005), has been introduced as another important factor in determining how interest develops and is maintained (Hidi, 2006). Neuroscientific evidence suggests that this seeking system is designed to actively engage the world and help integrate associated information about the environment through the emergence of cognitive maps, expectancies, and habit structures in order to increase the efficiency of behaviors (Vartuli, 2017).

Theories of Interest

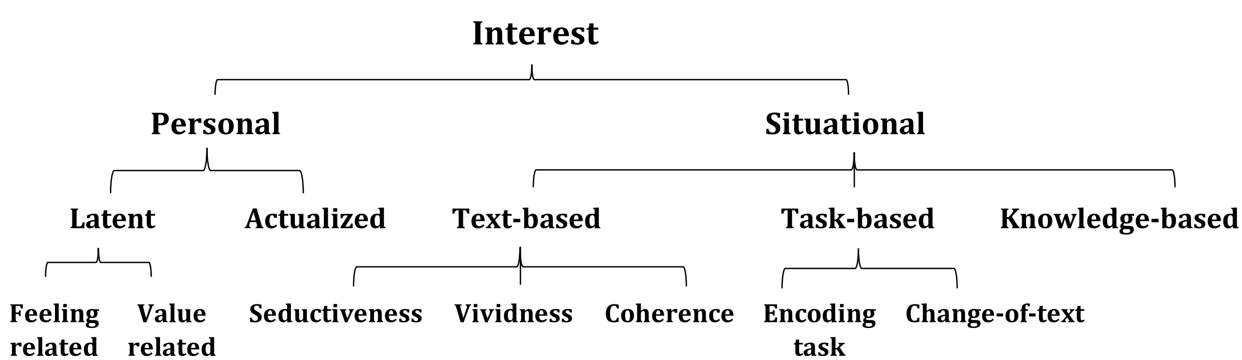
"Genuine interest is the accompaniment of the identification, through action, of the self with some object or idea, because of the necessity of that object or idea for the maintenance of a self-initiated activity" (Dewey, 1913, p. 14). Since the turn of the 20th century, various theories of interest in explaining attentional phenomena have been suggested by William James (1890), James Baldwin (1911), and Edward Thorndike (1935), in addition to John Dewey. However, it was only in the late 1970s when psychologists began to reiterate the central role interest has in relationship to an individual's acquisition of values and knowledge (e.g., Eckblad, 1981; Izard, 1979; Langsdorf, Izard, Rayias, & Hembree, 1983). An understanding of various theories and how they developed will help focus the work proposed in this dissertation, with specific attention given to the person-object-interest theory (POI) and the Four-Phase model of interest development.

Proposed by Ryan and Deci (2000), self-determination theory posits that two different types of motivation, intrinsic and extrinsic, compel individuals to act. Extrinsic motivation refers to taking action because of the intended outcome or results, whereas intrinsic motivation is the action caused by an individual's interest or enjoyment. This notion of intrinsic motivation is closely aligned with drive theory

(Hull, 1943), acknowledging that individuals are "driven" by a primary motivation, itself a composite of physiological or neurological needs and the associated behaviors which seeks to satisfy them.

Schraw & Lehman (2001) conducted a review of prior research, focusing on situational interest and distinguishing it from personal interest. While acknowledging the historical context that led to more current conceptualizations of interest, they breakdown situational interest into three main categories: text-based, task-based, and knowledge-based (see Figure 2.1). They also further distinguish between latent and actualized personal interest, in which latent interest refers to long-term orientation toward a particular topic and is assumed to be intrinsic, whereas actualized interest is a topic-specific motivational state and dictates how an individual engages in an activity (Schraw & Lehman, 2001).

Figure 2.1. Personal and situational interest



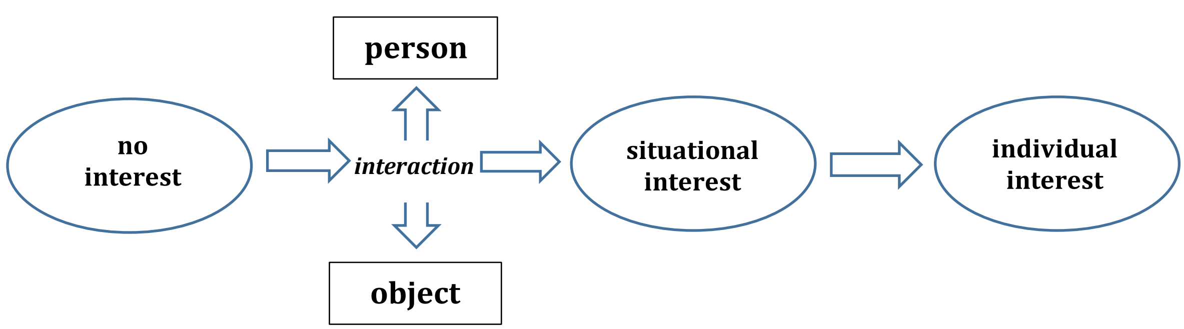
Eccles and Wigfield (2002) also conducted a review of research on motivation, beliefs, values, and goals focusing on developmental and educational psychology. Theories were arranged into four main categories—those focused on expectancy, those focused on the reasons for engagement, those integrating expectancy and value constructs, and those integrating motivation and cognition. It was noted that similarities and differences between categories stemming from the various intellectual traditions from which the theories originated created difficulty in placement. Despite this, POI, Hidi and Renninger's Four-Phase Model, and self-determination theory, were categorized as theories focused on reasons for engagement.

This all leads to the two dominant models of interest development: person-object-interest theory (POI) and Hidi and Renninger's Four-Phase interest model. Both models acknowledge interest as a unique motivational variable that is cognitive and affective in nature (e.g., Hidi & Renninger, 2006; Schiefele, Krapp, Prenzel, Heiland, & Kasten, 1983), and both believe that interest, unlike motivation, is content-specific and exists as a relationship between a person and something else such as facts, things, or domains—science, for example. Both also assert that interest operates through a dual process of situational and individual interest, as discussed earlier, and focus on learning and the role of interest in education.

Person-Object-Interest Theory (POI). According to Schiefele et al. (1983), prior to the early 1980s, little work had been done studying individuals' objective engagement with objects, despite earlier recognition that these relationships had considerable influence on personality development. The goal was to develop a useful theory of interest within a pedagogical framework so that it could be applied to studies of education. Originally called the educational theory of interest (Schiefele et al., 1983), it was developed using an action-theory framework rather than behavioral one. This stemmed from the fact while behavioral theories focus on conditional responses to stimuli, action theories rely upon comprehension of the situation and a choice between alternatives; they value the decision to become involved with the particular object based upon a given value structure; and they affect the emotional quality of the experience (Schiefele et al., 1983). An individual moves from a state of "minimal interest" (i.e., situational interest) to "ideal interest" (i.e., individual/personal interest) as he or she reengages with the object over time, developing a higher level of cognitive complexity that works in conjunction with the emotional attachment that has been created and the value orientation that has been placed on the relationship (Vartuli, 2017).

Figure 2.2 below depicts the three-phase model of POI. This model begins with an individual who has an initial experience and develops preliminary situational interest (the catch), through repeated interactions progresses to a more stabilized situational interest (the hold), followed by relatively enduring individual interest (Krapp, 2002).

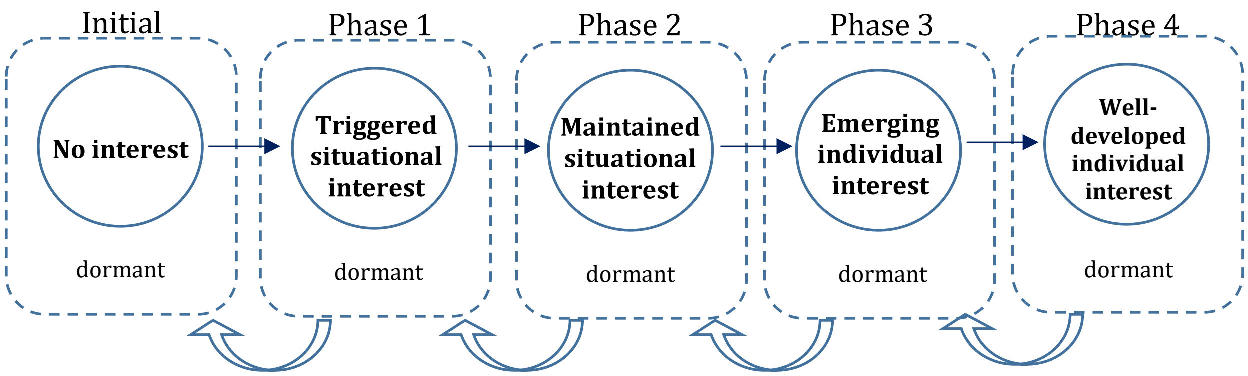
Figure 2.2. Phases of development, person-object theory of interest



Four-Phase Model of Interest Development. Proposed by Suzanne Hidi and Ann K. Renninger, the Four-Phase model builds directly off the three-phase model described in POI (see Figure 2.3). Here, early situational interest is triggered (the catch) by instructional conditions or learning environments, which progresses to maintained situational interest, characterized by persistence, focused attention, and reoccurrence to engage (Hidi & Renninger, 2006). Over time, emerging individual interest is fostered through positive feelings and stored knowledge and value, followed by maintained individual interest (the hold). For Hidi and Renninger, there are more distinct phases of development between the catch and the hold, as the last phase enables a person to sustain long-term endeavors, generates more and deeper levels of strategies for work, and persists despite frustration (Renninger & Hidi, 2002).

The Four-Phase model suggests that situational interest is largely temporary and affected almost exclusively by extrinsic factors, whereas individual interest is considered enduring and more intrinsic in nature. The authors believe that the progression of phases is sequential, although if unsupported, interest in any phase can go dormant or regress (2006).

Figure 2.3. Phases of development, Four-Phase model



Situational interest has been shown to positively influence cognitive performance, focus attention, enable integration of information with prior knowledge, and enhance learning. Individual interest has a positive effect on attention, recognition, and recall; academic motivation; persistence and effort; and levels of learning (Hidi & Renninger, 2006; Schiefele et al., 1983; Schiefele, Krapp, & Winteler, 1988). Both POI and the Four-Phase model stress the importance of cognition, affect, and value orientation in creating individual interest. Whereas POI divides interest into feeling- and value-related valences, the Four-Phase model posits that affect and knowledge work in concert to create value (Hidi & Renninger, 2006; Krapp, 2005). In consideration of student learning, creating environments that enhance situational interest which can then afford more enduring individual interest is necessary for an individual to persist in learning.

Many researchers believe that interest is associated with an internal awareness of beliefs, which supports the overwhelming use of self-reporting (i.e., through surveys and interviews) as a primary data collection method (Renninger, Nieswandt, & Hidi, 2015). Indeed, student interest in science has primarily been assessed this way, through Likert-style scale questions and open-ended responses. Teacher interviews and classroom observations are occasionally included for the triangulation of data. Other factors that have been theorized to characterize interest such as attitude, engagement, and persistence are also included in measurement (Hidi et al., 2004).

In their review of 12 years of research on interest and motivation in education and attitudes toward science and technology, Potvin and Hasni (2014) found that interest was most often operationally defined as an association with an "object of interest" associated with a specific domain preference. Following their review of the literature, Hasni and Potvin then created and validated a questionnaire that takes into account 18 components, including general interest in school science and technology ("school-S&T"), the utility of school-S&T, teaching methods preference, and perceived importance and preference for school-S&T with respect to other subjects (2015). Following administration of their newly validated instrument with more than 1,800 students in grades 5 through 11, they found that while general interest in school-S&T is high, few perceive utility in school-S&T for everyday life or intend to pursue S&T-related studies or careers. They did note significant grade level differences but not gender-related differences. Based on these findings, they recommend that schools intervene to promote S&T interest development, particularly at early grade levels to mitigate perpetuating gaps following elementary school, as well as promoting cultural activities relating to S&T (Hasni & Potvin, 2015).

One of the largest studies of student interest in science, the Programme for International Student Assessment (PISA) administered a domain-specific test focusing on science in 2006. With data from more than 400,000 students across 57 countries, the student achievement test also featured embedded interest items, defined by PISA as a willingness to engage in science-related issues and to reflect on scientific issues (OECD, 2007). This was significant, as Likert-response scale questions assessed students' levels of interest in subjects tied to specific contexts, enabling researchers to investigate students' specialized areas of scientific interest (Dreschel, Carstensen, & Prenzel, 2011). The data from this study have been used in follow-up studies to investigate new models for understanding variables affecting student performance and interest, as well as examine and compare country-level performance (Ainley & Ainley, 2011; Dreschel, Carstensen, & Prenzel, 2011; Lin, Lawrenz, Lin, & Hong, 2012; Olsen & Lie, 2011).

Summary

Fascination is a construct that draws on the three components of mastery goals, curiosity, and interest. Similar to the articulation of learning activation outlined in Chapter 1, in which activation enables success which further strengthens activation, so too do the subdimensions of mastery goals, curiosity, and interest reinforce one another through fascination. Conceptually, students who are fascinated with science are drawn to certain science topics or science as a process (curiosity), and through reinforcement, seek more opportunities related to learning science (interest), and become aware of what they don't know as they work to close knowledge gaps, continuing to build competence and skills (mastery goals). Chapter 3 focuses directly on the measurement of student fascination in science as part of the larger ALES:14 study, which is also discussed in detail.

Chapter 3: Methodology

This dissertation seeks to extend previous research done as part of the Activated Learning Enables Success study of 2014 (ALES:14) by explicitly investigating changes in student fascination in science over time. Specifically, the current study will address the following research questions:

RQ1: What is the average change in science fascination of middle school students over time? And how much does this change vary, on average?

RQ2: To what extent is predicted student science fascination associated with the instructional characteristics of the classroom, including student-centric teaching, hands-on methods, and classroom technology use? And are these effects sustained over time?

RQ3: To what extent are middle schools equally successful in sustaining science fascination for students whose gender, race, family support for science, or economic backgrounds differ?

In this chapter, I discuss the ALES:14 study and the sample of students, and go on to describe the technical aspects of the measures of activation used in this research. Because the dimension of fascination is of primary interest in this dissertation as well as the outcome variable in the three research questions above, I provide more detail on this dimension than I do on the others. I then provide a brief overview of hierarchical modeling and argue that this analytical approach is the most appropriate to answer the research questions presented above. Finally, I conclude with the development of the specific models that will be used to address each question while including the rationale for the variables that will be used and the functional forms that will be tested.

Description of ALES:14 & Sample Characteristics

The Activated Learning Enables Success study of 2014 (ALES:14) was conducted by the Research Group at the Lawrence Hall of Science at the University of California, Berkeley and the Learning Research and Development Center at the University of Pittsburgh, collectively called the Science Learning Activation Lab (www.activationlab.org). The intent of the study was to test the theory of science learning activation; specifically, that the four dimensions of activation—competency beliefs, fascination, scientific sensemaking, and values—are related to outcomes of success: choice, engagement, perceived success, and content learning (Dorph, Cannady, & Schunn, 2016).

ALES:14 followed two grade-level cohorts of students across two school years in two U.S. cities. More than 3,700 students in grades 6 and 8 were tracked into their 7th and 9th grade years, respectively, in schools in the San Francisco Bay Area and Western Pennsylvania, including five urban middle schools recruited from Berkeley, CA and six urban middle schools from Pittsburgh, PA. Schools were recruited by contacting middle school science teachers at in-service events, and teachers were offered compensation that varied according to the number of participating classes (Bathgate & Schunn, 2017). From teachers who agreed to participate, schools were chosen to span a range of socioeconomic backgrounds, as well as diverse types of science learning experiences. Selected schools had as few as two and as many as eight teachers; each teacher had as few as one and as many as five classes of participating students. Table 3.1 below summarizes the student demographics and teacher, class, and student samples by school site for the initial cohort.

Table 3.1. Sample student characteristics by school, Year 1

School	% Non-White	% Missing race*	% Female	% Missing gender**	# Teachers	# Classes	# Students
1	48.3%	7.3%	63.2%	11.1%	1 6th 1 8th	5 6th 4 8th	141 6th 120 8th
2	58.2%	39.8%	32.3%	38.2%	4 6th 1 8th	6 6th 6 8th	280 6th 280 8th
3	73.7%	21.3%	39.2%	23.5%	1 6th 2 8th	2 6th 5 8th	128 6th 229 8th
4	49.8%	21.6%	38.1%	24.5%	2 6th 2 8th	5 6th 10 8th	166 6th 372 8th
5	73.4%	15.5%	43.0%	20.6%	1 6th 1 8th	6 6th 3 8th	186 6th 130 8th
6	63.0%	2.2%	32.6%	4.4%	1 6th/8th	1 6th 2 8th	54 6th 81 8th
7	38.7%	19.8%	39.8%	19.5%	3 6th 2 8th	5 6th 9 8th	296 6th 355 8th
8	29.9%	17.9%	37.7%	19.8%	1 6th/8th	3 6th 2 8th	162 6th 156 8th
9	58.0%	27.5%	30.5%	26.9%	1 6th 1 8th	5 6th 5 8th	183 6th 174 8th
10	76.5%	23.5%	28.0%	28.8%	1 6th 1 8th	2 6th 3 8th	52 6th 80 8th
11	64.5%	20.0%	45.5%	18.2%	1 6th	2 6th	110 6th
TOTAL	**54.2%**	**22.0%**	**38.7%**	**23.4%**	**28**	**42 6th** **49 8th**	**1,758 6th** **1,977 8th**

* Includes students who selected "I don't know" for racial/ethnic background.
** Includes students who selected "Prefer not to answer" for gender.

Visible above, many students chose not to report either their racial/ethnic background or their gender identity; students frequently omitted both. In fact, while

22.0% had missing data for race/ethnicity, and 23.4% had missing data for gender, only 26.3% unique students had missing data across the two variables, implying a great degree of overlap between missingness on the two variables.

A number of instruments were used to collect data across two school years and five testing occasions. To gather the demographic information used to describe students above, a personal background questionnaire was administered at the outset of the study (T1), in the fall of Year 1 (Y1). In addition to gender and racial/ethnic background, this instrument also asked about the student's mother's and father's education and occupation, as well as questions pertaining to resources available at home and family support for school. This background questionnaire was offered a second time in the fall of Year 2 (T4) for students who either missed the previous administration or were new to the classes involved in the study.

Students were measured on all four dimensions of science activation five times over the course of the two school years: three times during Year 1 (fall (T1), winter (T2), and spring (T3) of the school year) and twice more in Year 2 (fall (T4) and spring (T5) of the school year). These assessments of fascination, values, competency beliefs, and scientific sensemaking were administered as self-report surveys, and will be described in additional detail in the next section of this chapter.

Finally, classroom observations paired with teachers' self-reports about patterns of instruction were collected in Year 1 only. Researcher observers visited each classroom anywhere from one to three times over the course of the year, conducting two 5-minute "sweeps" of classroom activities to assess the amount of student-centric (as opposed to textbook) teaching and the amount of hands-on activities provided. Teachers also completed instructional logs (as few as 1, as many as 83) detailing classroom activities pertaining to technology-use. As part of the original study, each of these measures—student-centric teaching, hands-on teaching, and classroom technology use—were distilled into a proportion of time spent employing the instructional strategies for each class.

Figure 3.1 shows the pattern of data collection for Years 1 and 2 including those measures relevant to the discussion of model development that follow.

Figure 3.1. ALES:14 measures over time

	2014-2015 6^{th} & 8^{th} grades			**2015-2016 7^{th} & 9^{th} grades**	
	Fall		**Spring**	**Fall**	**Spring**
Background Questionnaire	✓ T1	- -	- -	✓ T4	- -
Fascination	✓ T1	✓ T2	✓ T3	✓ T4	✓ T5
Classroom Observations	←	x 1-3	→	- -	- -

Note: "- -" in the figure above indicates the instrument/tool was not used in that time period and is therefore unavailable for inclusion.

The ALES:14 data set was chosen to address the stated research questions in this dissertation due to its longitudinal nature, the inclusion of a robust set of variables from which to select, and of course, the unique Activation framework measuring fascination in science. Over the last five years, the Activation Lab has produced a number of published works thanks to the quality of the research and resulting data (e.g., Bathgate & Schunn, 2016, 2017; Bonnette, Crowley, & Schunn, 2019), but to date, none have investigated changes in fascination or addressed the underlying relationships possibly present in the data as this work seeks to do.

Technical Aspects of the Activation Instruments

The activation measures used in the ALES:14 study were developed by researchers at the Activation Lab, having been tested and revised in a study two years prior. Each dimension—fascination, values, competency beliefs, and scientific sensemaking—was subjected to the same rigorous psychometric development procedures: first, a sequence of iterative exploratory factor analyses (EFA), followed by confirmatory factor analyses (CFA), and then item response theory (IRT) item-fit analyses. The final scale for each measure was then subjected to a differential item functioning (DIF) analysis and tested for measurement invariance (Moore, Bathgate, Chung, & Cannady, 2011). Importantly, each activation construct was conceived as semi-malleable, and therefore scores on each dimension are "amenable to intervention" and subject to change over time.

Fascination was defined as having "aspects of what many researchers have referred to as curiosity, interest in science both in and out of school, and mastery goals for science content" (Chung et al., 2016b, p.1). It was measured using an eight-item scale, presenting students with statements such as "I need to know how objects work" and "In general, I find science…" (see Appendix A for full instrumentation), each item with four response options. Activation Lab researchers reassessed the scale for internal reliability, unidimensionality, and Rasch model fit using a sample of more the 2,900 student responses. The fascination scale produced strong

Cronbach's and polychoric alpha coefficients (.88 and .90, respectively, with values 0.80 and above desirable), implying that individuals responded similarly across items and the scale has sufficient ordinal properties. Exploratory Factor Analysis revealed a single latent factor, and partial credit Rasch model fitting showed satisfactory infit and outfit statistics, with the exception of one item: "I wonder about how nature works…" Still, the overall person-separation reliability statistic was found to be satisfactory (EAP/PV = 0.868, again with reliability coefficients desirable at 0.80 and higher), establishing further evidence of internal validity (Chung et al., 2016b). The activation dimensions of values, competency beliefs, and scientific sensemaking were all subjected to the same rigorous reliability, unidimensionality, and Rasch model testing, and produced similarly acceptable results (Chung et al., 2016a; Chung et al., 2016c; Chung et al., 2017).

Table 3.2 presents the psychometric properties of all four activation dimensions, including scale length, reliability coefficient (Cronbach's alpha), root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker Lewis index (TLI), and model chi-square for measurement invariance. Reliability coefficients of 0.8 and above are usually deemed sufficient while those below but near 0.8 are considered marginal (Andrich, 1982). Desirable RMSEA values fall below 0.08, while CFI and TLI values should be greater than or equal to 0.90 and 0.95, respectively. The measurement invariance analysis tests occasions of the administration of the scale against one another to establish equality of the factor loadings, with the null hypothesis that the models fit perfectly. As shown below, the fascination scale is the only dimension with satisfactory reliability and CFA fit statistics; however, it is also the only scale shown to have a statistically significant mean difference between test occasions. When this is the case, it is often not possible to determine if observed changes are due to changes in perception, changes in the instrument, or a combination of both. Despite this, researchers deemed the scale "reasonably measurement invariant," and further stated, "Removal of (item 4) resulted in a scale that was measurement invariant and had higher reliability, but poorer model fit. It was retained in the final item set" (Moore et al., 2011, p. 28). Full instrumentation can be found in Appendix A.

Table 3.2. Fit statistics for all activation dimensions

Dimension	# Items	Reliability	RMSEA	CFI	TLI	Measurement Invariance
Fascination	8	0.88	0.065	0.983	0.977	$\chi^2 = 16.45$, $p = 0.02$, mean diff = 0.121
Values	4	0.70	0.089	0.987	0.962	$\chi^2 = 2.56$, $p = 0.46$
Competency Beliefs	9	0.90	0.105	0.971	0.961	$\chi^2 = 13.14$, $p = 0.11$
Scientific Sensemaking	12	0.75	0.037	0.976	0.970	$\chi^2 = 19.72$, $p = 0.07$

Variable Selection

In order to address the proposed research questions meaningfully, it is essential to identify key variables present in the data that are supported by previous research for studying fascination in science. Equally important, appropriate analysis of longitudinal data hinges on the outcome as a "continuous, psychometrically robust variable whose values change systematically over time" (Singer & Willett, 2003, p. 13). As detailed in the description of the ALES:14 data structure and the technical specifications of the measurement scales in Table 3.2 above, the fascination in science outcome variable has good psychometric properties including strong reliability (α=0.88) and small error of measurement (RMSEA=0.065), and was defined as semi-malleable and subject to change over time. Beyond the repeated measure of fascination, additional time-invariant student- and classroom-level variables are specifically listed in Research Questions 2 and 3 that warrant discussion.

At the student level in RQ3, gender and minority status are included to better understand potential mean differences in science fascination scores between groups as well as differences in growth rates over time between groups. Research continues to show that science attitudes, achievement, and career interests vary across gender identities (e.g., Baram-Tsabari & Yarden, 2011; Christidou, 2011; Jacobs-Priebe & Crowley, 2013) and racial/ethnic identities (e.g., Catsambis, 1995). The proposed analyses, however, do not seek to treat either variable (gender, minority status) as statistical controls, nor will they be given causal interpretations. Although Raudenbush and Bryk (2002) emphasize the importance of including statistical adjustments for individuals' demography to avoid biasing growth modeling, Spector and Brannick (2011) caution against the use of these control variables, noting that their inclusion is rarely theory-driven and may lead to erroneous inferences. Furthermore, Holland (2003) reminds us that race and gender variables are *not* causal even when included in causal modeling, as other underlying factors often are omitted in such models, and observed correlational patterns merely identify associations rather than causation. Race is not a cause, gender is not a cause; the effects of identifying as or being perceived as, for example, Black, or Female, however, may impact lived experiences in ways that are not equally distributed to others who are do not identify/are not perceived similarly. Here, gender is included as a variable of interest so that this research can describe changes in fascination scores over time for both boys and girls[2] included in the sample, and examine whether instructional practices have similar effects on fascination across gender groups; race is included similarly.

Given the potential role fascination in science may play in developing career interest in science, it is especially important to investigate changes in fascination scores over time separately for girls and those who identify as students of color, as historically both groups have been underrepresented in science professions (Archer, DeWitt, Osborne, Dillon, Willis, & Wong, 2012). Data from the National Science Foundation (NSF) show that while women comprised 50% of the college-educated workforce and 40% earned their highest degree in a science or engineering field, only 28% of science and engineering occupations were held by women in 2015. Hispanics, Blacks, and American Indians or Alaska Natives comprise 27% of the U.S. population age 21 and older, and yet represent only 11% of workers in science and engineering professions (National Science Board, 2018). As evidence of the importance of understanding and addressing these inequities, NSF recently released a new funding stream dedicated explicitly to combatting

[2] Although I do not conceive of gender as a binary construct, the dataset only includes information on boys and girls, lacking non-binary or other gender designations beyond “Prefer not to answer.”

systemic racism in STEM, called *Racial Equity in STEM Education* (National Science Foundation, 2021a).

Additional student variables include home resources for science and family support for science. Findings from TIMSS 2019 reinforce the powerful relationship between students' socioeconomic environment—measured through resources for learning in the home—and their educational achievement (Mullis et al., 2020). Further exploration of TIMSS data from prior cycles has shown that these home resources and attitudes toward science are also positively related (Geesa, Izci, Song, & Chen, 2019). Perhaps intuitively, empirical research has demonstrated there are both direct and indirect effects on students' attitudes about science and science learning associated with parental involvement, as measured by students' perception of parental support for class and school activities (George & Kaplan, 1998). Modeling ALES:14 scores for home resources for science and family support for science will help determine the extent to which student fascination in science is associated with these home-related variables.

Finally, at the classroom level in RQ2, instructional characteristics of the classroom, including student-centric teaching, hands-on methods, and classroom technology use are also included. While most educational research, including this dissertation, focuses on student-level outcomes, it is important to understand how teachers create conditions of involvement since so much educational time takes place in the classroom. Prior research has demonstrated that students in classrooms with instructional strategies that foster high involvement have significantly more positive affect toward learning (Turner, Meyer, Cox, Logan, DiCintio, & Thomas, 1998). Specifically, student-centered approaches (Kang & Keinonen, 2018), hands-on activities (Ornstein, 2006), and technology use in the classroom (OECD, 2015) have all been shown to influence student attitudes and classroom behaviors, and will be included in the development of the models.

Overview of Hierarchical Modeling

The longitudinal nature and nested data structure of ALES:14 affords the relatively unique opportunity of looking at changes in student fascination in science over time. Single-timepoint measurements of educational or psychological constructs, while capable of assessing students' states of being or abilities at the time of testing, suffer from the misinterpretation that scores relate to student potential (Dumas & McNeish, 2017). In fact, this critique of psychometric testing in education was first noted over 90 years ago by W.E.B. DuBois, who lamented the fact that many people conflated current ability and future capacity, which resulted in a self-fulfilling prophecy as low-scoring students were often denied the instruction to achieve a higher potential (1920/2013). In addition to the proliferation of reliable longitudinal educational achievement data in the U.S., recent advances in nonlinear growth modeling and statistical computing should help to address this major issue in educational measurement (Lohman, 2006).

One approach that has been utilized in this effort is dynamic assessment (DA) (Feuerstein, 1979). DA features multiple testing occasions integrated with instruction by a clinician; unfortunately, because of the substantial time investment required and disruption to the educational setting, this method has not been widely applied in U.S. states or school districts. Another approach is Dynamic Measurement Modeling (DMM), which was developed to estimate student capacity using large-scale longitudinal data and without the need for intensive clinical work like DA (McNeish & Dumas, 2017). Although capable of being flexibly applied to a number of research questions allowing for a diversity of growth curve shapes, DMMs incorporate individualized growth trajectories for *every* student in the data set in an effort to model an upper capacity asymptote as a prediction of potential (Dumas & McNeish, 2017). The additional parameterization necessary to model growth in this fashion adds a layer of complexity that is not necessary to address the stated research questions in this dissertation.

Another approach that is widely popular for appropriately analyzing nested data is multilevel (MLM) or hierarchical linear modeling (HLM). Although the "unnecessary ubiquity" of HLMs has been called out by some methodologists citing reasonable alternatives (e.g., Generalized Estimating Equations (GEEs), cluster robust-standard errors (CR-*SE*s)), there is a reason for their popularity: they provide fully modeled within-cluster correlations and cluster-specific interpretations of fixed and random effects (McNeish, Stapleton, & Silverman, 2017). Worth noting, there are multiple ways data structures can be nested and therefore appropriate for HLM analysis. Time-series data—repeated observations of the same variable(s) for an individual—have measures nested within persons, and as such, longitudinal data can be considered nested. Individuals may also be nested within groups, such as classrooms or schools or organizations. Higher-order groupings can also occur, such as individuals nested within schools and schools nested within states, for example (Huta, 2014). MLMs have distinct advantages:

> *Multilevel regression modeling does not correct bias in the regression coefficient estimates compared with an [Ordinary Least Squares] model; however, it produces unbiased estimates of the standard errors associated with the regression coefficients when the data are nested and easily allows group characteristics to be included in models of individual outcomes.* (O'Dwyer & Parker, 2014)

ALES:14 contains measures of fascination repeated within individuals, unique but unchanging individual attributes, and these individuals nested within classrooms/schools[3], so a three-level hierarchical model will be required to deal with similarities in unit-level observations. If unaccounted for by using a traditional Ordinary Least Squares (OLS) approach, the nature of these data could lead to correlated errors, an unaccounted for heterogeneity of regression slopes, and aggregation bias. A hierarchical (multilevel) approach, on the other hand, provides a better estimate of individual effects as it models effects across different levels, and partitions the model variance between the different levels (Raudenbush & Bryk, 2002). More importantly, HLMs are quite flexible and assume little about the

[3] In this analytical approach, classroom and school variability are confounded. As classroom-level observational data are entered into level 3, the uppermost nested structure represents students within classrooms. These classrooms are also nested within schools, which remain unmodeled as there are no school-level covariates to include (although fixed effects for schools could be added through the inclusion of dummy variables), meaning that school-to-school variability is confounded at the classroom level. For brevity and clarity, I will continue to refer to level 3 as the classroom level – rather than the classroom/school level – due to the inclusion of classroom instructional covariates at L3.

structure of the data, allowing for differences in the number of time point measures and the spacing of these observations, unlike Repeated-Measures ANOVA (Bickel, 2007).

Allowing for differences in the number of time point measures is especially advantageous in this longitudinal analysis spanning two school years and five administrations of the activation instruments. By employing a multilevel model with repeated measures at Level 1, all students, even those with missing activation scores, are included in analysis. Under an assumption of data missing at random (MAR), HLM is able to accommodate missing data by utilizing multiple model-based imputations to extract full information for analysis (Shin & Raudenbush, 2011). In considering the ALES:14 data, between 21.3% and 28.7% of students are missing fascination scores for each time point in Year 1 alone, and only 55.7% of students recorded fascination scores on all three measurement occasions. Rather than throwing out a 44.3% of the student sample, the hierarchical structure and multiple model-based imputation allow us to consider the full sample of fascination scores, even those that were never recorded, by imputing values based on existing information. However, this imputation method does not apply to student-level covariates at Level 2. The prediction of discrete variables such as demographic characteristics (e.g., gender, race/ethnicity, etc.) is difficult and controversial, and therefore students with these data missing will be excluded from analysis.

Developing the Models

When building HLMs, the null or "unconditional" model is specified first. This random coefficients model does not attempt to model variability at any level and the intercepts and growth rates are allowed to vary randomly across students (L2) and classes (L3). Furthermore, it assumes a linear function.

Level 1: $$Y_{tij} = \pi_{0ij} + \pi_{1ij}a_{tij} + e_{tij} \tag{3.1}$$

Level 2: $$\pi_{0ij} = \beta_{00j} + r_{0ij} \tag{3.2a}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij} \tag{3.2b}$$

Level 3: $$\beta_{00j} = \gamma_{000} + \mu_{00j} \tag{3.3a}$$

$$\beta_{10j} = \gamma_{100} + \mu_{10j} \tag{3.3b}$$

Mixed Model:

$$Y_{tij} = \gamma_{000} + \gamma_{100}a_{tij} + r_{0ij} + r_{1ij}a_{tij} + \mu_{00j} + \mu_{10j}a_{tij} + e_{tij} \tag{3.4}$$

In equation 3.1 above, the outcome variable (Y_{tij}) is fascination score at time *t* for individual *i* in class *j.* This L1 unconditional growth model includes one "within-person" variable (a_{tij}) to represent *time* between measurement occasions, a linear function. For all models, the time variable will be "rough-centered" to eliminate multicollinearity between terms when introducing quadratic and cubic terms to test non-linear models. By placing the Year 1 winter testing occasion (T2) as the intercept, this allows the instructional effects modeled at L3 to be observed over all five time-points (such that T1=-1, T2=0, T3=1, T4=2, and T5=3), since it is unlikely that instructional effects of Y1 are observable at the start of the academic year anyway.

This null model serves as a baseline to which subsequent models will be compared. From this initial modeling, the intraclass correlation coefficient (ICC) can be calculated, which captures the degree to which variance depends upon the nested structure of the data:

L2 ICC: $\rho = \frac{\tau_{00}}{\sigma^2 + \tau_{00} + \tau_{000}}$ (3.5)   L3 ICC: $\rho = \frac{\tau_{000}}{\sigma^2 + \tau_{00} + \tau_{000}}$ (3.6)

Equation 3.5 above shows the proportion of variance that can be attributed to individual differences, while equation 3.6 models the proportion of variance accounted for by classroom nesting. In developing the null model, I will test the ICCs to see if they are significantly different from 0, which would indicate that important variance is indeed being taken into account by the multilevel modeling approach.

From here, the model could be built-up using one of several procedures: forward selection, backward elimination, stepwise selection, or simultaneous block-entry approach. Regardless of the method of entering predictors, adding covariates at either L2 or L3 requires centering variables around the grand-mean or around group (L3) means in order to provide a true zero point for that variable (thus, centering is not necessary for dummy variables accounting for certain demographic categories, for example). While theory should always guide model development in terms of predictors added, parsimony is preferred to a more complex model.

Model development in this dissertation will employ a backward elimination process to ensure the necessary functional forms are tested using higher-order (i.e., quadratic, cubic) terms. In this process, the highest order polynomials are entered at the same time and avoid being excluded early due to other variable selection decisions. Determining the final functional form is performed by testing both the quadratic and cubic terms for statistical significance and removing the term associated with the highest p-value greater than α_{crit} (possibly both, retaining the linear model); remaining predictor variables are then assessed using the same method. In this case, α_{crit} may be set at .05 but is not required to be so by any convention as higher thresholds may work well. This is especially true when considering additional model-fit criterion: the deviance statistic, Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) are commonly evaluated across models, with larger models employing more parameters generally having a better fit (Anderson, 2012). As a measure of fit, the deviance statistic does not take into account complexity, while both AIC and BIC are penalized measures that add to the deviance based on the number of predictors entered; BIC is further contingent on sample size. Because sample sizes differ at different levels of a nested model, AIC often is recommended for its straightforward calculation and comparability (Boedeker, 2017). Thus, the best choice of model will balance size (parsimony) with model fit.

Once variables have been entered into the model, estimates of the additional variance explained by the predictor variables at each level are computed, comparing against the unconditional model. Because there is no direct measure of variance accounted for by HLMs, equations 3.7, 3.8, and 3.9 below are called "pseudo R^2" statistics:

L1 pseudo R^2: $$\frac{\sigma^2_{unconditional} - \sigma^2_{conditional}}{\sigma^2_{unconditional}} \quad (3.7)$$

L2 pseudo R^2: $$\frac{\tau_\pi(unconditional) - \tau_\pi(conditional)}{\tau_\pi(unconditional)} \quad (3.8)$$

L3 pseudo R^2: $$\frac{\tau_\beta(unconditional) - \tau_\beta(conditional)}{\tau_\beta(unconditional)} \quad (3.9)$$

Finally, model fit is estimated by comparing deviance statistics between models, with larger values indicating poorer fit. Typically, researchers build up levels sequentially rather than simultaneously, so that these deviance statistics take on meaning—another benefit of beginning with the null model—requiring that the same number of levels be used throughout. Thus, L1 would be developed, with deviance statistics compared to the null model, and "finalized" before beginning the development of L2; the same process would then be repeated for L2 and L3. As the difference between two deviance statistics follows a chi-square distribution (where *df*=difference in the number of parameters estimated), determining if a statistically significantly "better" model has been achieved simply entails evaluating the resulting chi-square statistic given the degrees of freedom.

Several assumptions regarding hierarchical growth models will also be investigated during the model development process – three concerning error structure and two concerning the predictors themselves:

1. L1 residuals are independent, and normally distributed (0, σ^2).
2. L2 & L3 random effect residuals are independent, and multivariate normal (0, τ^2).

3. Residuals between levels are independent (i.e., no covariance).
4. L1 predictors & residuals are independent; L2 predictors & residuals are independent; L3 predictors & residuals are independent.
5. Predictors at each level are independent of the random effects at other levels.

HLM software allows the residuals from multilevel analyses to be saved which then makes testing of the assumptions of residuals possible.

Developing the conditional model involves selecting relevant L1, L2, and L3 predictors using the backward elimination criteria described above. The addition of the L2 measured person characteristics specify the variation in the L1 intercepts and slopes (i.e., between individuals), while the addition of the L3 instructional characteristics further specify the variation in the L2 intercepts and slopes (i.e., between classes). The model will subsequently assume the form:

L1:
$$Y_{tij} = \pi_{0ij} + \pi_{1ij}a_{tij} + e_{tij} \tag{3.10}$$

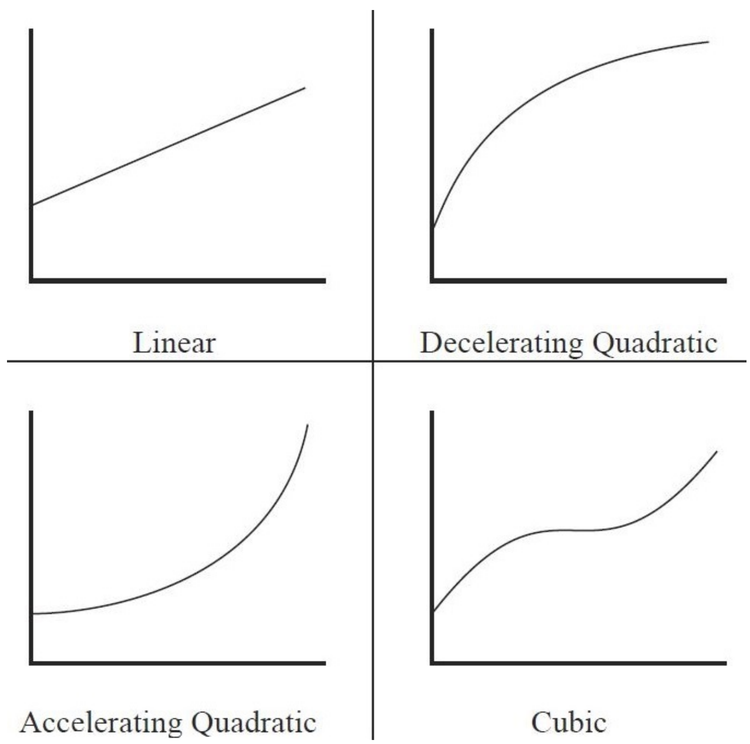
L2:
$$\left\{\begin{aligned} \pi_{0ij} &= \beta_{00j} + \sum_{q=1}^{Q_n} \beta_{nqj}X_{qij} + r_{0ij} && \text{(3.11a)} \\ \pi_{1ij} &= \beta_{10j} + \sum_{q=1}^{Q_n} \beta_{nqj}X_{qij} + r_{1ij} && \text{(3.11b)} \end{aligned}\right.$$

L3:
$$\left\{\begin{aligned} \beta_{00j} &= \gamma_{000} + \sum_{s=1}^{S_{nq}} \gamma_{nqs}W_{sj} + \mu_{00j} && \text{(3.12a)} \\ \beta_{10j} &= \gamma_{100} + \sum_{s=1}^{S_{nq}} \gamma_{nqs}W_{sj} + \mu_{10j} && \text{(3.12b)} \end{aligned}\right.$$

Visible above, L2 student-level covariates are represented by vector X_{qij} in equations 3.11a and 3.11b, while L3 classroom-level instructional characteristics are represented by vector W_{sj} in equations 3.12a and 3.12b.

This research does not assume a linear model; rather, one of the research questions of interest is to determine which functional form models the data best. Examples of higher-order polynomials include quadratic (decelerating or accelerating) and cubic, as depicted in Figure 3.2 below (Anderson, 2012).

Figure 3.2. Common non-linear functional forms



Generally, linear functions are assumed for growth models with three or fewer measurements over time, whereas a fourth data point enables quadratic testing, and a fifth, cubic (Anderson, 2012). As adherence to the proper functional form is one of the foremost validity concerns when drawing inferences from growth model results, and because the ALES:14 data contains five measures of fascination, testing each of these polynomial forms will be necessary. As described earlier, the backward elimination approach will be employed in which all higher-order polynomials are included and entered into the model at the same time, and then one by one, the highest order non-significant terms are eliminated if possible.

Lastly, because of the nature of longitudinal data in which students were tested three times during the first school year and twice in the second school year – with an obvious break in between either 6th and 7th or 8th and 9th grades for summer vacation – the discontinuous functional form will also be tested. This discontinuity can be evaluated for changes in *level*, but not slope (as in Figure 3.3), changes in *slope*, but not level (as in Figure 3.4), or changes in both slope *and* level (as in Figure 3.5) (Anderson, 2012).

Figure 3.3. Discontinuous growth: change in level

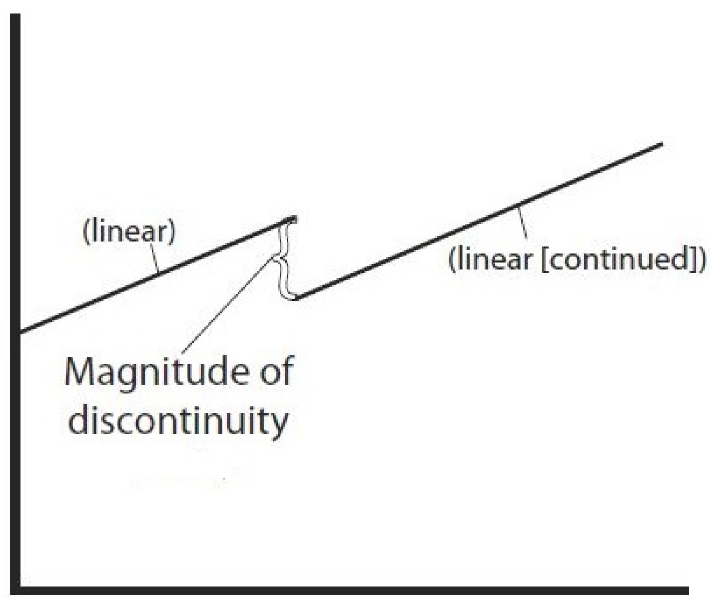


Figure 3.4. Discontinuous growth: change in slope

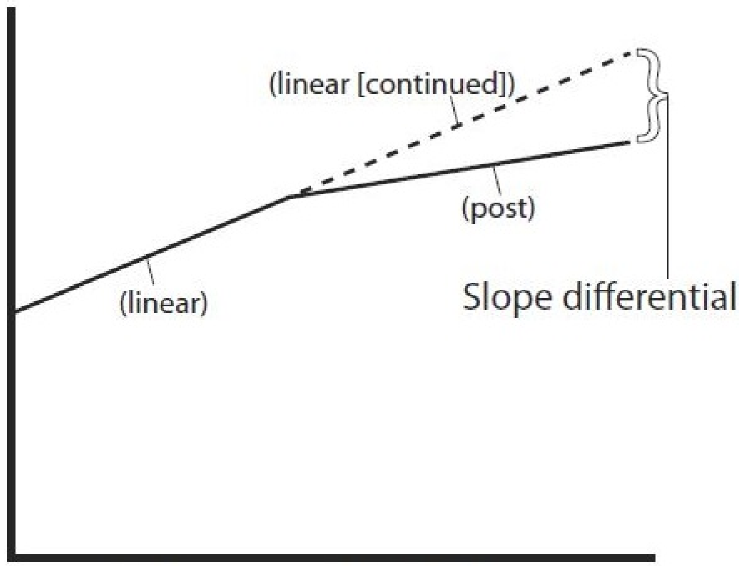
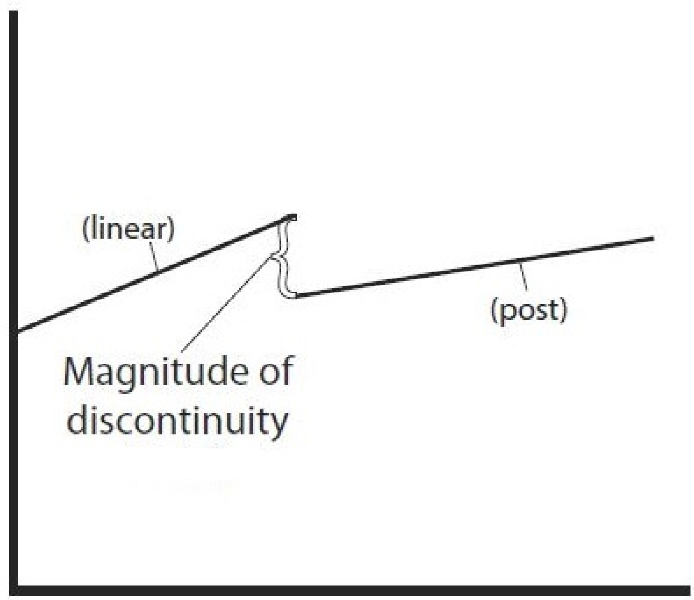


Figure 3.5. Discontinuous growth: change in level and slope



Addressing the Research Questions

Building a three-level growth model with fascination measures repeated within students and students nested within classrooms allows for the estimation of individual growth trajectories and thus an examination of patterns of change in science fascination over time. Incorporating additional student-level covariates at level 2 provides more information regarding whether and to what degree demographic variables and person-level characteristics contribute to variability in initial status and in their change in science fascination over time.

Moving beyond the null model and returning to the development of the conditional model presented in equations 3.10-12b above allows us to address RQ1: *What is the average change in science fascination of middle school students over time? And how much does this change vary, on average?* We begin building at L1 to establish the random coefficients growth model by including quadratic and cubic time terms to account for non-linear growth patterns in fascination over time:

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}a_{tij} + \pi_{2ij}a_{tij}^2 + \pi_{3ij}a_{tij}^3 + e_{tij} \tag{3.10'}$$

For equation 3.10':

- Y_{tij} represents the outcome: fascination score at time *t* for student *i* in class *j;*
- π_{0ij} represents the intercept: fascination score of student *i* in class *j* at time 0;
- a_{tij} represents the measurement occasion (time point *t*) for student *i* in class *j*; as noted earlier, these time points will be "rough-centered" at the second measurement occasion such that:

- at T1, a_t=-1, a_t^2=1, and a_t^3=-1
- at T2, a_t=0, a_t^2=0, and a_t^3=0
- at T3, a_t=1, a_t^2=1, and a_t^3=1
- at T4, a_t=2, a_t^2=4, and a_t^3=8
- at T5, a_t=3, a_t^2=9, and a_t^3=27

- π_{1ij} represents the linear component (a_{tij}, or time point to the first power), which is the instantaneous growth rate of student *i* in class *j* at time 0;
- π_{2ij} represents the quadratic component (a_{tij}^2, or time point squared), which is the acceleration (rate of change) in the growth trajectory of student *i* in class *j* at time 0;
- π_{3ij} represents the cubic component (a_{tij}^3, or time point to the third power), which is the change in the acceleration of growth of student *i* in class *j* at time 0; and
- e_{tij} represents random error (in the simplest case, we assume the errors are normally distributed, independent, with constant variance).

While there are now four level 2 equations, the L2 model is still unconditional, with no student-level covariates included, as seen in equations 3.11a'-d':

$$\pi_{0ij} = \beta_{00j} + r_{0ij} \tag{3.11a'}$$

$$\pi_{1ij} = \beta_{10j} + r_{1ij} \tag{3.11b'}$$

$$\pi_{2ij} = \beta_{20j} + r_{2ij} \tag{3.11c'}$$

$$\pi_{3ij} = \beta_{30j} + r_{3ij} \tag{3.11d'}$$

- β_{00j} represents the mean fascination score at time 0 (intercept) within class *j*; r_{0ij} represents the random student-level effect around mean fascination score within class *j*;
- β_{10j} represents the mean growth rate in fascination within class *j*; r_{1ij} represents the random student-level effect around mean growth rate within class *j*;
- β_{20j} represents the mean acceleration in the growth trajectory within class *j*; r_{2ij} represents the random student-level effect around mean acceleration in the growth rate within class *j*; and
- β_{30j} represents the mean change in acceleration of the growth trajectory within class *j*; r_{3ij} represents the random student-level effect around mean change in acceleration of the growth rate within class *j*.

Similarly, the L3 model with four equations is also unconditional, as seen in equations 3.12a'-d':

$$\beta_{00j} = \gamma_{000} + \mu_{00j} \tag{3.12a'}$$

$$\beta_{10j} = \gamma_{100} + \mu_{10j} \tag{3.12b'}$$

$$\beta_{20j} = \gamma_{200} + \mu_{20j} \tag{3.12c'}$$

$$\beta_{30j} = \gamma_{300} + \mu_{30j} \tag{3.12d'}$$

- γ_{000} represents the overall mean fascination score at time 0 (intercept) for all classes; μ_{00j} represents the random class-level effect around overall mean fascination score;

- γ_{100} represents the overall mean growth rate in fascination for all classes; μ_{10j} represents the random class-level effect around overall mean growth rate;
- γ_{200} represents the overall acceleration in growth trajectory for all classes; μ_{20j} represents the random class-level effect around overall mean acceleration in the growth rate; and
- γ_{300} represents the overall change in acceleration of the growth trajectory for all classes; μ_{30j} represents the random class-level effect around overall mean change in acceleration of the growth rate.

The reduced form expression for this random coefficients growth model, in which I allow the intercept (π_{0ij}), growth (π_{1ij}), acceleration (π_{2ij}), and cubic term (π_{3ij}) to vary randomly, is presented in equation 3.13:

$$Y_{tij} = \gamma_{000} + \gamma_{100}a_{tij} + \gamma_{200}a_{tij}^2 + \gamma_{300}a_{tij}^3 + r_{0ij} + r_{1ij}a_{tij} + r_{2ij}a_{tij}^2 + r_{3ij}a_{tij}^3 + \mu_{00j} + \mu_{10j}a_{tij} + \mu_{20j}a_{tij}^2 + \mu_{30j}a_{tij}^3 + e_{tij} \tag{3.13}$$

Running this initial random intercepts and slopes model will test whether the addition of the nonlinear components (time^2 and time^3) appropriately model the pattern of student fascination scores over time and significantly account for variance in fascination scores. These time components are the only fixed effects in the model, and we will test the following hypotheses:

1. Is the linear term significantly different from zero?

   $H_0: \pi_{1ij} = 0$ $\qquad$ $H_1: \pi_{1ij} \neq 0$

2. Is the quadratic term significantly different from zero?

$$H_0: \pi_{2ij} = 0 \qquad\qquad H_1: \pi_{2ij} \neq 0$$

3. Is the cubic term significantly different from zero?

$$H_0: \pi_{3ij} = 0 \qquad\qquad H_1: \pi_{3ij} \neq 0$$

Decisions on these three hypotheses will allow us to specify L1 appropriately in subsequent models.

Importantly, estimating random effects at levels two and three allow for an investigation of the covariance structures of these random components. To determine if there is residual variation between students, we analyze the L2 variance-covariance matrix:

$$T_\pi = \begin{bmatrix} \tau_{\pi_{00}} & & \\ \tau_{\pi_{01}} & \tau_{\pi_{11}} & \\ \tau_{\pi_{02}} & \tau_{\pi_{12}} & \tau_{\pi_{22}} \end{bmatrix}$$

This allows us to test if the variability in the intercepts is statistically different from zero:

$$H_0: \tau_{\pi_{00}} = 0 \qquad\qquad H_1: \tau_{\pi_{00}} \neq 0$$

Finally, to determine if there is residual variation between classes, we analyze the L3 variance-covariance matrix:

$$T_\beta = \begin{bmatrix} \tau_{\beta_{00}} & & \\ \tau_{\beta_{01}} & \tau_{\beta_{11}} & \\ \tau_{\beta_{02}} & \tau_{\beta_{12}} & \tau_{\beta_{22}} \end{bmatrix}$$

This allows us to test if the variability in the intercepts is statistically different from zero:

$$H_0: \tau_{\beta_{00}} = 0 \qquad\qquad H_1: \tau_{\beta_{00}} \neq 0$$

Once the average change in science fascination has been modeled by establishing the functional form at L1 (i.e., linear, quadratic, or cubic), we can address RQ3*: To what extent are middle schools equally successful in sustaining science fascination for students whose gender, race, family support for science, or economic backgrounds differ?* To do this, we build up the vector X_{qij} of student-level covariates specified equations 3.11a and 3.11b that are hypothesized to be associated with fascination "on top" of the correctly specified L1. These individual background variables include student gender (e.g., Baram-Tsabari & Yarden, 2011), racial/ethnic minority status (e.g., Catsambis, 1995), availability of resources at home, and family value of education (e.g., George & Kaplan, 1998).

To address RQ2: *To what extent is predicted student science fascination associated with the instructional characteristics of the classroom, including student-centric teaching, hands-on methods, and classroom technology use? And are these effects sustained over time?* we build up the vector W_{sj} of classroom-level variables from equations 3.12a and 3.12b "on top" of the L2 covariates, such that L3 variables now include the average proportion of time spent providing hands-on instruction, the average proportion of time spent doing student-centric activities, and the average proportion of time spent using technology in the classroom.

Several HLMs will be tested for both sixth and eighth graders, evaluating non-linear terms while building levels 2 and 3 and comparing against discontinuous models, with the intention to create parsimony whenever possible. Because the nature of the data remains consistent across models (i.e., same number of levels), it will be possible to compare the accuracy afforded by each as measured by variance explained, the precision of the estimates as measured by the standard errors of the regression coefficients, and of course, the overall significance of the models and relative impact of each predictor in influencing science fascination.

Chapter 4: Results

The previous chapter described the analysis plan for the current study, the results of which are presented here. This chapter begins with a discussion of the data, including missing data analysis followed by descriptive statistics for the final sample, as well as a description of each of the variables beyond the theoretical justifications presented in Chapter 3. The subsequent sections describe in detail the model development process for the 6^{th} and 8^{th} grade cohorts, respectively: beginning with an exploration of potential functional forms at level 1, followed by the construction of the models at levels 2 and 3, and finally model checking using residual analyses to test the assumptions of HLM laid out in the previous chapter. The chapter concludes by offering an interpretation of findings from the final models before leading into the final chapter, which will explicitly address each of the proposed research questions:

RQ1: What is the average change in science fascination of middle school students over time? And how much does this change vary, on average?

RQ2: To what extent is predicted student science fascination associated with the instructional characteristics of the classroom, including student-centric teaching, hands-on methods, and classroom technology use? And are these effects sustained over time?

RQ3: To what extent are middle schools equally successful in sustaining science fascination for students whose gender, race, family support for science, or economic backgrounds differ?

Discussion of Data

While sample student characteristics are presented in Table 3.1 of Chapter 3, that preliminary presentation offers only a glimpse into the data prior to analysis. In fact, of the 3,735 students in the data file, only 2,309 cases (62% of the initial Year 1 sample) were retained for final analyses due to missing data. HLM remains flexible for missing data among level 1 observations—in this dissertation, any of the 5 instances of fascination measurement—however, the software excludes cases with missing data at levels-2 and -3 (Raudenbush, Bryk, Cheong, Congdon Jr., & du Toit, 2019, p. 77). At the class level (L3), six teachers had missing observational data for the percentage of student-centric teaching (one of whom was also missing self-report data for hands-on teaching/classroom technology use), affecting 16 classes and 488 students. Missingness at the student level (L2) impacted another 826 students: 176 either neglected to self-identify in terms of gender or race or selected "Prefer not to answer" for one or both questionnaire items, while the remaining 650 students were missing more than one of the race, gender, family support for science, or home resources variables. Sixty (60) students selected "I don't know" as their only racial identity category, and were therefore uncategorizable. Finally, 52 students had no recorded fascination scores (L1) and were therefore omitted from analyses.

Although imputation of data is possible for cases with missing variables, it was decided to include only cases with "full" data in model building—those with at least one recorded *fascination* score at L1 and all variables observed at L2 and L3—using complete case analysis. One of the primary reasons to keep cases with missing data is to avoid the loss of statistical power, which in turn affects the efficiency of the estimates. Even with substantial data loss (38% of the person-level observations), the remaining sample sizes provide more than adequate power for two unique grade level models. Power analyses conducted using Optimal Design software for repeated measures analyses show that roughly 575 individuals over 5 time points are required to obtain a power of 0.80, assuming a Type I error rate of 0.05, a small minimum detectable effect size ($\delta \geq 0.24$), and variability estimates obtained from the preliminary unconditional models (detailed in the next section, $\sigma^2 = 0.13$ and $\tau = 0.12$). Given that the retained grade 6 sample is 999 students and the grade 8 sample is over 1,300, sufficient data are clearly present to conduct the proposed analyses. Still, if the data are *not* missing randomly as theorized, the estimates obtained will be biased.

Table 4.1 below presents descriptive statistics for all relevant variables employed in model development for the 2,309 students with complete data. At grade 6, 999 students are nested within 41 classrooms, for an average of just over 24 students per class. At grade 8, 1,310 students are nested within 52 classrooms, for an average of just over 25 students per class. Studies suggesting rules of thumb for minimum sample sizes in multilevel modeling include at least 30 observations per group and at least 30 groups (Hox, 2002, as cited in Bickel, 2007), and, more recently, at least 50 groups (Maas & Hox, 2005, as cited in Bickel, 2007). Given these guidelines, the average number of students per class may be some cause for concern in terms of unbiased standard error estimates as neither grade 6 nor grade 8 classrooms average more than 30 students. Both grades exceed the minimum class number threshold of 30, while grade 8 alone exceeds the threshold of 50.

Table 4.1. Descriptive statistics, final sample

L1 Variables	N	Mean	SD	Min	Max
fascination	6,883	2.51	0.52	1.00	4.00
grade 6	3,043	2.58	0.54	1.00	4.00
grade 8	3,840	2.46	0.49	1.00	4.00
L2 Variables					
HR	2,309	3.33	0.55	1.00	4.00
grade 6	999	3.28	0.56	1.14	4.00
grade 8	1,310	3.37	0.53	1.00	4.00
FS	2,309	3.44	0.54	1.00	4.00
grade 6	999	3.51	0.49	1.00	4.00
grade 8	1,310	3.38	0.56	1.00	4.00
URM	2,309	0.41	0.49	0.00	1.00
grade 6	999	0.44	0.50	0.00	1.00
grade 8	1,310	0.39	0.49	0.00	1.00
female	2,309	0.50	0.50	0.00	1.00
grade 6	999	0.51	0.50	0.00	1.00
grade 8	1,310	0.49	0.50	0.00	1.00
L3 Variables					
student-centric	93	0.55	0.17	0.18	0.93
grade 6	41	0.50	0.15	0.22	0.89
grade 8	52	0.59	0.18	0.18	0.93
hands-on	93	0.35	0.20	0.00	0.66
grade 6	41	0.29	0.16	0.00	0.56
grade 8	52	0.40	0.21	0.00	0.66
class-tech	93	0.05	0.09	0.00	0.32
grade 6	41	0.09	0.11	0.00	0.32
grade 8	52	0.03	0.06	0.00	0.24

The sole time-varying measure at level 1 is *fascination*, with technical details provided earlier in Chapter 3. Fascination scores were calculated as the mean of eight items as displayed in full in Appendix A1, with statements concerning interest, curiosity, and mastery goals pertaining to science and response options ranging from 1 to 4 measured five times over the course of the two-year study (see Figure 3.1). Visible above, sixth graders had a slightly higher mean fascination mean scores (i.e., the average across all testing occasions) (2.58) compared to their eighth grade counterparts (2.46). While all 2,309 students had at least one observed fascination score, only 336 (15% of the overall sample) had all five fascination scores; sixth graders had on average 3.05 observed scores, while eighth graders had on average 2.93. The means for each measurement occasion are presented in Table 4.2, and plotted in Figure 4.1. Notably, the graphical display of these means clearly depicts a pattern of change inconsistent with strictly linear growth, with the lowest scores observed at T3 nearly indistinguishable between grade levels.

Table 4.2. Fascination means at each measurement occasion

Fascination	N	Mean	SD	Min	Max
T1	2161	2.66	0.58	1.00	4.00
grade 6	934	2.83	0.56	1.00	4.00
grade 8	1227	2.53	0.55	1.00	4.00
T2	1901	2.57	0.57	1.00	4.00
grade 6	797	2.63	0.60	1.00	4.00
grade 8	1104	2.52	0.54	1.00	4.00
T3	1861	2.30	0.32	1.38	3.75
grade 6	771	2.31	0.33	1.38	3.29
grade 8	1090	2.30	0.31	1.50	3.75
T4	489	2.47	0.43	1.38	4.00
grade 6	280	2.47	0.43	1.38	4.00
grade 8	209	2.47	0.43	1.50	3.88
T5	471	2.48	0.45	1.00	4.00
grade 6	261	2.43	0.42	1.00	4.00
grade 8	210	2.55	0.49	1.38	4.00

Figure 4.1. Fascination change over time, grades 6 & 8

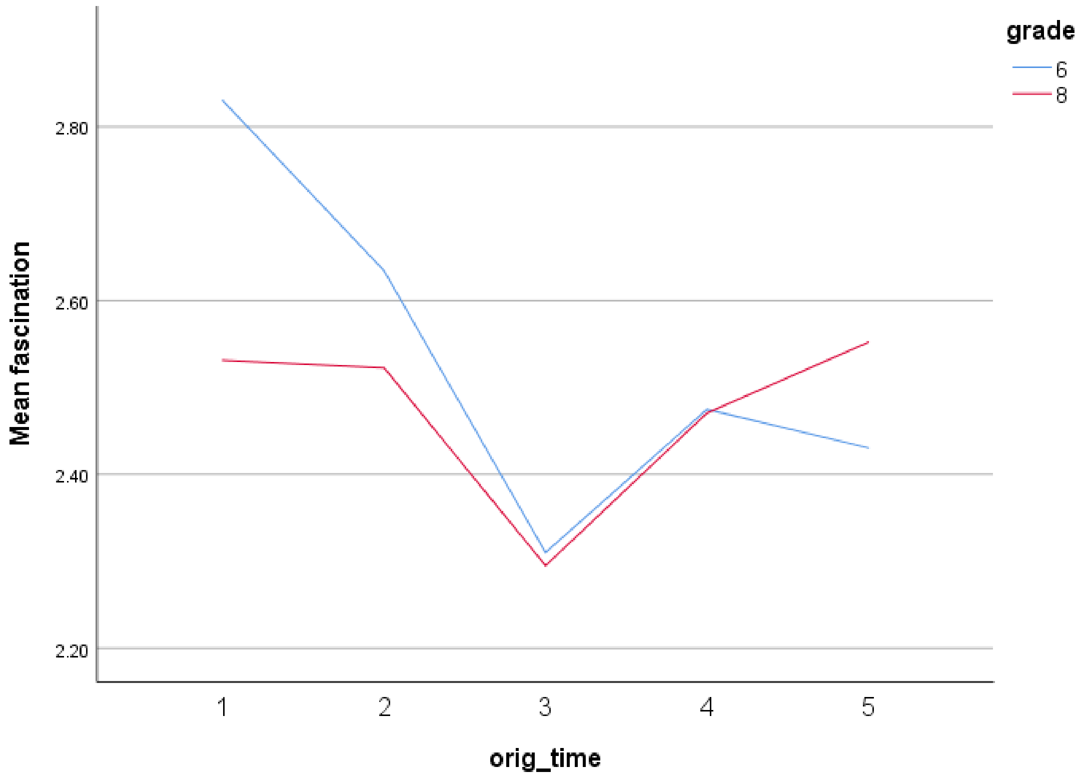


Table 4.2 above shows a distinct pattern of attrition: far more fascination scores are observed in Year 1 (T1/T2/T3) than in Year 2 (T4/T5) for both grade levels. Due to the nature of recruiting teachers for study participation and including full classes of students in Year 1, complexity in data collection arose in Year 2 as these full classes of students dispersed into multiple classes the following year. (This is also why the observed instructional variables are also present for year 1 only.) Students were asked to come during lunch or after school in Year 2 of the study to take the various Activation measures, a far more complicated request of middle and high school students than taking class time for participation. Exactly 50% of students across both grade levels had fully observed fascination scores in Year 1 and no observed scores in Year 2, while only 24% of the sample had any

recorded Year 2 fascination score. Of the 550 students who recorded fascination scores in Year 2, 410 (75%) contributed scores at both T4 and T5. Appendix D1 offers a comparison of average fascination for Y1-only students compared to Y1&2 students.

Table 4.3 presents the coding of the time variable used in modeling. Noted in the previous chapter, the time variable was "rough-centered" at the second measurement occasion such that T2=0, effectively allowing the instructional effects modeled at level 3 to be observed over all five time-points, since it is unlikely that such effects are observable at the start of start of the academic year (T1). As non-linear models were theorized and capable of being tested given the five time-points, quadratic and cubic terms are also shown. The remaining two variables, *grade* and *post*, are included for testing discontinuous functional forms, with the dashed line indicating the break between school years and the hypothesized point of discontinuity (between T3 and T4, the spring of Y1 and the fall of Y2). *Grade* is a dummy variable, included in a linear model, to represent the magnitude of discontinuity between Years 1 and 2—a change in level—as represented in Figure 3.3 in the previous chapter. *Post* is another time-varying *time* variable that posits a meaningful change in slope following the point of discontinuity, represented in Figure 3.4 in the previous chapter. Including both variables in a linear model would test the functional form represented in Figure 3.5 in the previous chapter: a change in both level and slope at the point of discontinuity.

Table 4.3. Time-point coding for model development

Measurement occasion	School year	*time*	*time*^{2}	*time*^{3}	*grade*	*post*
1	Y1 (fall)	-1	1	-1	0	0
2	Y1 (winter)	0	0	0	0	0
3	Y1 (spring)	1	1	1	0	0
4	Y2 (fall)	2	4	8	1	1
5	Y2 (spring)	3	9	27	1	2

Four level 2 variables were included in modeling: home resources (*HR*), family support for science (*FS*), race, coded as under-represented minority (*URM*), and gender, included as a dummy variable for girls (*female*), each of which was asked on the Background Questionnaire administered at the start of Years 1 and 2 (T1 and T4). The measure for home resources was obtained by asking the extent to which various supports for learning were available at home using seven items and response options from 1 ("Never") to 4 ("Always"), with higher values indicating more availability; the mean across the seven responses was included as the student measure of home resources. As shown in Table 4.1, both grade 6 and grade 8 students had relatively high mean scores for *HR*, 3.28 and 3.37, respectively, indicating that home resources for learning were available somewhere between "most of the time" and "always," on average. The full set of items used for measuring home resources is presented in Appendix A2.

Similarly, family support for science was measured using five items on the Background Questionnaire asking students' perceptions of the extent to which someone in their family supported science learning, with response options also ranging from 1 ("NO!") to 4 ("YES!"). The mean across the five items was used as the measure for family support for science, and once again, students in both grades had relatively high scores for *FS*, with grade 6 students reporting slightly higher home support on average (3.51 compared to 3.38 at grade 8). The full set of items used for measuring family support for science is presented in Appendix A3.

Race was measured on the Background Questionnaire by asking, "Which of the following describes your racial/ethnic background? You may choose more than one." Response categories included White, Black or African American, Asian, Indian/Middle Eastern, Native American/Pacific Islander, Hispanic/Latino/Mexican, I don't know, and Other (please specify). Students who omitted a response to this question or selected "I don't know" to the Background Questionnaire were considered to have missing racial identity data and were not included in analyses, as previously discussed with regard to missing data. Following the justification provided in Chapter 3 regarding the underrepresentation of people who identify as Hispanic, Black, Native American or Pacific Islander in science and STEM professions, the race data were coded into the binary variable of underrepresented minority (*URM*) using the definition provided by the National Science Foundation. "This category comprises three racial or ethnic minority groups ([people membered] blacks, Hispanics, and American Indians or Alaska Natives) whose representation in S&E [science & engineering] education or employment is smaller than their representation in the U.S. population." (National Science Foundation, 2021b). As seen in Table 4.1, across both 6th and 8th graders, 41% identified with a race/ethnicity that membered them as *URM*, with slightly more 6th (44%) than 8th graders (39%) falling into this category.

Gender was measured on the Background Questionnaire by asking, “Are you a girl or a boy?” with three possible response options: girl, boy, or prefer not to answer. Once again acknowledging possible non-binary gender identities, the Background Questionnaire lacked such response categories and therefore gender was treated as binary, with those students selecting “Prefer not to answer” as having missing gender data. For model development, the gender variable was recoded into *female*, a dummy variable with those who selected boy as the base case (0) and those who selected girl as equal to 1. As seen in Table 4.1, half of all students included in analyses identified as girls.

Lastly, three level 3 variables were included in model development: student-centric teaching (*student-centric*), hands-on methods (*hands-on*), and classroom technology use (*class-tech*). As discussed in the previous chapter, these measures were obtained from self-reported instructional logs over the course of Year 1 only. Each measure represents the average proportion of time a teacher spent employing one of three strategies—student-centric teaching, hands-on methods, and classroom technology—within a given class. Worth noting, student-centric teaching and hands-on methods were two of three instructional categories, the third being textbook teaching (omitted from modeling), and as such are mutually exclusive; these three measures sum to 100% within classes. Table 4.1 shows that, on average, student-centric teaching occurred 55% of the time (slightly more often in grade 8 than grade 6), while hands-on teaching occurred 35% of the time on average (again, slightly more often in grade 8 than grade 6), implying that textbook teaching occurred roughly 10% of the time, on average. Classroom technology use was independent of both student-centric and hands-on teaching methods. Table 4.1 illustrates that classroom technology was employed relatively infrequently, only 5% of instructional time on average, although was more prevalent in 6^{th} grade (9%) than in 8^{th} grade (3%). Appendix A4 displays the Teacher Log used to measure these instructional techniques.

Model Building

The next sections present each stage of the model building process as outlined in the “Developing the Models” section of Chapter 3, referencing previous equations as those for grades 6 and 8 are developed. Specifically, the null model is first developed, in which unconditional linear growth serves as the starting model to which all subsequent models are compared. This enables a determination about the extent of nesting and an evaluation of the variance components at each of the three levels. Non-linear terms are then added and evaluated at level 1, while keeping levels 2 and 3 unconditional. Next, discontinuous models are tested and compared to prior models in order to determine the functional form at L1 from which to proceed. After determining the final growth model specifications—linear, non-linear, or discontinuous—level 2 is specified, followed by level 3.

In each phase of model building, backward elimination variable selection is employed, in which all relevant predictors for a given level are entered simultaneously and evaluated for significance, with non-significant terms removed. This is particularly important for the specification of functional form at level 1, as each higher order polynomial depends on the term below it (e.g., a quadratic function must include both quadratic and linear terms to be appropriately specified).

Although the time-varying terms at L1 are rough-centered with a meaningful value of 0 as displayed in Table 4.3, all L2 and L3 variables have been grand-mean centered to reflect dispersion around the level 3 mean. This decision was made with the recognition that L3 terms can only be grand-mean centered, and two of the four L2 variables are dichotomous and therefore need not be centered at all. Therefore, the only two variables for which this decision has any impact are home resources (*HR*) and family support for science (*FS*). In the interest of keeping results straightforward with respect to interpretation of terms, grand-mean centering holds the most appeal. In equation writing, grand-mean centering is most commonly denoted as $(X_{ij} - \bar{X}_{\bullet\bullet})$; an individual's observed value minus the grand-mean. To simplify notation moving forward, grand-mean centered variables will be **bolded** to denote the centering.

Level 1: Functional Form

Referencing the mixed model in equation 3.4 and inserting *time* as the lone L1 term specifying linear growth produces the null model, shown in equation 4.1:

$$fascination_{tij} = \gamma_{000} + \gamma_{100}time_{tij} + \\ r_{0ij} + r_{1ij}time_{tij} + \mu_{00j} + \mu_{10j}time_{tij} + e_{tij} \quad (4.1)$$

This unconditional model sets up the baseline from which model comparisons are drawn. Building upon this to include non-linear growth as shown for L1 in equation

3.10’ and the mixed model in 3.13, we obtain equation 4.2 including the quadratic and cubic terms:

$$fascination_{tij} = \gamma_{000} + \gamma_{100}time_{tij} + \gamma_{200}time_{tij}^2 + \gamma_{300}time_{tij}^3 + r_{0ij} + r_{1ij}time_{tij} + r_{2ij}time_{tij}^2 + r_{3ij}time_{tij}^3 + \mu_{00j} + \mu_{10j}time_{tij} + \mu_{20j}time_{tij}^2 + \mu_{30j}time_{tij}^3 + e_{tij} \quad (4.2)$$

From here, we can begin to refine the model to find the one that is best fitting/most parsimonious by removing all non-significant fixed and random effects.

Finally, we re-envision the functional form, not as a polynomial equation with quadratic and cubic terms, but instead with discontinuity terms specifying potential changes in level and slope, as coded in Table 4.3. Still, the resulting mixed model strongly resembles the non-linear equation of 4.2, as there are once again three terms at L1—*time*, *grade*, and *post*—each of which are subsequently modeled with no predictors and an error term at L2 and L3. Equation 4.3 shows the beginning discontinuous model, testing change in both level and slope after the point of discontinuity:

$$fascination_{tij} = \gamma_{000} + \gamma_{100}time_{tij} + \gamma_{200}grade_{tij} + \gamma_{300}post_{tij} + r_{0ij} + r_{1ij}time_{tij} + r_{2ij}grade_{tij} + r_{3ij}post_{tij} + \mu_{00j} + \mu_{10j}time_{tij} + \mu_{20j}grade_{tij} + \mu_{30j}post_{tij} + e_{tij} \quad (4.3)$$

Similarly, from here the model can be refined by removing non-significant effects, possibly reducing the discontinuity from a change in both slope and level to a change in either slope or level, or removing non-significant random effects.

Table 4.4 displays the results of the models displayed in equations 4.1, 4.2, and 4.3 for grade 6, and in the case of models 4.2' and 4.3', the final non-linear and discontinuous model after iterative refinement (while noting that one to two models were run before finalizing that are not displayed in the table). Table 4.5 displays the results of the same models for grade 8 and will be discussed separately.

Grade 6 Functional Form

Beginning with the null model 4.1 in Table 4.4, there are several things to note as the model develops, the first of which are the calculations of the intraclass correlation coefficients for levels 2 and 3. Using equations 3.5 and 3.6 and substituting the estimated variance components for their respective parameters:

L2 ICC: $\hat{\rho} = \frac{\hat{\tau}_{00}}{\hat{\sigma}^2+\hat{\tau}_{00}+\hat{\tau}_{000}} = \frac{.122}{.134+.122+.004} = 0.468$

L3 ICC: $\hat{\rho} = \frac{\hat{\tau}_{000}}{\hat{\sigma}^2+\hat{\tau}_{00}+\hat{\tau}_{000}} = \frac{.004}{.134+.122+.004} = 0.015$

These results indicate that about 47% of the variance in fascination scores is between sixth-grade students at L2, while less than 2% of the variance is between sixth-grade classrooms. Furthermore, the remaining variance (1-.468-.015) leaves about 52% of the variance in fascination within students. In other words, more than half of the variance in observed fascination scores can be attributed to within-person time varying factors influencing fascination that have not been accounted for.

Looking at preliminary model fit statistics to which subsequent models will be compared, we see that the deviance statistic is 4019.5 with nine estimated parameters, the Akaike Information Criterion (AIC) (equal to the deviance statistic plus twice the number of parameters as "penalization") is 4037.5, and the model took 17 iterations to converge. Raudenbush and Bryk acknowledge that the number of iterations required for model estimates to converge can often be used as a diagnostic reference: "When the number of iterations required is large…this indicates that the estimation is moving toward a boundary condition" (i.e., the variance estimate for τ is approaching zero) (Raudenbush et al., 2019, p. 200). Lastly, all fixed and random effect estimates were found to be statistically significant. This indicates that intercept, representing the expected fascination score of a sixth grade student at time 0, is significantly different from zero, that the linear time component significantly contributes to the explanation of change in student fascination scores, and that these two components (intercept and linear change over time) significantly vary between students and classrooms, all else constant.

Model 4.2 of Table 4.4 presents the results after including the non-linear $time^2$ and $time^3$ terms at level 1, while allowing the intercepts and slopes for both terms to vary randomly across L2 and L3 units. Initial model fit has improved—the AIC has been reduced considerably to 3743.8, and the deviance statistic is statistically significantly different from model 4.1 (χ^2=325.7, *df*=16, *p*<.001)—although the number of iterations required for convergence was quite large, 3,896. Of course, "if the added complexity better models the observed data, it is likely worth the increased computational demands, as we will obtain a better representation of the relations among the data" (Anderson, 2012, p. 54).

Model 4.2' represents the refinement of model 4.2 after removing the level 2 random effects (r_2 and r_3) associated with the non-linear terms, which were observed to be non-significant, while retaining the non-significant fixed effect associated with the quadratic term *time*^{2} in order to ensure the cubic effect was fully specified by including all lower level polynomial terms in the model. The removal of these non-significant terms created a more parsimonious model with reduced complexity, as only 18 parameters were estimated compared to the prior 25, and the number of iterations required for convergence dropped to 965—still a large number, but far fewer than the 3,896 for model 4.2. Though both the deviance statistic and AIC (3803.7 and 3839.7, respectively) reflect a significant improvement from model 4.1, comparisons show that the fit statistics are higher for model 4.2, indicating that model 4.2' may be a poorer fit for the data than 4.2.

Removing the non-linear terms introduced in model 4.2 to evaluate a potential discontinuity between school years, model 4.3 retains the initial linear *time* component and adds the variables *grade* and *post* at L1 to see if there is a change in level or slope as 6th grade students enter 7th grade. Levels 2 and 3 once again are unconditional, with random effects entered at both levels for all level 1 terms. As displayed in Table 4.4, preliminary results for model 4.3 indicate a better fit than the unconditional model 4.1, with a significant decrease in the deviance statistic (χ^2=340.6, *df*=16, *p*<.001) and lower AIC (3728.9). The number of iterations required for convergence was quite high (3,421) and several of the random effects were determined to be non-significant, indicating room for improvement in model specification.

Model 4.3' shows the final discontinuous model, having removed the random effects associated with both discontinuous terms at L2 (r_2 and r_3) and the random effect associated with *post* at L3 (u_{30}). Similar to the pattern of model refinement that occurred moving from model 4.2 to 4.2', the fit statistics have once again slightly increased from the prior model 4.3. However, model 4.3' remains a significant improvement from the null model when comparing the deviance statistic of 3738.4 (χ^2=281.1, *df*=5, *p*<.001) and has a lower AIC of 3766.4. Furthermore, these same statistics can be used to compare the fit between the final non-linear and discontinuous models. Compared to model 4.2', model 4.3' has a significantly lower deviance statistic (χ^2=65.3, *df*=4, *p*<.001), a lower AIC (3766.4 compared to 3839.7), and took fewer iterations to converge. The discontinuous model also doesn't retain any statistically non-significant fixed effects for proper specification as is the case in the non-linear model, and all remaining random effects were found to be significant.

Table 4.4. Grade 6 level 1 model development - functional form

Fixed effects		Model 4.1		Model 4.2		Model 4.2’		Model 4.3		Model 4.3’	
		Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Intercept	γ_{000}	2.62	0.017	2.60	0.026	2.60	0.026	2.60	0.017	2.60	0.017
time	γ_{100}	-0.17	0.015	-0.25	0.015	-0.25	0.015	-0.26	0.013	-0.26	0.013
time²	γ_{200}	--	--	0.02*	0.022	0.02*	0.022	--	--	--	--
time³	γ_{300}	--	--	0.02	0.007	0.02	0.007	--	--	--	--
grade	γ_{200}	--	--	--	--	--	--	0.21	0.056	0.20	0.058
post	γ_{300}	--	--	--	--	--	--	0.21	0.026	0.22	0.028
Random effect estimates											
L1, w/in-student	e_{tij}	0.134		0.097		0.119		0.114		0.117	
L2 intercept	r_{0ij}	0.122		0.207		0.126		0.117		0.126	
L2 slope – *time*	r_{1ij}	0.015		0.039		0.016		0.025		0.016	
L2 slope – *time²*	r_{2ij}	--		0.036		*removed*		--		--	
L2 slope – *time³*	r_{3ij}	--		0.004 *		*removed*		--		--	
L2 slope – *grade*	r_{2ij}	--		--		--		0.021 *		*removed*	
L2 slope – *post*	r_{3ij}	--		--		--		0.018 *		*removed*	
L3 intercept	u_{00j}	0.004		0.015		0.016		0.005		0.005	
L3 slope – *time*	u_{10j}	0.006		0.005		0.005		0.003		0.002	
L3 slope – *time²*	u_{20j}	--		0.013		0.012		--		--	
L3 slope – *time³*	u_{30j}	--		0.001		0.001		--		--	
L3 slope – *grade*	u_{20j}	--		--		--		0.007		0.046	
L3 slope – *post*	u_{30j}	--		--		--		0.052		*removed*	
Deviance (parameters)		4019.5 (9)		3693.8 (25)		3803.7 (18)		3678.9 (25)		3738.4 (14)	
Chi-square (df)		--		325.7 (16)		215.8 (9)		340.6 (16)		281.1 (5)	
AIC		4037.5		3743.8		3839.7		3728.9		3766.4	
Iterations		17		3,896		965		3,421		868	
Reliability estimates											
L1 intercept	π_0	0.70		0.71		0.73		0.73		0.73	
L1 – *time*	π_1	0.27		0.45		0.30		0.29		0.31	
L1 – *time²*	π_2	--		0.31		*removed*		--		--	
L1 – *time³*	π_3	--		0.31		*removed*		--		--	
L1 – *grade*	π_2	--		--		--		0.03		*removed*	
L1 – *post*	π_3	--		--		--		0.06		*removed*	
L2 intercept	β_{00}	0.35		0.54		0.61		0.45		0.41	
L2 – *time*	β_{10}	0.71		0.56		0.58		0.42		0.39	
L2 – *time²*	β_{20}	--		0.68		0.71		--		--	
L2 – *time³*	β_{30}	--		0.64		0.65		--		--	
L2 – *grade*	β_{20}	--		--		--		0.45		0.63	
L2 – *post*	β_{30}	--		--		--		0.22		*removed*	
Pseudo R^2 calculations											
L1		--		0.281		0.111		0.152		0.126	
L2		--		-0.706		-0.036		0.034		-0.036	
L3		--		-2.768		-3.202		-0.356		-0.253	

* Coefficient is *not* significant, p>.05; all other values significant.
“--” indicates fixed/random effect was not entered into the model.
NOTE: Chi-square statistics compare models to the unconditional model 4.1.

Finally, referencing equations 3.7, 3.8, and 3.9, we can compare the pseudo R^2 statistics for each model in relation to the unconditional model specified in 4.1. As presented at the bottom of Table 4.4, the pseudo R^2 statistics for levels 2 and 3 are mostly negative, perhaps unsurprising given that variables have not yet been added at those levels to explain variance; negative values are largely uninterpretable and therefore should not be used (Holden, Kelley, & Agarwal, 2008). Comparing the pseudo R^2 statistics for level 1, however, we see that model 4.2' results in a reduction of 11% in the unexplained L1 variance, e_{ijk}, while model 4.3' reduces this variance by 13%. Although a relatively small difference of less than 2%, this implies that model 4.3' accounts for more of the previously unexplained variance in fascination scores at level 1 through the incorporation of discontinuous terms when compared to the non-linear model.

For those reasons—lower deviance and AIC fit statistics, more parsimony (14 parameters estimated), all remaining terms significant, and a higher level 1 pseudo R^2 statistic—the final functional form that will be used to continue model development for grade 6 is the discontinuous model presented in equation 4.3':

$$fascination_{tij} = \gamma_{000} + \gamma_{100}time_{tij} + \gamma_{200}grade_{tij} + \gamma_{300}post_{tij} + r_{0ij} + r_{1ij}time_{tij} + \mu_{00j} + \mu_{10j}time_{tij} + \mu_{20j}grade_{tij} + e_{tij} \qquad (4.3')$$

Fully interpreting the results of this model are contingent upon building up levels 2 and 3. It is worth highlighting, however, that in Table 4.4 all of the random effects were estimated with sufficient reliability (>0.10), indicating that the precision of the estimates for the randomly varying slopes and intercepts (π_0, π_1, β_{00}, β_{10}, and β_{20}) is not a cause for concern. Preliminary estimates of the fixed effects at the top of the table suggest that, on average, the initial fascination score for all students is 2.60 midway through 6th grade (*time*=0), down from 2.86 at the start of the school year. By the end of the school year, average fascination scores decrease further to 2.34 (out of 4.00). Holding all else constant, fascination scores "bounce back" by the beginning of 7th grade, as the magnitude of discontinuity (*grade*) predicts a starting point of 2.50 to begin the following school year; the change in slope (*post*) indicates that fascination scores decrease once again by the end of the year to 2.46.

Grade 8 Functional Form

Following the same procedure that was used to decide upon the L1 functional form for grade 6, Table 4.5 presents the results of the unconditional linear model, two versions of the non-linear model, and two versions of the discontinuous model for grade 8. To simplify the discussion, this section will only focus on the null model briefly, followed by a comparison of the final two models, 4.2'' and 4.3'', to evaluate model fit and decide on the grade 8 functional form.

Calculating the intraclass correlation coefficients for levels 2 and 3 using the estimated variance components of the unconditional model 4.1 for grade 8 produces the following:

L2 ICC: $\hat{\rho} = \frac{\hat{\tau}_{00}}{\hat{\sigma}^2+\hat{\tau}_{00}+\hat{\tau}_{000}} = \frac{.105}{.111+.105+.005} = 0.477$

L3 ICC: $\hat{\rho} = \frac{\hat{\tau}_{000}}{\hat{\sigma}^2+\hat{\tau}_{00}+\hat{\tau}_{000}} = \frac{.005}{.111+.105+.005} = 0.021$

These results are extremely similar to the null model run for grade 6, and indicate that about 48% of the variance in fascination scores is between eighth graders at L2, while only 2% of the variance is between eighth-grade classrooms. This leaves roughly 50% of the variance in fascination within-students; that is, unexplained within-person time varying factors influencing fascination scores that are not modeled.

Both models 4.2'' and 4.3'' have fit statistics that indicate significant improvement over the unconditional model 4.1, and following their refinement from the starting non-linear and discontinuous models, have fixed and random effects terms that are all statistically significant. In comparing the two models to one another, it initially appears that the non-linear model is a better fit for the data. The deviance statistic for model 4.2'' is significantly smaller than that for model 4.3'' (χ^2=158.7, df=8, p<.001) and the AIC is also smaller in comparison (4296.1 vs. 4438.8). The pseudo R^2 statistics also support the non-linear model over the discontinuous model, as the L1 variance reduction in model 4.2'' was calculated to be 17%, whereas that of model 4.3'' was only 4%. (Once again, the L2 and L3 pseudo R^2 statistics were found to be negative and are ignored pending further model development.) Despite this, model 4.3'' reached convergence with far fewer iterations of the estimates—1,812 fewer!—and is more parsimonious, with only 10 parameters estimated compared to the 18 in model 4.2''.

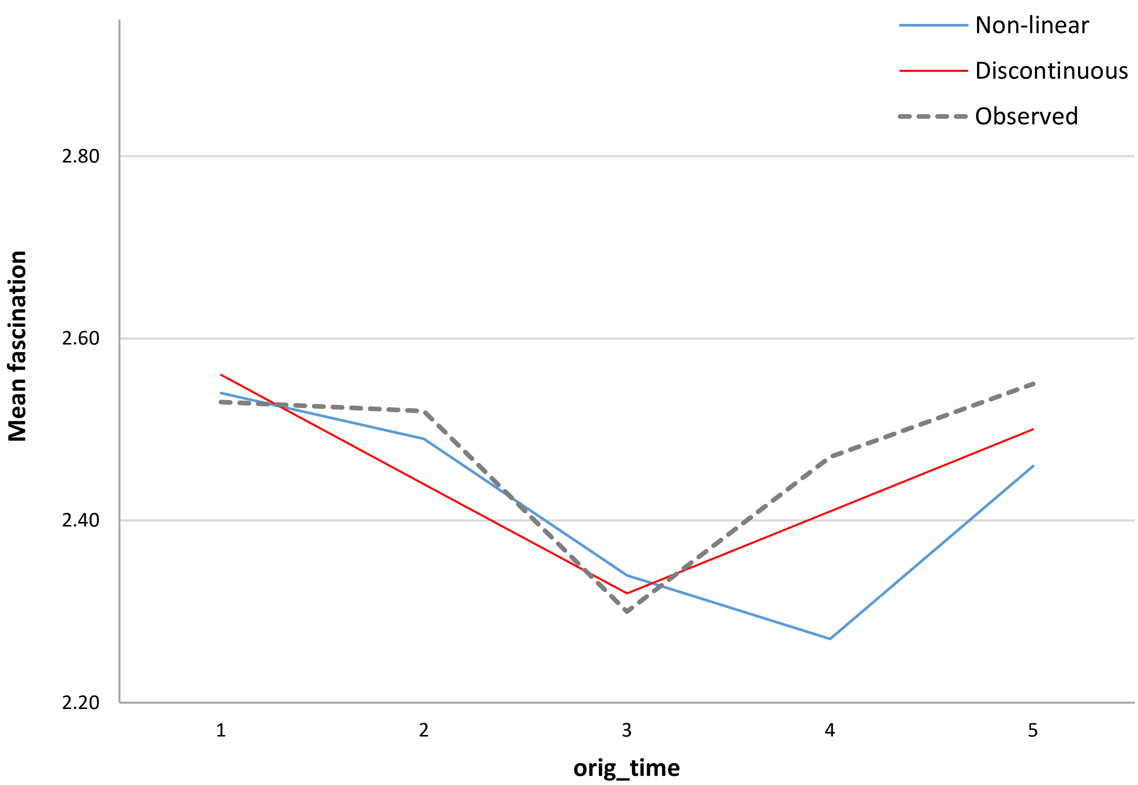
Table 4.5. Grade 8 level 1 model development - functional form

Fixed effects		Model 4.1 Coeff.	Model 4.1 SE	Model 4.2 Coeff.	Model 4.2 SE	Model 4.2'' Coeff.	Model 4.2'' SE	Model 4.3 Coeff.	Model 4.3 SE	Model 4.3'' Coeff.	Model 4.3'' SE
Intercept	γ_{000}	2.46	0.014	2.49	0.017	2.49	0.017	2.44	0.014	2.44	0.014
time	γ_{100}	-0.06	0.011	-0.14	0.014	-0.13	0.013	-0.11	0.013	-0.12	0.013
time²	γ_{200}	--	--	-0.05	0.011	-0.05	0.011	--	--	--	--
time³	γ_{300}	--	--	0.04	0.005	0.03	0.004	--	--	--	--
grade	γ_{200}	--	--	--	--	--	--	0.16	0.052	*removed*	
post	γ_{300}	--	--	--	--	--	--	0.17	0.026	0.26	0.034
Random effect estimates											
L1, w/in-student	e_{tij}	0.111		0.089		0.092		0.102		0.106	
L2 intercept	r_{0ij}	0.105		0.170		0.168		0.099		0.107	
L2 slope – *time*	r_{1ij}	0.024		0.049		0.048		0.030		0.021	
L2 slope – *time²*	r_{2ij}	--		0.027		0.025		--		--	
L2 slope – *time³*	r_{3ij}	--		0.004		0.004		--		--	
L2 slope – *grade*	r_{2ij}	--		--		--		0.003 *		*removed*	
L2 slope – *post*	r_{3ij}	--		--		--		0.031 *		*removed*	
L3 intercept	u_{00j}	0.005		0.005		0.005		0.005		0.005	
L3 slope – *time*	u_{10j}	0.003		0.005		0.004		0.005		0.004	
L3 slope – *time²*	u_{20j}	--		0.001 *		*removed*		--		--	
L3 slope – *time³*	u_{30j}	--		< 0.001 *		*removed*		--		--	
L3 slope – *grade*	u_{20j}	--		--		--		0.045 *		*removed*	
L3 slope – *post*	u_{30j}	--		--		--		0.005 *		*removed*	
Deviance (parameters)		4606.9 (9)		4237.9 (25)		4260.1 (18)		4250.2 (25)		4418.8 (10)	
Chi-square (df)		--		369.03 (16)		346.77 (9)		356.68 (16)		188.18 (1)	
AIC		4624.9		4287.9		4296.1		4300.2		4438.8	
Iterations		23		5046		1836		6199		24	
Reliability estimates											
L1 intercept	π_0	0.71		0.67		0.66		0.72		0.72	
L1 – *time*	π_1	0.35		0.49		0.48		0.36		0.33	
L1 – *time²*	π_2	--		0.25		0.23		--		--	
L1 – *time³*	π_3	--		0.30		0.28		--		--	
L1 – *grade*	π_2	--		--		--		0.01		*removed*	
L1 – *post*	π_3	--		--		--		0.11		*removed*	
L2 intercept	β_{00}	0.43		0.31		0.48		0.45		0.45	
L2 – *time*	β_{10}	0.55		0.49		0.62		0.56		0.63	
L2 – *time²*	β_{20}	--		0.11		*removed*		--		--	
L2 – *time³*	β_{30}	--		0.23		*removed*		--		--	
L2 – *grade*	β_{20}	--		--		--		0.26		*removed*	
L2 – *post*	β_{30}	--		--		--		0.10		*removed*	
Pseudo R^2 calculations											
L1		--		0.194		0.170		0.078		0.041	
L2		--		-0.612		-0.597		0.062		-0.014	
L3		--		-0.088		-0.056		0.004		-0.047	

* Coefficient is *not* significant, p>.05; all other values significant.
"--" indicates fixed/random effect was not entered into the model.
NOTE: Chi-square statistics compare models to the unconditional model 4.1.

Figure 4.2 below plots the suggested fascination means for each measurement occasion at grade 8 based on non-linear model 4.2'', discontinuous model 4.3'', and the observed fascination scores as displayed in Table 4.2. Though the non-linear model does appear to match the pattern of change in fascination scores particularly well in Year 1 (time points 1-3), there is a serious departure from the observed pattern in Year 2 at time points 4 and 5. Meanwhile, the discontinuous model displays a decreasing linear pattern in Year 1 that is less aligned with the observed data particularly at time point 2 mid-way through grade 8, but the Year 2 pattern appears to be more similar to the observed change in Year 2 than the non-linear prediction.

Figure 4.2. Grade 8 models 4.2'' and 4.3'' compared to observed mean fascination



Admittedly, the model comparison seems to suggest the non-linear model is a better fit for the grade 8 data. Parsimony is preferred when the model statistics suggest a good fit, so it is worth highlighting that model 4.3'' does fit the data significantly better than the null model (deviance comparison, χ^2=188.2, *df*=1, *p*<.001, AIC 186.1 lower). Because there is no single statistic or rule that determines the decision when it comes to model development and some amount of subjectivity is involved, the discontinuous model will serve as the final functional form for grade 8, for several reasons. First, the significant random effects found in the non-linear model at L2 (r_2 and r_3) suggest that, for each student, the rate of deceleration (*time*^{2}) and change in this rate of change (*time*^{3}) varies for each student. Statistically, this seems to be compatible with model 4.2'' based on significance, but practically, modeling this randomness adds complexity that does not seem worthwhile. Second, theory does not suggest that a continuous (in this case, non-linear) growth model should be expected for change in fascination, and a discontinuous model does make sense given the gap between grades 8 and 9. Finally, keeping the grade 6 and grade 8 models similar in their functional form (discontinuous) assists the forthcoming analysis in the interpretation of terms and estimates. Resultantly, the final functional form that will be used to continue model development for grade 8 is the discontinuous model presented in equation 4.3'':

$$fascination_{tij} = \gamma_{000} + \gamma_{100}time_{tij} + \gamma_{200}post_{tij} + r_{0ij} + r_{1ij}time_{tij} + \mu_{00j} + \mu_{10j}time_{tij} + e_{tij} \qquad (4.3'')$$

Compared to the grade 6 functional form identified in equation 4.3', the grade 8 functional form importantly differs in that the change in level variable (*grade*) was not found to have significant fixed or random effects and therefore is not modeled (such that the discontinuity present is only in the change in slope, represented with the variable *post*).

Level 2: Student-level Covariates

Now that the discontinuous functional forms at level 1 have been decided upon, simultaneous entry of the student-level variables *HR, FS, URM,* and *female* for each level 2 equation allows us to build up the models, dramatically increasing the number of level 3 equations and thus the complexity. Similar to the preceding discussion of grade 8 models, the next sections detailing level 2 development, first for grade 6 and then for grade 8, will focus on the starting models employing backwards elimination of relevant predictor variables and then a comparison of the final level 2 model selected.

Grade 6 Level 2 Model Development

Equations 4.4a-y show the level 1, 2, and 3 models for grade 6. Building from model 4.3', we see that level 2 equations 4.4d and 4.4e lack their respective error terms (r_2 and r_3), as the level 1 model development process informed us that there was no random variation in the Year 2 effects (*grade*, *post*) between students. Similarly, equation 4.4u lacks an error term (μ_{30}), as we earlier saw there was no significant variation in the change in slope between 6th and 7th grades between 6th

grade classrooms at level 3. However, all other L3 models have a random error component, testing to see if there is significant variation between classrooms for each of the L2 predictors entered and their effect on the growth parameters *time*, *grade*, and *post*.

Level 1:

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}time_{tij} + \pi_{2ij}grade_{tij} + \pi_{3ij}post_{tij} + e_{tij} \tag{4.4a}$$

Level 2:

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}\boldsymbol{HR}_{ij} + \beta_{02j}\boldsymbol{FS}_{ij} + \beta_{03j}URM_{ij} + \beta_{04j}female_{ij} + r_{0ij} \tag{4.4b}$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}\boldsymbol{HR}_{ij} + \beta_{12j}\boldsymbol{FS}_{ij} + \beta_{13j}URM_{ij} + \beta_{14j}female_{ij} + r_{1ij} \tag{4.4c}$$

$$\pi_{2ij} = \beta_{20j} + \beta_{21j}\boldsymbol{HR}_{ij} + \beta_{22j}\boldsymbol{FS}_{ij} + \beta_{23j}URM_{ij} + \beta_{24j}female_{ij} \tag{4.4d}$$

$$\pi_{3ij} = \beta_{30j} + \beta_{31j}\boldsymbol{HR}_{ij} + \beta_{32j}\boldsymbol{FS}_{ij} + \beta_{33j}URM_{ij} + \beta_{34j}female_{ij} \tag{4.4e}$$

Level 3:

$$\beta_{00j} = \gamma_{000} + \mu_{00j} \tag{4.4f}$$

$$\beta_{01j} = \gamma_{010} + \mu_{01j} \tag{4.4g}$$

$$\beta_{02j} = \gamma_{020} + \mu_{02j} \tag{4.4h}$$

$$\beta_{03j} = \gamma_{030} + \mu_{03j} \tag{4.4i}$$

$$\beta_{04j} = \gamma_{040} + \mu_{04j} \tag{4.4j}$$

$$\beta_{10j} = \gamma_{100} + \mu_{10j} \tag{4.4k}$$

$$\beta_{11j} = \gamma_{110} + \mu_{11j} \tag{4.4l}$$

$$\beta_{12j} = \gamma_{120} + \mu_{12j} \tag{4.4m}$$

$$\beta_{13j} = \gamma_{130} + \mu_{13j} \tag{4.4n}$$

$$\beta_{14j} = \gamma_{140} + \mu_{14j} \tag{4.4o}$$

$$\beta_{20j} = \gamma_{200} + \mu_{20j} \tag{4.4p}$$

$$\beta_{21j} = \gamma_{210} + \mu_{21j} \tag{4.4q}$$

$$\beta_{22j} = \gamma_{220} + \mu_{22j} \tag{4.4r}$$

$$\beta_{23j} = \gamma_{230} + \mu_{23j} \quad (4.4s)$$

$$\beta_{24j} = \gamma_{240} + \mu_{24j} \quad (4.4t)$$

$$\beta_{30j} = \gamma_{300} \quad (4.4u)$$

$$\beta_{31j} = \gamma_{310} + \mu_{31j} \quad (4.4v)$$

$$\beta_{32j} = \gamma_{320} + \mu_{32j} \quad (4.4w)$$

$$\beta_{33j} = \gamma_{330} + \mu_{33j} \quad (4.4x)$$

$$\beta_{34j} = \gamma_{340} + \mu_{34j} \quad (4.4y)$$

Tables 4.6 and 4.7 below show the results of the grade 6 level 2 model development process for fixed and random effect estimates, respectively. Beginning with the model summary and fit statistics at the bottom of Table 4.6 for the initial model (4.4), convergence required 3,682 iterations due to the increased parameterization. Compared to the null model (4.1), the addition of level 2 predictors has improved overall model fit (χ^2=438.7 *df*=205 *p*<.001). A cursory glance at the fixed effects estimated in Table 4.6 show a number of non-significant terms; specifically, neither home resources (*HR*) nor under-represented minority status (URM) meaningfully explained fascinations scores in relation to any of the L1 components. Likewise, the dummy variable *female* was found to be non-significant for all L1 time-varying components but was significant in its relationship to the L1 intercept. The family support for science variable (*FS*), on the other hand, was found to be significant in its relationship to three of four L1 components, including the intercept, Year 1 slope variable (*time*, linear), and Year 2 slope variable (*post*). Looking at the estimates for model 4.4 in Table 4.7, we see that none of the random effects for any of these newly entered student-level variables was found to be significant; as well, the addition of these newly entered terms into the model caused the random effect associated with the L2 intercept for the *time* variable to also lose significance.

Table 4.6. Grade 6 level 2 model development - fixed effects

Fixed effects		Model 4.4		Model 4.4'	
		Coeff.	SE	Coeff.	SE
Intercept	β_{00}				
intercept	γ_{000}	2.64	0.026	2.65	0.019
HR	γ_{010}	0.05 *	0.031	--	--
FS	γ_{020}	0.14	0.034	0.16	0.029
URM	γ_{030}	0.04 *	0.025	--	--
female	γ_{040}	-0.14	0.027	-0.11	0.02
time	β_{10}				
intercept	γ_{100}	-0.28	0.017	-0.26	0.013
HR	γ_{110}	-0.01 *	0.019	--	--
FS	γ_{120}	-0.09	0.023	-0.09	0.018
URM	γ_{130}	0.02 *	0.021	--	--
female	γ_{140}	0.02 *	0.020	--	--
grade	β_{20}				
intercept	γ_{200}	0.19	0.077	0.20	0.056
HR	γ_{210}	-0.05 *	0.071	--	--
FS	γ_{220}	<0.01 *	0.094	--	--
URM	γ_{230}	-0.09 *	0.080	--	--
female	γ_{240}	0.09 *	0.072	--	--
post	β_{30}				
intercept	γ_{300}	0.25	0.035	0.22	0.026
HR	γ_{310}	0.02 *	0.051	--	--
FS	γ_{320}	0.18	0.053	0.13	0.040
URM	γ_{330}	0.07 *	0.048	--	--
female	γ_{340}	-0.08 *	0.047	--	--
Deviance (parameters)		3580.8 (214)		3670.3 (15)	
Chi-square (df) – null		438.7 (205)		349.2 (6)	
Chi-square (df) – 4.4		--		-89.5 (209)	
AIC		4008.8		3700.3	
Iterations		3682		610	

* Coefficient is *not* significant, p>.05; all other values significant.
"—" indicates fixed effect was not entered into the model.

Table 4.7. Grade 6 level 2 model development - random effects

			Model 4.4	Model 4.4'
Random effects estimates	L1, w/in-student	e_{tij}	0.113	0.118
	L2 intercept	r_{0ij}	0.106	0.116
	L2 *time*	r_{1ij}	0.012	0.016
	L3 intercept – int.	u_{00j}	0.013	0.004
	L3 intercept – *HR*	u_{01j}	0.015 *	--
	L3 intercept – *FS*	u_{02j}	0.017 *	--
	L3 intercept – *URM*	u_{03j}	0.007 *	--
	L3 intercept – *female*	u_{04j}	0.010 *	--
	L3 *time* – int.	u_{10j}	0.002 *	--
	L3 *time* – *HR*	u_{11j}	0.003 *	--
	L3 *time* – *FS*	u_{12j}	0.006 *	--
	L3 *time* – *URM*	u_{13j}	0.006 *	--
	L3 *time* – *female*	u_{14j}	0.003 *	--
	L3 *grade* – int.	u_{20j}	0.081	0.038
	L3 *grade* – *HR*	u_{21j}	0.022 *	--
	L3 *grade* – *FS*	u_{22j}	0.089 *	--
	L3 *grade* – *URM*	u_{23j}	0.056 *	--
	L3 *grade* – *female*	u_{24j}	0.026 *	--
	L3 *post* – *HR*	u_{31j}	0.014 *	--
	L3 *post* – *FS*	u_{32j}	0.020 *	--
	L3 *post* – *URM*	u_{33j}	0.023 *	--
	L3 *post* – *female*	u_{34j}	0.014 *	--
Reliability Estimates	L1 intercept	π_0	0.71	0.71
	L1 – *time*	π_1	0.27	0.31
	L2 intercept – int.	β_{00}	0.41	0.40
	L2 intercept – *HR*	β_{01}	0.29	--
	L2 intercept – *FS*	β_{02}	0.26	--
	L2 intercept – *URM*	β_{03}	0.15	--
	L2 intercept – *female*	β_{04}	0.24	--
	L2 *time* – int.	β_{10}	0.20	--
	L2 *time* – *HR*	β_{11}	0.14	--
	L2 *time* – *FS*	β_{12}	0.20	--
	L2 *time* – *URM*	β_{13}	0.24	--
	L2 *time* – *female*	β_{14}	0.17	--
	L2 *grade* – int.	β_{20}	0.49	0.73
	L2 *grade* – *HR*	β_{21}	0.06	--
	L2 *grade* – *FS*	β_{22}	0.15	--
	L2 *grade* – *URM*	β_{23}	0.15	--
	L2 *grade* – *female*	β_{24}	0.11	--
	L2 *post* – *HR*	β_{31}	0.09	--
	L2 *post* – *FS*	β_{32}	0.09	--
	L2 *post* – *URM*	β_{33}	0.17	--
	L2 *post* – *female*	β_{34}	0.17	--

* Coefficient is *not* significant, p>.05; all other values significant.
"—" indicates random effect was not entered into the model.

Removing the non-significant fixed and random effects resulted in model 4.4’ (no additional refinements were necessary to finalize the level 2 model). Requiring far fewer iterations than the starting L2 model (610 compared to 3,682), the deviance statistic was observed to actually be larger than the previous model; however, due to the reduction of 199 parameters, the AIC reflects the parsimony preferred in this reduced model and is considerably lower (3700.3 compared to 4008.8). Once again, the model was observed to be a significant improvement from model 4.1 when no predictors had yet been entered (χ^2=349.2, *df*=6, *p*<.001). Calculating the pseudo R^2 statistics, we obtain the following:

L1 pseudo R^2: $\frac{0.134-0.118}{0.134} = 0.121$

L2 pseudo R^2: $\frac{0.122-0.116}{0.122} = 0.045$

L3 pseudo R^2: $\frac{0.0040-0.0041}{0.0040} = -0.037$

The addition of the level 2 variables in model 4.4’ have slightly reduced the additional level 1 variance contribution from model 4.3’ lacking L2 variables (previously, a 12.6% reduction down to 12.1%), but still reflects an improvement over the initial null model in 4.1. However, adding student-level variables has contributed to a 4.5% reduction in the level 2 variance (while we once again ignore the negative L3 pseudo R^2). The final model for grade 6 after the refinement of level 2 variables is presented in equation 4.4’:

$$fascination_{tij} = \gamma_{000} + \gamma_{010}\boldsymbol{FS}_{ij} + \gamma_{020}female_{ij} + \gamma_{100}time_{tij} + \gamma_{110}time_{tij} * \boldsymbol{FS}_{ij} + \gamma_{200}grade_{tij} + \gamma_{300}post_{tij} + \gamma_{310}post_{tij}\boldsymbol{FS}_{ij} + r_{0ij} + r_{1ij}time_{tij} + \mu_{00j} + \mu_{20j}grade_{tij} + e_{tij} \quad (4.4')$$

Final interpretation of the grade 6 model will hinge on the addition of level 3 predictors.

Grade 8 Level 2 Model Development

As seen in equation 4.3'', the functional form for grade 8 purposefully reflects the discontinuous model selected for grade 6, with the key difference that the change in level (represented by the *grade* variable at L1) was not found to be significant and therefore was not retained. Equations 4.4a' through 4.4s' below show the L1, L2, and L3 models for grade 8, noting the missing random error terms for the *post* variable (r_2 and μ_{20}) based on the level 1 model development process:

Level 1:

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}time_{tij} + \pi_{2ij}post_{tij} + e_{tij} \quad (4.4a')$$

Level 2:

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}\boldsymbol{HR}_{ij} + \beta_{02j}\boldsymbol{FS}_{ij} + \beta_{03j}URM_{ij} + \beta_{04j}female_{ij} + r_{0ij} \quad (4.4b')$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}\boldsymbol{HR}_{ij} + \beta_{12j}\boldsymbol{FS}_{ij} + \beta_{13j}URM_{ij} + \beta_{14j}female_{ij} + r_{1ij} \quad (4.4c')$$

$$\pi_{2ij} = \beta_{20j} + \beta_{21j}\boldsymbol{HR}_{ij} + \beta_{22j}\boldsymbol{FS}_{ij} + \beta_{23j}URM_{ij} + \beta_{24j}female_{ij} \quad (4.4d')$$

Level 3:

$$\beta_{00j} = \gamma_{000} + \mu_{00j} \quad (4.4e')$$

$$\beta_{01j} = \gamma_{010} + \mu_{01j} \quad (4.4f')$$

$$\beta_{02j} = \gamma_{020} + \mu_{02j} \quad (4.4g')$$

$$\beta_{03j} = \gamma_{030} + \mu_{03j} \quad (4.4h')$$

$$\beta_{04j} = \gamma_{040} + \mu_{04j} \quad (4.4i')$$

$$\beta_{10j} = \gamma_{100} + \mu_{10j} \quad (4.4j')$$

$$\beta_{11j} = \gamma_{110} + \mu_{11j} \quad (4.4k')$$

$$\beta_{12j} = \gamma_{120} + \mu_{12j} \quad (4.4l')$$

$$\beta_{13j} = \gamma_{130} + \mu_{13j} \quad (4.4m')$$

$$\beta_{14j} = \gamma_{140} + \mu_{14j} \quad (4.4n')$$

$$\beta_{20j} = \gamma_{200} \quad (4.4o')$$

$$\beta_{21j} = \gamma_{210} + \mu_{21j} \quad (4.4p')$$

$$\beta_{22j} = \gamma_{220} + \mu_{22j} \quad (4.4q')$$

$$\beta_{23j} = \gamma_{230} + \mu_{23j} \quad (4.4r')$$

$$\beta_{24j} = \gamma_{240} + \mu_{24j} \quad (4.4s')$$

Tables 4.8 and 4.9 show the results of the grade 8 level 2 models for fixed and random effect estimates, respectively. Although the model estimates far fewer parameters than the grade 6 level 2 starting model due to the removal of *grade* at L1, convergence for the grade 8 level 2 starting model still required a large number of iterations (2,538) due to increased parameterization (124). Compared to the null model, the addition of level 2 predictors has unsurprisingly improved overall model fit (χ^2=416.9 *df*=115 *p*<.001). Visible in Table 4.9, the race variable *URM* was found to be non-significant for all L1 terms (intercept, *time,* and *post*); the random effects, however, display a very similar pattern to grade 6 as only one of the newly added predictors (family support for science) was found to be significant.

Iterative refinements involving the removal of non-significant fixed and random effects resulted in model 4.4'', a more parsimonious model requiring 91 iterations to converge. Once again, we observe a slight increase in the deviance statistic compared to the starting L2 model, but with far fewer parameters estimated, the AIC reflects an improvement compared to the model 4.4 starting point (4303.9 vs. 4438.0). Compared to model 4.1, the result is a significantly improved model overall (χ^2=341.0 *df*=10 *p*<.001).

Table 4.8. Grade 8 level 2 model development - fixed effects

Fixed effects		Model 4.4 Coeff.	Model 4.4 SE	Model 4.4'' Coeff.	Model 4.4'' SE
Intercept	β_{00}				
intercept	γ_{000}	2.50	0.018	2.50	0.016
HR	γ_{010}	0.06	0.021	0.07	0.023
FS	γ_{020}	0.11	0.021	0.11	0.023
URM	γ_{030}	-0.01 *	0.023	--	--
female	γ_{040}	-0.10	0.023	-0.11	0.021
time	β_{10}				
intercept	γ_{100}	-0.12	0.018	-0.12	0.012
HR	γ_{110}	-0.06	0.019	-0.07	0.018
FS	γ_{120}	-0.07	0.016	-0.05	0.014
URM	γ_{130}	0.01 *	0.020	--	--
female	γ_{140}	0.01 *	0.017	--	--
post	β_{20}				
intercept	γ_{200}	0.27	0.044	0.23	0.028
HR	γ_{210}	0.18	0.041	0.24	0.043
FS	γ_{220}	0.10	0.043	--	--
URM	γ_{230}	< -0.01 *	0.038	--	--
female	γ_{240}	-0.09	0.037	--	--
Deviance (parameters)		4190.0 (124)		4265.9 (19)	
Chi-square (df) – null		416.9 (115)		341.0 (10)	
Chi-square (df) – 4.4		--		-75.9 (105)	
AIC		4438.0		4303.9	
Iterations		2538		91	

* Coefficient is *not* significant, p>.05; all other values significant.
"—" indicates fixed effect was not entered into the model.

Table 4.9. Grade 8 level 2 model development - random effects

			Model 4.4	Model 4.4''
Random effects estimates	L1, w/in-student	e_{tij}	0.105	0.106
	L2 intercept	r_{0ij}	0.090	0.098
	L2 *time*	r_{1ij}	0.015	0.019
	L3 intercept – int.	u_{00j}	0.007	0.004
	L3 intercept – *HR*	u_{01j}	0.006 *	--
	L3 intercept – *FS*	u_{02j}	0.006 *	--
	L3 intercept – *URM*	u_{03j}	0.006 *	--
	L3 intercept – *female*	u_{04j}	0.010 *	--
	L3 *time* – int.	u_{10j}	0.007	0.003
	L3 *time* – *HR*	u_{11j}	0.006 *	--
	L3 *time* – *FS*	u_{12j}	0.003 *	--
	L3 *time* – *URM*	u_{13j}	0.005 *	--
	L3 *time* – *female*	u_{14j}	0.005 *	--
	L3 *post* – *HR*	u_{21j}	0.036	0.020
	L3 *post* – *FS*	u_{22j}	0.024	--
	L3 *post* – *URM*	u_{23j}	0.015 *	--
	L3 *post* – *female*	u_{24j}	0.008 *	--
Reliability Estimates	L1 intercept	π_0	0.69	0.70
	L1 – *time*	π_1	0.28	0.31
	L2 intercept – int.	β_{00}	0.28	0.39
	L2 intercept – *HR*	β_{01}	0.16	--
	L2 intercept – *FS*	β_{02}	0.17	--
	L2 intercept – *URM*	β_{03}	0.16	--
	L2 intercept – *female*	β_{04}	0.28	--
	L2 *time* – int.	β_{10}	0.48	0.59
	L2 *time* – *HR*	β_{11}	0.25	--
	L2 *time* – *FS*	β_{12}	0.15	--
	L2 *time* – *URM*	β_{13}	0.23	--
	L2 *time* – *female*	β_{14}	0.28	--
	L2 *post* – *HR*	β_{21}	0.27	0.32
	L2 *post* – *FS*	β_{22}	0.20	--
	L2 *post* – *URM*	β_{23}	0.20	--
	L2 *post* – *female*	β_{24}	0.14	--

***** Estimate is *not* significant, p>.05; all other values significant.
"—" indicates random effect was not entered into the model.

The pseudo R^2 statistics are calculated as follows:

L1 pseudo R^2: $\frac{0.111-0.106}{0.111} = 0.046$

L2 pseudo R^2: $\frac{0.105-0.098}{0.105} = 0.069$

L3 pseudo R^2: $\frac{0.005-0.003}{0.005} = 0.237$

The addition of the level 2 variables in model 4.4'' have slightly increased the additional level 1 variance contribution from model 4.3'' lacking L2 variables (a reduction of 4.6%, up from 4.1%). Furthermore, we now see that the student-level predictors have contributed an additional 3.3% to the variance explained at level 2 (6.9% of the initial 47.7% L2 ICC calculated from the null model). Although the pseudo R^2 calculated for L3 is encouraging, this will be re-evaluated after the addition of level 3 predictors.

The final model for grade 8 after the refinement of level 2 variables is presented in equation 4.4'':

$$\begin{aligned} fascination_{tij} &= \gamma_{000} + \gamma_{010}\boldsymbol{HR}_{ij} + \gamma_{020}\boldsymbol{FS}_{ij} + \gamma_{030}female_{ij} + \gamma_{100}time_{tij} + \\ &\gamma_{110}time_{tij} * \boldsymbol{HR}_{ij} + \gamma_{120}time_{tij} * \boldsymbol{FS}_{ij} + \gamma_{200}post_{tij} + \gamma_{210}post_{tij}\boldsymbol{HR}_{ij} + \\ &r_{0ij} + r_{1ij}time_{tij} + \mu_{00j} + \mu_{10j}time_{tij} + \mu_{21j}post_{tij} * \boldsymbol{HR}_{ij} + e_{tij} \qquad (4.4'') \end{aligned}$$

Prior to making substantive interpretations of the results presented in model 4.4'' above, classroom-level covariates will be added to level 3.

Level 3: Class-level Covariates

The model building process concludes with the addition of classroom-level covariates at level 3. Similar to the process at level 2, simultaneous entry of the instructional variables student-centric teaching (*student*), hands-on instruction (*hands-on*), and classroom technology use (*classtech*) allows a determination of the significance for each term for all level 3 equations. Once again, the sections that follow will focus on the starting models employing backwards elimination of relevant predictor variables and then a comparison of the final level 3 model selected, first for grade 6 and then for grade 8.

Grade 6 Level 3 Model Development

Equations 4.5a through 4.5m show the grade 6 models, now with level 3 predictors added. While 4.5a through 4.5e remain unchanged based on the results presented in Tables 4.6 and 4.7, the remaining equations (4.5f-4.5m) now show bolded terms for the grand-mean centered instructional variables, testing to see if these explain meaningful variance in fascination scores based on the discontinuous growth model that has been developed thus far:

Level 1:

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}time_{tij} + \pi_{2ij}grade_{tij} + \pi_{3ij}post_{tij} + e_{tij} \qquad (4.5a)$$

Level 2:

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}\boldsymbol{FS}_{ij} + \beta_{02j}female_{ij} + r_{0ij} \tag{4.5b}$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}\boldsymbol{FS}_{ij} + r_{1ij} \tag{4.5c}$$

$$\pi_{2ij} = \beta_{20j} \tag{4.5d}$$

$$\pi_{3ij} = \beta_{30j} + \beta_{31j}\boldsymbol{FS}_{ij} \tag{4.5e}$$

Level 3:

$$\beta_{00j} = \gamma_{000} + \gamma_{001}\boldsymbol{student} + \gamma_{002}\boldsymbol{handson} + \gamma_{003}\boldsymbol{classtech} + \mu_{00j} \tag{4.4f}$$

$$\beta_{01j} = \gamma_{010} + \gamma_{011}\boldsymbol{student} + \gamma_{012}\boldsymbol{handson} + \gamma_{013}\boldsymbol{classtech} \tag{4.4g}$$

$$\beta_{02j} = \gamma_{020} + \gamma_{021}\boldsymbol{student} + \gamma_{022}\boldsymbol{handson} + \gamma_{023}\boldsymbol{classtech} \tag{4.4h}$$

$$\beta_{10j} = \gamma_{100} + \gamma_{101}\boldsymbol{student} + \gamma_{102}\boldsymbol{handson} + \gamma_{103}\boldsymbol{classtech} \tag{4.4i}$$

$$\beta_{11j} = \gamma_{110} + \gamma_{111}\boldsymbol{student} + \gamma_{112}\boldsymbol{handson} + \gamma_{113}\boldsymbol{classtech} \tag{4.4j}$$

$$\beta_{20j} = \gamma_{200} + \gamma_{201}\boldsymbol{student} + \gamma_{202}\boldsymbol{handson} + \gamma_{203}\boldsymbol{classtech} + \mu_{20j} \tag{4.4k}$$

$$\beta_{30j} = \gamma_{300} + \gamma_{301}\boldsymbol{student} + \gamma_{302}\boldsymbol{handson} + \gamma_{303}\boldsymbol{classtech} \tag{4.4l}$$

$$\beta_{31j} = \gamma_{310} + \gamma_{311}\boldsymbol{student} + \gamma_{312}\boldsymbol{handson} + \gamma_{313}\boldsymbol{classtech} \tag{4.4m}$$

Tables 4.10 and 4.11 below show the results for the starting level 3 model (4.5) based on the equations above and, following iterative refinements, the final model that was selected (4.5') to explain change in fascination over time for the grade 6 cohort. Similar to the pattern seen in the level 2 development for both grades, model 4.5' took fewer iterations to converge than model 4.5, is more parsimonious with only 18 parameters estimated (compared to 39), and though the deviance statistic is slightly larger, the AIC is lower, reflecting the tradeoff of predictors and parsimony. We also see that model 4.5' is a significant improvement compared to the null model (χ^2=360.1 *df*=9 *p*<.001). Table 4.10 shows that the majority of level 3 predictors were found to be non-significant, with only 3 of the 24 newly added terms retained in model 4.5'. Equation 4.5' presents the final mixed model for grade 6:

$$
\begin{aligned}
fascination_{tij} &= \gamma_{000} + \gamma_{010}\boldsymbol{FS}_{ij} + \gamma_{020} female_{ij} + \gamma_{021} female_{ij} * \boldsymbol{student}_j + \\
&\gamma_{100} time_{tij} + \gamma_{110}\boldsymbol{FS}_{ij} * time_{tij} + \gamma_{200} grade_{tij} + \gamma_{300} post_{tij} + \gamma_{310}\boldsymbol{FS}_{ij} * post_{tij} + \\
&\gamma_{311}\boldsymbol{FS}_{ij} * \boldsymbol{handson}_j * post_{tij} + \gamma_{312}\boldsymbol{FS}_{ij} * \boldsymbol{classtech}_j * post_{tij} + \\
&r_{0ij} + r_{1ij} time_{tij} + \mu_{00j} + \mu_{20j} grade_{tij} + e_{tij} \qquad (4.5')
\end{aligned}
$$

Using the variance estimates provided in Table 4.11, the pseudo R^2 statistics are calculated as follows:

L1 pseudo R^2: $\frac{0.134-0.118}{0.134} = 0.122$

L2 pseudo R^2: $\frac{0.122-0.116}{0.122} = 0.048$

L3 pseudo R^2: $\frac{0.0040-0.0041}{0.0040} = -0.030$

Comparable to the pseudo R^2 statistics obtained after finalizing level 2, the addition of the level 3 variables has resulted in a final model in which level 1 variance has been reduced by over 12%, as well as almost 5% of the level 2 variance compared to the starting null model. While the contribution to level 3 variance explained is negligible and somewhat disappointing, this is not altogether surprising, given that less than 2% of the overall variance was found to reside at level 3 between classrooms, and the majority of the variables initially entered at the class-level were found to be non-significant.

Table 4.10. Grade 6 level 3 model development - fixed effects

Fixed effects		Model 4.5		Model 4.5'	
		Coeff.	SE	Coeff.	SE
Intercept	β_{00}				
intercept	γ_{000}	2.65	0.019	2.65	0.020
student	γ_{001}	0.17 *	0.117	--	--
hands-on	γ_{002}	-0.04 *	0.111	--	--
class-tech	γ_{003}	0.25 *	0.163	--	--
FS int.	γ_{010}	0.16	0.029	0.16	0.029
student	γ_{011}	-0.02 *	0.189	--	--
hands-on	γ_{012}	-0.01 *	0.175	--	--
class-tech	γ_{013}	0.30 *	0.227	--	--
female int.	γ_{020}	-0.11	0.020	-0.11	0.020
student	γ_{021}	-0.33	0.116	-0.20	0.088
hands-on	γ_{022}	0.17 *	0.120	--	--
class-tech	γ_{023}	0.12 *	0.215	--	--
time	β_{10}				
intercept	γ_{100}	-0.26	0.012	-0.26	0.013
student	γ_{101}	0.05 *	0.085	--	--
hands-on	γ_{102}	0.02 *	0.088	--	--
class-tech	γ_{103}	-0.18 *	.0106	--	--
FS int.	γ_{110}	-0.09	0.018	-0.09	0.018
student	γ_{111}	<0.01 *	0.126	--	--
hands-on	γ_{112}	0.02 *	0.100	--	--
class-tech	γ_{113}	0.01 *	0.196	--	--
grade	β_{20}				
intercept	γ_{200}	0.33	0.054	0.21	0.057
student	γ_{201}	-0.01 *	0.391	--	--
hands-on	γ_{202}	-1.35	0.333	--	--
class-tech	γ_{203}	-1.73	0.428	--	--
post	β_{30}				
intercept	γ_{300}	0.17	0.031	0.22	0.027
student	γ_{302}	-0.03 *	0.265	--	--
hands-on	γ_{303}	0.45 *	0.241	--	--
class-tech	γ_{304}	0.84	0.259	--	--
FS int.	γ_{310}	0.20	0.042	0.20	0.037
student	γ_{311}	0.28 *	0.400	--	--
hands-on	γ_{312}	-0.78	0.332	-0.54	0.122
class-tech	γ_{313}	-0.95	0.474	-0.75	0.261
Deviance (parameters)		3637.5 (39)		3659.4 (18)	
Chi-square (df) – null		382.0 (30)		360.1 (9)	
Chi-square (df) – 4.5		--		-21.9 (21)	
AIC		3715.5		3695.4	
Iterations		486		37	

* Coefficient is *not* significant, p>.05; all other values significant.
"—" indicates fixed effect was not entered into the model.

Table 4.11. Grade 6 level 3 model development - random effects

			Model 4.5	Model 4.5'
Random effects estimates	L1, w/in-student	e_{tij}	0.117	0.118
	L2 intercept	r_{0ij}	0.116	0.116
	L2 *time*	r_{1ij}	0.015	0.015
	L3 intercept – int.	u_{00j}	0.003	0.004
	L3 *grade* – int.	u_{20j}	0.028	0.041
Reliability Estimates	L1 intercept	π_0	0.71	0.71
	L1 – *time*	π_1	0.30	0.30
	L2 intercept – int.	β_{00}	0.35	0.40
	L2 *grade* – int.	β_{20}	0.66	0.74

Grade 8 Level 3 Model Development

Equations 4.5a' through 4.5m' show the grade 8 models with grand-mean centered level 3 predictors added. As with grade 6, the equations for levels 1 and 2 remain unchanged following model 4.4'', while the level 3 equations have added the three class-level covariates as grand-mean centered terms:

Level 1:

$$Y_{tij} = \pi_{0ij} + \pi_{1ij} time_{tij} + \pi_{2ij} post_{tij} + e_{tij} \quad (4.5a')$$

Level 2:

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}\boldsymbol{HR}_{ij} + \beta_{02j}\boldsymbol{FS}_{ij} + \beta_{03j} female_{ij} + r_{0ij} \quad (4.5b')$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}\boldsymbol{HR}_{ij} + \beta_{12j}\boldsymbol{FS}_{ij} + r_{1ij} \quad (4.5c')$$

$$\pi_{2ij} = \beta_{20j} + \beta_{21j}\boldsymbol{HR}_{ij} \quad (4.5d')$$

Level 3:

$$\beta_{00j} = \gamma_{000} + \gamma_{001}\boldsymbol{student} + \gamma_{002}\boldsymbol{handson} + \gamma_{003}\boldsymbol{classtech} + \mu_{00j} \quad (4.5e')$$

$$\beta_{01j} = \gamma_{010} + \gamma_{011}\boldsymbol{student} + \gamma_{012}\boldsymbol{handson} + \gamma_{013}\boldsymbol{classtech} \quad (4.5f')$$

$$\beta_{02j} = \gamma_{020} + \gamma_{021}\boldsymbol{student} + \gamma_{022}\boldsymbol{handson} + \gamma_{023}\boldsymbol{classtech} \quad (4.5g')$$

$$\beta_{03j} = \gamma_{030} + \gamma_{031}\boldsymbol{student} + \gamma_{032}\boldsymbol{handson} + \gamma_{033}\boldsymbol{classtech} \quad (4.5h')$$

$$\beta_{10j} = \gamma_{100} + \gamma_{101}\boldsymbol{student} + \gamma_{102}\boldsymbol{handson} + \gamma_{103}\boldsymbol{classtech} + \mu_{10j} \quad (4.5i')$$

$$\beta_{11j} = \gamma_{110} + \gamma_{111}\boldsymbol{student} + \gamma_{112}\boldsymbol{handson} + \gamma_{113}\boldsymbol{classtech} \tag{4.5j'}$$

$$\beta_{12j} = \gamma_{120} + \gamma_{121}\boldsymbol{student} + \gamma_{122}\boldsymbol{handson} + \gamma_{123}\boldsymbol{classtech} \tag{4.5k'}$$

$$\beta_{20j} = \gamma_{200} + \gamma_{201}\boldsymbol{student} + \gamma_{202}\boldsymbol{handson} + \gamma_{203}\boldsymbol{classtech} \tag{4.5l'}$$

$$\beta_{21j} = \gamma_{210} + \gamma_{211}\boldsymbol{student} + \gamma_{212}\boldsymbol{handson} + \gamma_{213}\boldsymbol{classtech} + \mu_{21j} \tag{4.5m'}$$

Tables 4.12 and 4.13 present the results for the starting level 3 model (4.5) based on the equations above and, following iterative refinements, the final model that was selected (4.5'') to explain change in fascination over time for the grade 8 cohort. Once again, we observe a pattern in which model 4.5'' took fewer iterations to converge than the starting level 3 model (23 compared to 155), has fewer parameters (16 compared to 46), but a larger deviance statistic (4306.9 compared to 4212.9). In this instance, unlike the final model at grade 6, model 4.5'' was found to have a slightly *larger* AIC value (4338.9 compared to 4304.9), suggesting that the parsimony provided by the final model is not discounting the model misfit as much as the starting model. However, unlike model 4.5, all of the fixed and random effects that have been retained are statistically significant, and compared to the unconditional model presented in 4.1, the model is a significantly better fit (χ^2=360.1 *df*=9 *p*<.001).

Table 4.12 below shows that almost all of the level 3 variables entered into the grade 8 model were found to be non-significant, with only 1 of the 27 newly added terms retained in model 4.5''. Interestingly, once level 3 was modeled with predictors, this resulted in a non-significant fixed parameter estimated for level 2 (β_{21}) which had previously been retained in level 2 modeling. In other words, once the data were modeled with level 3 predictors, home resources no longer significantly influenced the Year 2 slope trajectory, *post*.

Using the variance estimates provided in Table 4.13, the pseudo R^2 statistics are calculated as follows:

L1 pseudo R^2: $\frac{0.111-0.106}{0.111} = 0.040$

L2 pseudo R^2: $\frac{0.105-0.097}{0.105} = 0.080$

L3 pseudo R^2: $\frac{0.005-0.004}{0.005} = 0.224$

While the pseudo R^2 statistic for level 1 is slightly lower than that which was obtained after finalizing level 2, we still observe a final model in which the unexplained variance in the random parameter, *e*, has been reduced by 4%, and unexplained variance in the level 2 intercept, r_0, has been reduced by 8%. Importantly, modeling at level 3 has led to a reduction in the level 3 unexplained variance, u_{00}, by 22%. In contrast to model 4.5' for grade 6, this finding is indeed surprising, as only one predictor was retained at level 3. Given that the variance accounted for at level 3 was only found to be 2.1% based on the initial unconditional model, however, this additional variance explained still accounts for less than 0.5% of the total variation in fascination scores.

Table 4.12. Grade 8 level 3 model development - fixed effects

Fixed effects		Model 4.5 Coeff.	Model 4.5 SE	Model 4.5'' Coeff.	Model 4.5'' SE
Intercept	β_{00}				
intercept	γ_{000}	2.50	0.015	2.50	0.015
student	γ_{001}	0.12 *	0.101	--	--
hands-on	γ_{002}	-0.06 *	0.081	--	--
class-tech	γ_{003}	0.12 *	0.127	--	--
HR int.	γ_{010}	0.07	0.021	0.08	0.023
student	γ_{011}	-0.08 *	0.130	--	--
hands-on	γ_{012}	0.11 *	0.102	--	--
class-tech	γ_{013}	0.42 *	0.363	--	--
FS int.	γ_{020}	0.11	0.022	0.11	0.022
student	γ_{021}	0.12 *	0.116	--	--
hands-on	γ_{022}	-0.18	0.089	--	--
class-tech	γ_{023}	0.67	0.309	0.61	0.216
female int.	γ_{030}	-0.11	0.021	-0.12	0.021
student	γ_{031}	-0.07 *	0.129	--	--
hands-on	γ_{032}	-0.11 *	0.121	--	--
class-tech	γ_{033}	-0.12 *	0.224	--	--
time	β_{10}				
intercept	γ_{100}	-0.12	0.011	-0.12	0.012
student	γ_{101}	-0.10 *	0.073	--	--
hands-on	γ_{102}	0.06 *	0.073	--	--
class-tech	γ_{103}	-0.06 *	0.115	--	--
HR int.	γ_{110}	-0.07	0.018	-0.02	0.012
student	γ_{111}	0.04 *	0.109	--	--
hands-on	γ_{112}	-0.03 *	0.091	--	--
class-tech	γ_{113}	-0.02 *	0.303	--	--
FS int.	γ_{120}	-0.05	0.013	-0.05	0.014
student	γ_{121}	-0.06 *	0.081	--	--
hands-on	γ_{122}	0.02 *	0.074	--	--
class-tech	γ_{123}	-0.16 *	0.149	--	--
post	β_{20}				
intercept	γ_{200}	0.25	0.046	0.26	0.033
student	γ_{201}	0.26 *	0.359	--	--
hands-on	γ_{202}	0.15 *	0.314	--	--
class-tech	γ_{203}	-0.81 *	0.425	--	--
HR int.	γ_{210}	0.11 *	0.079	--	--
student	γ_{211}	-0.87 *	0.777	--	--
hands-on	γ_{212}	1.35 *	0.781	--	--
class-tech	γ_{213}	0.65 *	0.945	--	--
Deviance (parameters)		4212.9 (46)		4306.9 (16)	
Chi-square (df) – null		394.0 (37)		300.1 (7)	
Chi-square (df) – 4.5		--		-94.0 (30)	
AIC		4304.9		4338.9	
Iterations		155		23	

* Coefficient is *not* significant, p>.05; all other values significant.
"—" indicates fixed effect was not entered into the model.

Table 4.13. Grade 8 level 3 model development - random effects

			Model 4.5	Model 4.5''
Random effects estimates	L1, w/in-student	e_{tij}	0.105	0.106
	L2 intercept	r_{0ij}	0.097	0.097
	L2 *time*	r_{1ij}	0.019	0.020
	L3 intercept – int.	u_{00j}	0.002	0.004
	L3 *time* – int.	u_{10j}	0.002	0.004
	L3 *grade – HR*	u_{21j}	0.008	--
Reliability Estimates	L1 intercept	π_0	0.70	0.90
	L1 – *time*	π_1	0.31	0.32
	L2 intercept – int.	β_{00}	0.30	0.39
	L2 *time* – int.	β_{10}	0.46	0.63
	L2 *post – HR*	β_{21}	0.18	--

Equation 4.5'' presents the final mixed model for grade 8:

$$fascination_{tij} = \gamma_{000} + \gamma_{010}\boldsymbol{HR}_{ij} + \gamma_{020}\boldsymbol{FS}_{ij} + \gamma_{021}\boldsymbol{FS}_{ij} * \boldsymbol{classtech}_j + \gamma_{030}female_{ij} + \gamma_{100}time_{tij} + \gamma_{110}\boldsymbol{HR}_{ij} * time_{tij} + \gamma_{120}\boldsymbol{FS}_{ij} * time_{tij} + \gamma_{200}post_{tij} + r_{0ij} + r_{1ij}time_{tij} + \mu_{00j} + \mu_{10j}time_{tij} + e_{tij} \quad (4.5'')$$

Residual Analyses

Diagnostic checks were performed regarding the specifications of the hierarchical models, based on the assumptions laid out in Chapter 3 concerning the error structures and predictor variables. Using residuals from final models 4.5' and 4.5'' for grades 6 and 8, respectively, Appendices B and C present a graphical analysis for the error terms at each level, while tables examining the correlations between the residuals and predictors are also included.

Specific to grade 6 model 4.5', Appendix B first presents a series of histograms to test for normally distributed residuals with a mean of zero at all three levels. The figures displayed in B1-B5 confirm that the mean for each residual term

(e, r_0, r_1, μ_{00}, and μ_{20}) approximates zero, and for the most part, the histograms follow the normal curve. This supports the error structure assumptions regarding normality. The figures in B6-B12 plot the error terms against categorical predictors at each level to determine if significant differences exist in the residuals, all of which follow similar normal distribution patterns. The table displayed in B13 shows the correlations between the predictors and residuals, with no statistically significant relationships present. This supports the predictor variable assumptions regarding independence.

Appendix C presents the same set of histograms specific to grade 8 model 4.5'' in the same order. We first see the standard normal curve with mean 0 overlaid against the residual terms (e, r_0, r_1, μ_{00}, and μ_{10}), although there are slight irregularities in the distributional pattern of the level 3 residuals associated with the *time* variable, μ_{10}. Despite this, the remaining histograms (C6-C11) show relatively uniform distributions across categorical variables at levels 1 and 2. The table of correlations in C12 shows only one significant level 3 relationship between the classroom-technology predictor and the residuals associated with the *time* variable, μ_{10}. Correlated predictors and residuals within level signal potential model misspecification, however, the variable *classtech* is only modeled in one interaction term without constituting a large share of explained variance.

Conclusion

In order to clarify the technical findings presented in this chapter, an overall interpretation of each final model in the context of change in fascination scores is warranted prior to answering the research questions posed in this dissertation in the next chapter. Beginning at grade 6 and substituting the observed fixed parameter estimates from model 4.5' into equation 4.5' allows us to better interpret each of the terms. Equation 4.6 below presents the final grade 6 predictive model for student fascination in science:

$$\widehat{fascination}_{tij} = 2.65 + 0.16\boldsymbol{FS}_{ij} - 0.11 female_{ij} - 0.20 female_{ij} * \boldsymbol{student}_j - 0.26 time_{tij} - 0.09\boldsymbol{FS}_{ij} * time_{tij} + 0.21 grade_{tij} + 0.22 post_{tij} + 0.20\boldsymbol{FS}_{ij} * post_{tij} - 0.54\boldsymbol{FS}_{ij} * \boldsymbol{handson}_j * post_{tij} - 0.75\boldsymbol{FS}_{ij} * \boldsymbol{classtech}_j * post_{tij} \quad (4.6)$$

Because this equation now models *predicted* fascination scores, the error terms $(r_{0ij}, r_{1ij}, \mu_{00j}, \mu_{20j}, e_{tij})$ are notably missing; however, it is assumed that each student's individual observed fascination score would have some amount of deviance from the prediction above that is captured in the model through the residuals.

The initial term representing the intercept indicates that the expected fascination score for boys in the middle of their 6th grade school year (*time*=0, measurement point 2) is 2.65, on average (i.e., those who have family support equal to the grand-mean). Girls, however, have an expected fascination score that is 0.11 points (or roughly two-tenths of a standard deviation) lower across all time points. For both boys and girls, fascination decreases linearly from the start of the school year (average fascination scores at *time*=-1: 2.91 and 2.80, respectively) until the end of the school year (average fascination scores at *time*=1: 2.39 and 2.28, respectively). Due to the discontinuous nature of the model—including both a change in level (new intercept at the start of grade 7) and change in slope (new growth trajectory across the grade 7 measurement occasions)—scores at the start of Year 2 come back up slightly, on average (2.56, 2.45) but are once again depressed by the end of the year (2.52, 2.41). In other words, the change in level between grades 6 and 7 was observed to be 0.17 fascination scale score points, while the change in slope was found to be -0.04 fascination scale score points (implying a slight *increase* compared to the prior year, given the negative slope initially observed).

The γ_{010}, γ_{110}, and γ_{310} estimates related to the grand-mean centered *FS* suggests that students with above average family support have higher fascination scores, on average, and that this effect fluctuates over time based on the interactions with *time* and *post.* Holding all else constant, those who score 1 point higher than the grand mean on the *FS* scale are expected to be 0.25 points higher on the fascination score scale at the start of grade 6 (*time*=-1); this influence is reduced by 0.09 scale points at each remaining time point in Year 1 (0.16 points higher at *time*=0, 0.07 points higher at *time*=1). In Year 2 (grade 7), the trend reverses such that each incremental scale point higher in family support translates to an additional 0.11 fascination scale score points above the 0.07 observed at the end of Year 1 (0.18 points higher at *time*=2, 0.29 points higher at *time*=3). Given that the family support mean for grade 6 students was observed to be 3.51 (see Table 4.1) with a standard deviation of 0.49 and a restricted range (1.00-4.00), a more likely scenario involves students scoring one-tenth of a point above or below the grand-mean, and predicted changes one-tenth the magnitude of those described above.

The remaining coefficient estimates all pertain to covariates at level 3 and involve interactions with lower level terms. Specifically, γ_{021}, the coefficient related to the *female* **student* term in equation 4.6, demonstrates the additional reduction in expected fascination scores for girls for cases in which student-centric teaching occurs above average. In the case of γ_{311} and γ_{312}, the coefficients for *FS*handson*post* and *FS*classtech*post*, students in Year 2 (grade 7, when *post* ≠ 0) with above grand-mean scores for family support and above average classroom hands-on instruction or classroom technology use are expected to see *decreases* in fascination scores.

Transitioning to the final grade 8 model, we now substitute the observed fixed parameter estimates from model 4.5'' into equation 4.5'', allowing us to replicate this interpretive process for eighth graders. Equation 4.7 presents the final predictive model for student fascination in science at grade 8 (while again, residuals are not included in the predictive notation):

$$\begin{aligned} \widehat{fascination}_{tij} = {} & 2.50 + 0.08\boldsymbol{HR}_{ij} + 0.11\boldsymbol{FS}_{ij} + 0.61\boldsymbol{FS}_{ij} * \boldsymbol{classtech}_j - \\ & 0.12 female_{ij} - 0.12 time_{tij} - 0.02\boldsymbol{HR}_{ij} * time_{tij} - \\ & 0.05\boldsymbol{FS}_{ij} * time_{tij} + 0.26 post_{tij} \end{aligned} \tag{4.7}$$

Here, we observe similarities to the sixth grade model; namely, the negative slope coefficients associated with the variables *time* and *female*. The intercept, 2.50, represents the expected fascination score for boys in the middle of their 8th grade school year with grand-mean equivalents for home resources and family support for science, while girls have an expected fascination score that is 0.12 points lower (roughly 0.25 s.d.) at all time points. Fascination again decreases linearly from the start of the school year (average fascination scores at *time*=-1: 2.62 and 2.50, respectively) until the end of the school year (average fascination scores at *time*=1: 2.38 and 2.26, respectively). Although the nature of the discontinuity for the grade 8 model differs in that there is no change in level at summer break (i.e., the *grade* variable is not present in the final model), we still observe a change in slope in which fascination scores are expected to rise in Year 2. By the end of ninth grade, average fascination scores for both boys and girls are slightly higher than the beginning of eighth grade (2.66 and 2.54, respectively). In other words, the γ_{200} coefficient estimate for *post* of 0.26, in concert with the original γ_{100} coefficient for *time* of -0.12 results in a change in slope of +0.14.

The remaining terms suggest an increase in fascination scores for all students with above average home resources (γ_{010}, γ_{110}) or family support for science (γ_{020}, γ_{120}), the influence of each of which decreases as time goes on due to their observed negative interaction coefficients with the *time* variable. Lastly, students with above average family support scores who receive more than the grand-mean of instructional classroom technology use (γ_{021}=0.61) correspond with increased fascination scores.

The next chapter extends these interpretations by linking them directly to the research questions. More specifically, the next chapter offers reflections on the data and analyses herein, while discussing the implications of findings toward improving student fascination in science more broadly.

Chapter 5: Discussion

This dissertation sought to answer three research questions, each of which focused on different aspects of students' fascination in science and potential influences:

RQ1: What is the average change in science fascination of middle school students over time? And how much does this change vary, on average?

RQ2: To what extent is predicted student science fascination associated with the instructional characteristics of the classroom, including student-centric teaching, hands-on methods, and classroom technology use? And are these effects sustained over time?

RQ3: To what extent are middle schools equally successful in sustaining science fascination for students whose gender, race, family support for science, or economic backgrounds differ?

These questions were explored for two cohorts of students, one composed of students who began the study in grade 6 and the second composed of students who began the study in grade 8. For both cohorts of students, hierarchical models were iteratively constructed to address these questions. This chapter discusses model results in relation to each research question, making connections to prior research as applicable. After a description of the limitations of the current work, implications for researchers and practitioners are offered as concluding remarks.

Change in Middle School Students' Science Fascination

The first research question asked specifically about the amount of change—and implicitly, about the nature of this change—in student fascination scores:

> *What is the average change in science fascination of middle school students over time? And how much does this change vary, on average?*

Detailed in the previous chapter, multilevel equations were constructed to model student science fascination scores for both grade-level cohorts. In a linear growth model, the answer to this question would involve a discussion of the estimated L1 growth parameter (coefficient π_{1ij} for the linear slope variable, *time*) along with the error term associated with the unspecified L2 equation for the slope variable, r_{1ij}, prior to the addition of student- and class-level covariates at levels 2 and 3. However, with five measurement occasions and the discontinuous patterns of change identified at both grade levels, a simple linear equation does not adequately model change in fascination. Instead, the interpretation of change over time hinges on simultaneously analyzing the linear *time* component and the discontinuous terms for change in level (*grade*) and change in slope (*post*). Likewise, we must take into account the combination of growth parameters in the discussion of variance by analyzing the L2 error terms associated with all three variables: r_{1ij} for the linear time component, r_{2ij} for the *grade* change-in-level discontinuous component, and

r_{3ij} for the *post* change-in-slope discontinuous component.[4] Both discontinuous grade-level models were found to be statistically significantly better fits for the data compared to their respective null models, which included only the linear *time* variable (for grade 6, χ^2=281.1, *df*=5, *p*<.001; for grade 8, χ^2=188.2, *df*=1, *p*<.001).

Returning to equations 4.3' and 4.3'' for grades 6 and 8, respectively, removing terms associated with stochastic error to evaluate predicted fascination scores after the level 1 modeling phase, and inserting coefficient estimates from Tables 4.4 and 4.5 for each, we observe equation 5.1 for grade 6 and equation 5.2 for grade 8 which are necessary to address this first research question:

$$\widehat{fascination}_{tij} = 2.60 - 0.26 * time_{tij} + 0.20 * grade_{tij} + 0.22 * post_{tij} \qquad (5.1)$$

$$\widehat{fascination}_{tij} = 2.44 - 0.12 * time_{tij} + 0.21 * post_{tij} \qquad (5.2)$$

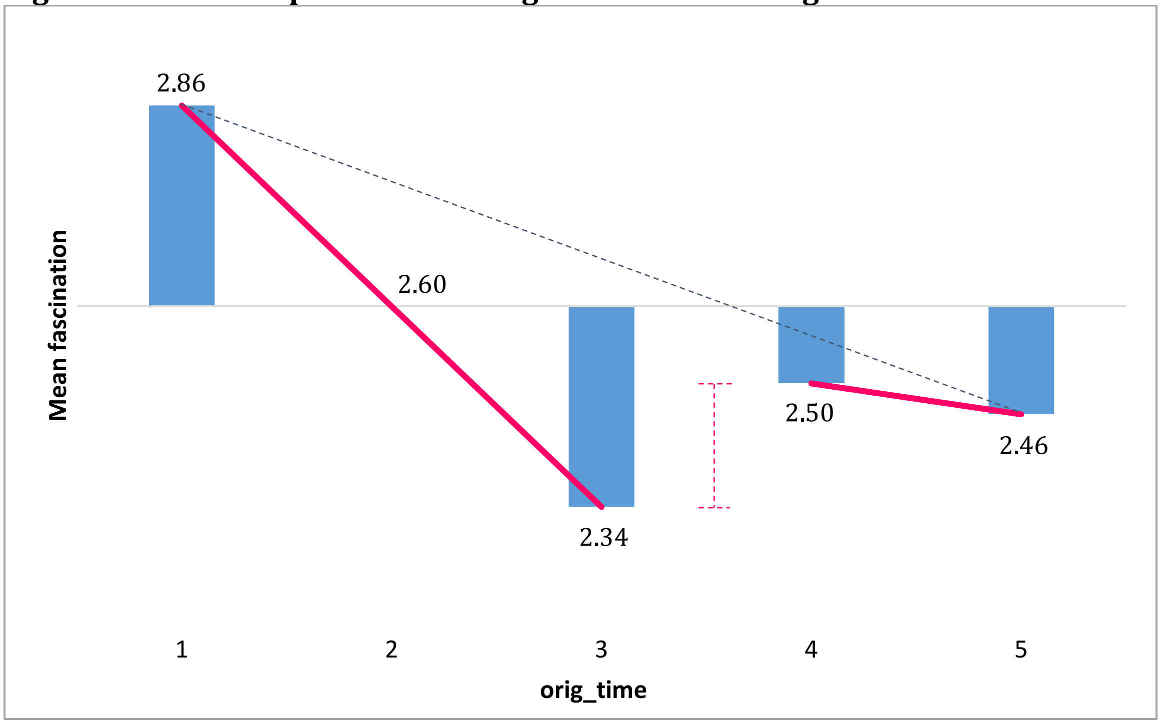
Using the formulas above, Figures 5.1 and 5.2 below detail the predicted average change in fascination over time for grade 6 students and grade 8 students, respectively.

Visible in Figure 5.1, grade 6 predicted average fascination scores are at their peak of 2.86 (out of 4.00) at the start of Year 1, which is the first measurement occasion (orig_time=1, *time*=-1). Compared to their lower scoring peers, sixth grade students with mean fascination scores near 2.86 are considerably more likely to

[4] Additional variance at level 3 could also be discussed, pertaining to differences between classes (u_{10j}, u_{20j}, and u_{30j}). However, for the purposes of addressing RQ1 and the variance in change in fascination between students, this section solely includes the level 2 variance components to focus on differences between students.

endorse items such as “After a really a really interesting science activity is over, I look for more information about it,” and “I want to know everything about science.” (See Appendix A1 for full *Fascination in Science* scale.) Throughout the school year, scores then decrease, bottoming out at 2.34 by the end of grade 6 (orig_time=3, *time*=1). At the start of grade 7 (orig_time=4, *time*=2), scores rebound slightly to 2.50 but continue to fall to 2.46 by the end of the school year (orig_time=5, *time*=3). As the discontinuous terms (*grade* and *post*) were both found to be significant in the grade 6 model, the dashed vertical red line in the graph represents the change in level (*grade*) between grades 6 and 7 representing 0.16 fascination scale score points. Meanwhile the change in slope (*post*) is evident by the slightly less steep line connecting average predicted fascination between measurement occasions 4 and 5, having been reduced from -0.26 in Year 1 to -0.04 in Year 2. For the average student, the change in fascination from the start of grade 6 to the end of grade 7 is -0.40 scale score points, observed in Figure 5.1 as the dashed black line representing the difference between predicted average fascination at the first measurement occasion (2.86) and the fifth measurement occasion (2.46).

Figure 5.1. Grade 6 predicted average fascination change

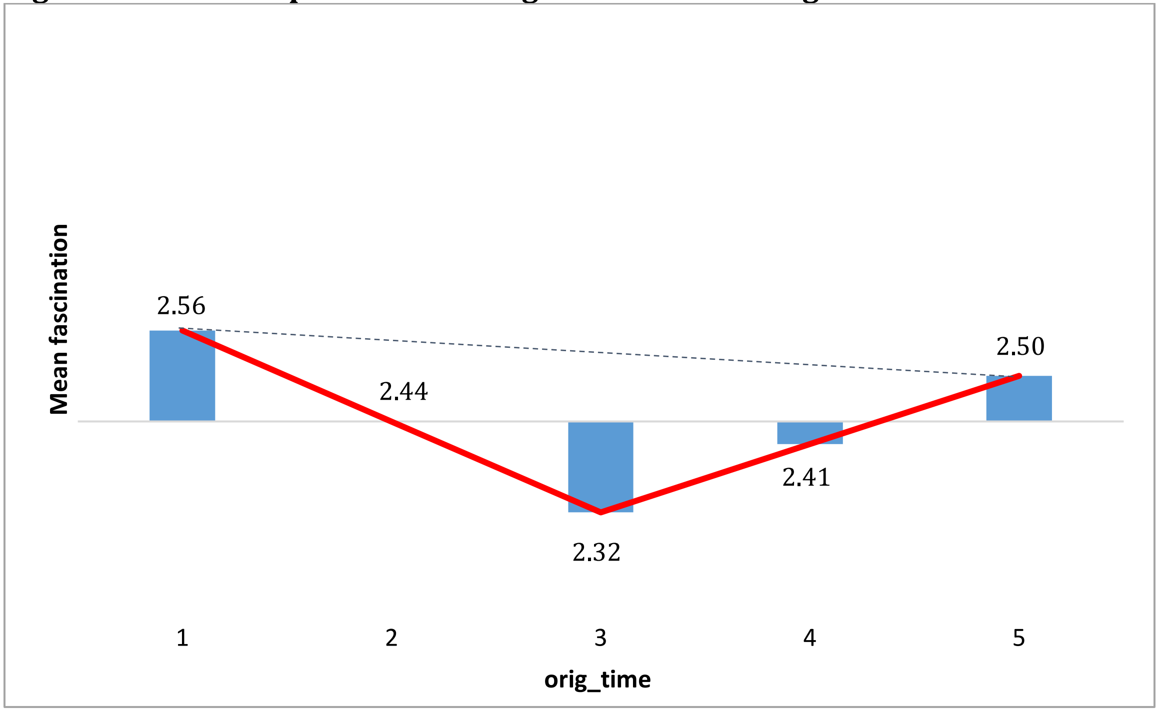


Next, we examine the estimates for the variances of the growth parameters for *time*, *grade*, and *post* at level 2. Model 4.3' in Table 4.4 reveals that, of these three random effects, two (r_{2ij} and r_{3ij}) were not retained in the final modeling of functional form at L1. This indicates that the resulting χ^2 statistic for each homogeneity of variance test for these terms was sufficiently low, that the null hypotheses specified in Chapter 3 pertaining to the L2 variance-covariance matrices were retained (e.g., $H_0: \tau_{\pi_{20}} = 0$), and that there is no true variation in those growth parameters. Table 4.4 also reveals that the remaining variance estimate (r_{1ij}) *was* retained in modeling of the final functional form, and that the opposite is true: the χ^2 test of homogeneity was found to be statistically significant (p<.05), and the null hypothesis that the variation in the growth parameter is zero is rejected. The conclusion is that there is significant student-to-student variation in the linear component of fascination growth in Year 1, but not in the discontinuous nature of change in level or change in slope. Table 4.4 shows the variance estimate for the *time* linear growth parameter, $\hat{\tau}_{\pi_{10}}$, to be 0.016, the square root of which produces an estimated standard deviation of 0.126. Thus, a sixth grade student whose change in fascination is one standard deviation above average is expected to have a flatter, but still negative slope of -0.13 between measurement occasions in Year 1 (-0.259 + 0.126 = -0.133). With no significant variation in observed change in level or change in slope, this produces a total change in fascination from the start of grade 6 to the end of grade 7 of -0.15 scale score points. Meanwhile, a sixth grade student whose change in fascination is one standard deviation below average would be expected to have a corresponding drop in fascination in Year 1 of 0.25 scale score points, resulting in a total change in fascination across the two years of -0.65 scale score points.

Given this variance in estimated average change in fascination in science scores (s.d.=0.126), the observed difference of -0.40 scale points over the two year study represents quite a loss in fascination—a drop of more than three standard deviations in growth. In fact, at the beginning of Year 1, fewer than 24% of grade 6 students had mean fascination scores of 2.46 or lower.

Focusing now on eighth graders, Figure 5.2 shows the graphical results of equation 5.2 and predicted change in fascination scores from the start of grade 8 to the end of grade 9. We see that predicted average fascination scores begin at 2.56 (out of 4.00) at the start of Year 1 (orig_time=1, *time*=-1). Compared to their lower scoring peers, eighth grade students with mean fascination scores near 2.56 are considerably more likely to endorse the two items on the fascination scale pertaining to the subdimension of interest: "In general, when I work on science I…" and "In general, I find science…" (See Appendix A1 for full *Fascination in Science* scale). Similar to the pattern observed at grade 6, scores decrease to their lowest observed point at the end of the first school year, 2.32 (orig_time=3, *time*=1). Although there is no change-in-level term (*grade*) representing a discontinuous intercept at the start of Year 2, the change-in-slope term (*post*) reverses the pattern of decreasing fascination scores in an upward linear fashion throughout grade 9. The positive Year 2 slope does not completely restore fascination scores to their Year 1 starting point by the end of the year but comes close, with an expected average of 2.50. This reflects a Year 2 slope of 0.09, a stark contrast to the Year 1 observed slope of -0.12. For the average student, the change in fascination from the start of grade 8 to the end of grade 9 is -0.06 scale score points, observed in Figure 5.2 as the dashed black line representing the difference between predicted average fascination at the first measurement occasion (2.56) and the fifth measurement occasion (2.50).

Figure 5.2. Grade 8 predicted average fascination change



Finally, we examine the grade 8 estimates for the variances of the growth parameters for *time* and *post* at level 2 (*grade* was not retained in the final functional form model). Model 4.3'' in Table 4.5 shows that the random effect for the discontinuous term *post* (r_{2ij}) was not statistically significant, indicating there is no true variation in Year 2 slope between students. As the random effect estimate associated with the Year 1 slope variable *time*, r_{1ij}, was retained in final modeling, we once again conclude that there is significant individual variation in fascination growth in Year 1, but not in the discontinuous nature of change in slope. Table 4.5 shows the variance estimate for the *time* growth parameter, $\hat{\tau}_{\pi_{10}}$, to be 0.021, the square root of which produces an estimated standard deviation of 0.145. Thus, an eighth grade student whose change in fascination is one standard deviation above average is actually expected to have a positive slope between measurement

occasions in Year 1 (-0.116 + 0.145 = 0.029), ending middle school with a predicted average fascination score of 2.61. With no significant variation in observed change in slope in Year 2, this produces a total change in fascination from the start of grade 8 to the end of grade 9 of 0.23 scale score points (0.29 higher than the average student). Meanwhile, a sixth grade student whose change in fascination is one standard deviation below average would be expected to have a corresponding drop in fascination in Year 1 of 0.29 scale score points, resulting in a total change in fascination across the two years of
-0.35 scale score points.

Although disappointing to observe another downward trend in fascination in science, the predicted drop of 0.06 scale score points on average for eighth graders is far less severe than the predicted loss for the average sixth grader. This change in fascination represents less than half a standard deviation in growth, and 49% of grade 8 students had mean fascination scores of 2.50 or lower at the beginning of Year 1.

To summarize and more succinctly answer the first research question, the average change in science fascination from the beginning of sixth to the end of seventh grade is predicted to be -0.40 scale score points, with a standard deviation of 0.126. The average change in science fascination from the beginning of eighth grade to the end of ninth grade is predicted to be -0.06 scale score points, with a standard deviation of 0.145. While not particularly promising at either grade level, the pattern of change in fascination for the grade 8 cohort is considerably more optimistic than that for the grade 6 cohort, particularly when considering the increase in fascination scores observed in Year 2.

Influence of Classroom Instruction on Students' Science Fascination

The second research question asked how classroom-level variables influence student fascination scores:

> *To what extent is predicted student science fascination associated with the instructional characteristics of the classroom, including student-centric teaching, hands-on methods, and classroom technology use? And are these effects sustained over time?*

Before looking at model results to determine the relationship of the level 3 classroom predictors with student fascination in science, it is worth revisiting the calculations of the intraclass correlation coefficients based on the variance estimates produced by the initial null models. Specifically, the L3 ICC for grade 6 was found to be 0.015, and for grade 8 was found to be 0.021. In both cases, the proportion of variance attributable to the classroom nesting structure was minimal, roughly 2%. Thus, any instructional characteristics that play a role in influencing student fascination would only be able to account for a very small proportion of variance explained in fascination scores. Keeping this in mind, it is therefore not surprising to see that Tables 4.10 and 4.12 show very few significant relationships between the level 3 predictors—student-centric teaching, hands-on methods, and classroom technology use—and student fascination in science.

For 6th graders, each variable was found to have only one significant relationship. Specifically, student-centric teaching (*student*) was found to have a significant relationship with the variable *female* and its effect on the intercept term (γ_{021} = -0.20, *se* = 0.09, t(953) = -2.33, p = 0.02), indicating that, while above average student-centric teaching time was not found to correlate with higher fascination scores, increases in student-centric teaching above the grand mean lead to additional decreases in expected fascination for girls, on average. Hands-on methods and classroom technology use were each found to have a significant relationship with the level 2 family support variable and its effect on the Year 2 *post* variable (*handson:* γ_{311} = -0.54, *se* = 0.12, t(959) = -4.40, $p < 0.01$; *classtech:* γ_{312} = -0.75, *se* = 0.26, t(959) = -2.87, $p < 0.01$). These results indicate that, following the summer discontinuity, 7th graders with above grand-mean scores for family support and above average classroom hands-on instruction or classroom technology use are expected to see *decreases* in fascination scores. Although statistically significant, the standard errors for each of these terms were high relative to the other *se* estimates in the final model, and given the magnitude of the estimated coefficient parameters, the implication is that each of these terms is estimated with less precision than the other terms retained in the final model.

For 8th graders, only one level 3 predictor was retained in the final model, *classtech*. The single statistically significant relationship detected was an interaction with the level 2 family support variable and its effect on the intercept term ($\gamma_{021} = 0.61, se = 0.22, t(1200) = 2.83, p = 0.005$). These results indicate that students with above average family support who receive more than the grand-mean of instructional classroom technology use have a corresponding increase in predicted fascination scores mid-way through Year 1 (*time*=0).

Overall, these results suggest that classroom instructional characteristics are not meaningfully associated with predicted science fascination. Despite a few significant relationships found in the final models at both grade levels, the influence on changes in student fascination scores are minimal given the low proportions of variance attributable to classroom-level nesting. This is most noticeable when reviewing the L3 pseudo R^2 statistics for each of the final models and their contributions toward reducing previously unexplained variance in predicted fascination scores. The negative L3 pseudo R^2 calculated for the final grade 6 model is uninterpretable, while the pseudo R^2 calculated for the final grade 8 model of 0.22 translates to a reduction of only 2% in previously unexplained variance.

To the extent that significant relationships were observed at either grade level, the effects were not sustained over time. For both grade levels, there were associations with initial predicted fascination scores (i.e., the intercepts) and diminishing effects over time based on interactions with the discontinuous Year 2 terms. This suggests that, however minimal the observed impacts of instructional characteristics on fascination scores are in Year 1, they quickly fade as students enter new classrooms and new grades in Year 2.

Student Characteristics and their Relationship to Students' Science Fascination

The third and final research question posed in this dissertation asked how student-level variables influence student fascination scores:

> *To what extent are middle schools equally successful in sustaining science fascination for students whose gender, race, family support for science, or economic backgrounds differ?*

If schools were in fact *equally successful* in sustaining science fascination for students whose individual characteristics differed (in terms of gender, race, family support for science, or economic backgrounds), we would find no significant relationships between the modeled L2 predictors and predicted fascination scores. Unlike the nesting structure of classrooms at level 3, close to 50% of the variation in fascination scores at both grade levels was found to be related to person-to-person differences modeled at level 2, providing more opportunity for modeled student-level variables to find patterns of association.

At grade 6, several significant relationships were found, exclusively with variables related to family support for science (*FS*) and gender (*female*). Specifically, both covariates were found to have an effect on initial predicted fasciation values (*FS:* γ_{010} = 0.16, *se* = 0.03, $t(953)$ = 5.33, $p < 0.01$; *girl:* γ_{020} = -0.11, *se* = 0.02, $t(953)$ = -5.63, $p < 0.01$), while family support alone was found to have a continuing impact over time on the Y1 growth term *time* (γ_{110} = -0.09, *se* = 0.02, $t(953)$ = -5.86,

p < 0.01) and the Y2 change-in-slope term *post* (γ_{310} = 0.20, *se* = 0.04, t(959) = 5.25, p < 0.01). As noted earlier, the association between *female* and *fascination* indicates that, on average, sixth grade girls have an expected initial fascination score that is 0.11 points (or $^{2}/_{10}$ s.d.) lower, and compared to boys, this reduction holds throughout the two-year period. And though the influence of *FS* fluctuates across measurement occasions due to the discontinuous model, students who have above grand-mean family support have higher fascinations scores, on average, at each observed time point.

At grade 8, similar relationships were found between predicted fascination scores and student-level variables. Family support for science and gender once again had an effect on the intercept term (*FS:* γ_{020} = 0.11, *se* = 0.02, t(1200) = 4.93, p < 0.01; *girl:* γ_{030} = -0.12, *se* = 0.02, t(1200) = -5.55, p < 0.01), as did home resources (γ_{010} = 0.08, *se* = 0.02, t(1200) = 3.38, p < 0.01). Home resources and family support for science were also found to have a relationship with the Year 1 slope variable, *time* (*HR:* γ_{110} = -0.02, *se* = 0.01, t(1200) = -2.08, p = 0.04; *FS:* γ_{120} = -0.05, *se* = 0.01, t(1200) = -3.55, p < 0.01). In all instances, students with scores above the grand mean for family support for science and home resources are predicted to have higher fascination scores than their peers, on average, although the influence of *HR* and *FS* wanes over time, as noted in Chapter 4.

These results highlight some important trends across grade levels. First, as the variable *female* was found to be significantly and negatively related to initial fascination scores in both final models, it is evident that prior experiences lead females to begin each year with lower levels of fascination in science. In addition, differences in the level of fascination in science stay the same over the two-year period of study, all other things being equal. While the present analyses do not support claims about the impact prior schooling has had on students' levels of fascination in science, these results do suggest that schools are not effective at narrowing or eliminating these differences in the level of fascination in science for middle school students entering high school. Girls have consistently lower fascination scores at all time points in both grade level models, allowing us to conclude that schools are not equally successful in supporting fascination for students whose gender differs.

Secondly, home background variables were also found to be significantly related to fascination scores, most notably family support for science. For both sixth and eighth graders, fascination scores were higher for students with above average family support for science, a pattern that persists into grades 7 and 9 to a certain degree. While not necessarily surprising, this suggests that schools also are not equally successful in supporting for students with less familial support, and that students who experience lower levels of support at home could be better supported at school to mitigate these differences.

Finally, the race variable *URM* was not found to have a significant relationship with any of the modeled characteristics of change in fascination over time. At face value, this indicates that students of all racial backgrounds are equally supported by schools in terms of science fascination, and that the impact schooling has on students' fascination in science seems to vary more for other characteristics previously mentioned—gender, family support for science, and home resources.

Summary of Findings

In addressing the first research question regarding the average change in science fascination of middle school students, the observed average change for the sixth grade cohort was found to be -0.40 scale score points, representing a drop of more than three standard deviations in the growth parameter (0.126). In contrast, the average change in fascination scores for the eighth grade cohort was found to be -0.06 scale score points, a minimal drop in comparison to the observed standard deviation in growth of 0.145. In addressing the second research question regarding the relationship between instructional characteristics and predicted student science fascination, meaningful associations were not observed due to the low proportion of variance in fascination scores at the classroom level. In addressing the third and final research question regarding student-level covariates and their relationships to changes in science fascination, a significant negative association was found with the gender variable *female* in both grade-level models, indicating that females begin with lower levels of fascination in science than boys at each grade; these observed differences remained throughout the two-year study. Family support for science was also found to be significantly related to fascination, and students with above average family support in both grades were observed to have higher average fascination scores than their peers with lower family support. No significant relationship was observed between the race variable, *underrepresented minority* (*URM*), and changes in fascination over time.

Limitations

This dissertation attempted to provide a rigorous approach to addressing the stated research questions, including the use of high-quality data and the employment of an appropriate methodology, although there are several caveats to the inferences that can be drawn from this work. Among them are subjective decisions made about variable coding that could have influenced results and measurement issues stemming from secondary analyses, each of which will be discussed in more detail.

The decision to recode racial/ethnic data into a singular *underrepresented minority* (*URM*) variable was made to both simplify the level 2 model development process by reducing the number of prospective variables, while also strongly aligning with theory helping to guide this work. The NSF definition provided in the previous chapter—"…groups whose representation in S&E education or employment is smaller than their representation in the U.S. population" (NSF, 2021b)—is precisely the consideration this variable and its coding intended to provide in the modeling of fascination in science. Given the complexity of true racial/ethnic backgrounds and the multitude of combinations that exist from an identity standpoint, the possibility of variable coding was quite vast, despite the search for parsimony in model development. Still, it is striking that the results presented in both grade-level models suggest no statistical relationship between (dummy-coded) race and fascination scores in science, contrary to previous work (e.g., Archer et al., 2012; Catsambis, 1995). It stands to reason that the *URM* variable could be misspecified, while alternate coding schemes would have highlighted significant relationships that went undetected with the binary coding scheme utilized in the developed models.

Limitations imposed by the initial research design raise questions and offer guidance for future work. Specifically, end-of-year fascination scores in Year 1 for both grade levels are low enough to raise suspicions about the timing of measurements in relation to state-imposed standardized testing in the two states from which schools participated, California and Pennsylvania. There are well-documented negative impacts of high-stakes tests which impact the affective side of learning (e.g., Madaus & Russell, 2009/2010), and measurement occasions even prior to, but in anticipation of, standardized tests could artificially deflate fascination scores, potentially differentially for subgroups of students. However, because there was no definitive indication that this was universally the case, there was no reason to exclude this measurement occasion and drastically reduce the usable sample; future studies should ensure end-of-year measures are not influenced by assessment schedules.

Relatedly, the magnitude of attrition entering Year 2 of the study, although potentially unavoidable, raises additional questions about patterns of missingness (recall that 76% of the final sample had no Year 2 fascination scores). As detailed in Chapter 3, HLM remains a robust solution for repeated-measures studies, requiring only one observation at L1 for cases to be retained in the sample. Appendix D1 presents a comparison of students who participated only in Year 1 of the study against those who participated in both years. Importantly, the pattern of change depicted in Year 1 looks similar: both groups start with their highest observed fascination scores at the outset and decrease to almost identical scale points by the end of the first school year. However, the gap between the two lines suggests the possibility that students who opted out of participation in Year 2 may have been those with lower fascination in science. While some amount of attrition should be expected, future studies should work to mitigate the loss of individual participation as much as possible.

Of course, secondary analyses also remain limited in that there remains the possibility of unmodeled yet significant relationships relating to changes in fascination scores due to variables unavailable in the data. ALES:14 was an ambitious endeavor based on the triangulation of instruments (student measures, teacher logs, researcher observations), the length of study (two school years, five measurement occasions), and the rigor and length of instrumentation (three Activation constructs in addition to fascination along with measures of success; see Figure 1.1). And while this dissertation intently focused on fascination and variables related based on prior research, extending analyses to include a wider range of potential influences on fascination beyond what was included in the dataset remains a critical next step.

While the discontinuous models selected as "final" in equations 4.5' and 4.5'' were both significant improvements from the unspecified starting growth model (grade 6, χ^2=281.1, *df*=5, *p*<.001; grade 8, χ^2=188.2, *df*=1, *p*<.001), they objectively fail to capture the majority of variance in fascination scores. To analyze the total variance in fascination scores explained by the final models, we subtract the summed variance components of the conditional models (4.5' and 4.5'') from the summed variance components of the unconditional models (4.1) and compute a ratio to the total unconditional variance, as shown in equation 5.3:

$$\text{Prop. variance explained: } \frac{(\hat{\sigma}^2(uncon.)+\hat{\tau}_\pi(uncon.)+\hat{\tau}_\beta(uncon.))-(\hat{\sigma}^2(cond.)+\hat{\tau}_\pi(cond.)+\hat{\tau}_\beta(cond.))}{(\hat{\sigma}^2(uncon.)+\hat{\tau}_\pi(uncon.)+\hat{\tau}_\beta(uncon.))} \tag{5.3}$$

For grade 6: $$\frac{(.13424+.12157+.00396)-(.11780+.11572+.00408)}{(.13424+.12157+.00396)} = 0.0853$$

For grade 8: $$\frac{(.11084+.10536+.00464)-(.10644+.09704+.00360)}{(.11084+.10536+.00464)} = 0.0623$$

Here, we observe that the inclusion of discontinuous growth terms and statistically significant individual- and class-level covariates only explain 8.5% of the total variance in fascination scores at grade 6, and even less (6.2%) at grade 8. It is evident that additional explanatory variables are necessary to more adequately model change in fascination over time.

Finally, it should be noted that, despite the appropriateness of HLM in investigating change in fascination scores over time, the method itself does not identify causal relationships, and this work has not identified any "effects" leading to higher fascination scores, per se. As a descriptive study, the findings reported are suggestive and not definitive, in the hopes that future research can build on this dissertation work and continue to identify relationships that positively influence science fascination.

Implications

Despite the limitations noted above, this dissertation provides important insight into changes in middle school students' fascination in science. The purpose of this research was to investigate potential patterns and magnitudes of change in fascination scores, while identifying student and classroom covariates that may be related to change over time. The study sought to address three research questions using a large, longitudinal sample collected through the Activated Learning Enables Success study (ALES:14), while making connections to earlier research done on the primary subdimension of fascination, interest. Results indicate that discontinuous growth models fit the data significantly better than linear growth models, highlighted by changes in level and slope in Year 2 for sixth graders entering seventh grade, and change in slope alone for eighth graders entering high school. These same results suggest that, on average, student fascination in science decreases over time, regardless of grade level.

Were this study limited to two measurement occasions in its analysis of change over time—say, the beginning of Year 1 and end of Year 1, or even the beginning of Year 1 and end of Year 2—a natural conclusion might be limited to repeat what has been said before. School-based experiences repress students' opportunities for pursuing curiosity (e.g., Engel, 2011); students lose interest in school-based science right around the time they enter high school (Barmby, Kind, & Jones, 2008; Christidou, 2011; Krapp & Prenzel, 2011); students become disenfranchised and lose interest over time (Gottfried, Fleming, & Gottfried, 2001; Osborne, Simon, & Collins, 2003; Simpson & Oliver, 1990). However, with five measurement occasions and discontinuous growth, it is apparent that changes in fascination in science are more complex for middle schoolers.

Most notably, the observed patterns of change at both grade levels suggest that science fascination scores "rebound" in some way after Year 1. Newly initiated 7th graders display an increase in fascination in science following their summer vacation compared to their end-of-year 6th grade selves (even if that "new shine" wears off over the course of the year), while 9th graders begin to see increases in fascination scores contrary to their final year of middle school. While these findings do not contradict prior findings such as "students lose interest in school-based science right around the time they enter high school" since fascination scores continuously declines throughout grade 8, it actually appears as though entering high school students have increasing fascination in science. Additionally, while early middle school students show declining fascination scores throughout grade 6 and again in grade 7, these declines are somewhat mitigated by their summer experiences. Therefore, future research should focus on out-of-school experiences that bolster fascination in science, particularly over summer break, while studying the elements of ninth grade (e.g., factors related to a new school building, new students, choice in classes, etc.) that contribute to increased fascination in high school.

Still, these results may be somewhat disheartening from an educational policy perspective: in grades 6 and 8, fascination scores, on average, decline in a linear fashion from the start of the academic year to its end. The instructional characteristics modeled in this study were not found to augment these losses in a substantial way, and even prior to building the level 3 model including classroom covariates, the proportion of variance in fascination scores available to be "explained" through teaching or classroom nesting was minimal (~2%). Future research should place emphasis on the context of instruction and academic content relative to individuals' situational interest (e.g., Hidi & Harackiewicz, 2000), rather than simply focusing on gross instructional categories.

This work also supports earlier findings related to the role parents and caregivers play in stimulating fascination in science (e.g., George & Kaplan, 1998), as family support for science was consistently found to have an association with increased fascination scores, and at grade 8, home resources was as well. These observed relationships, although intuitive, are critical in helping to steer action on the part of parents and even science educators who are able to provide guidance to caregivers of middle-schoolers: support your child's science learning at home. Not only were above average family support for science scores found to be associated with higher fascination scores, but prior research has shown additional relationships between attitudinal aspects like feelings toward science and interest and advanced science achievement (e.g., Mullis et al., 2020; Harackiewicz, Barron, Tauer, & Elliot 2002).

Another problematic finding supported by this work, as suggested in prior research (e.g., Archer et al., 2012; Baram-Tsabari & Yarden, 2011), is the disadvantage faced by middle school girls in terms of science fascination. Models for both sixth and eighth grade students demonstrate the negative relationship between gender and initial fascination status, with no observed closure in this science fascination gap over time. The implication here is that more must be done, in and out of school, to ensure that girls don't prematurely decide that science isn't for them based on lower interest, curiosity, or mastery goals.

Given the important role that these student characteristics play in observed changes in science fascination, future research should also explore potential within-level interactions. While the models developed in this dissertation explicitly account for cross-level interactions (e.g., the influence of the level 3 student-centric teaching variable on the level 2 gender variable's effect on the level 1 intercept), they do not account for effects specific to combinations of within-level terms. In particular, the intersection of racialized identity and family support for science should be explored to determine if differing levels of at-home support influence the science fascination of students of all backgrounds equally.

Inevitably, increasing fascination in science alone is not enough to improve levels of scientific literacy across the populace vastly. The theory of Science Learning Activation appropriately implicates additional interrelated dispositions and skills as necessary in driving success through choice, engagement, perceived success, and science learning (see Figure 1.1). Future research using methodologies capable of testing relationships among multiple outcomes, such as Structural Equation Modeling, may prove fruitful in building on this dissertation work by incorporating other Activation constructs.

Particularly when considering the current state of the world and broad outcomes many see as desirable (such as fighting the impacts of climate change), it is helpful to revisit the OECD's definition of scientific literacy: "the capacity…to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity" (National Science Board, 2012, pp. 132-33). Students with increased levels of fascination in science are more likely to continue choosing participation in science activities when given the chance, and in turn, should be more prone to making evidence-based claims when presented with socio-scientific decisions in the future. Understanding when fascination is likely to change as children develop is critical to our efforts to improve science-related outcomes.

References

Ainley, M., & Ainley, J. (2011). Student Engagement with Science in Early Adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychologist, 36*, 4-12. doi:10.1016/j.cedpsych.2010.08.001

Anderson, D. (2012). *Hierarchical Linear Modeling (HLM): An Introduction to Key Concepts Within Cross-Sectional and Growth Modeling Frameworks* (Technical Report #1308). Eugene, OR: Behavioral Research and Teaching.

Andrich, D. (1982). An Extension of the Rasch Model for Ratings Providing Both Location and Dispersion Parameters. *Psychometrika, 47*(1), 105-113.

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2012). "Balancing acts": Elementary school girls' negotiations of femininity, achievement, and science. *Science Education, 96*, 967-989. doi: 10.1002/sce.21031.

Baldwin, J. (1911). *Thought and Things: A study of the development and meaning of thought* (Vol. 3). New York: Macmillan.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 191-215.

Baram-Tsabari, A., & Yarden, A. (2011). Quantifying the Gender Gap in Science Interests. *International Journal of Science and Mathematics Education, 9*(3), 523-550.

Barmby, P., Kind, P., & Jones, K. (2008). Examining Changing Attitudes in Secondary School Science. *International Journal of Science Education, 30*(8), 1075-1093. doi: 10.1080/0950069070134966.

Bathgate, M. & Schunn, C. (2016). Disentangling Intensity from Breadth of Science Interest: What predicts learning behaviors? *Instructional Science, 44,* 423-440. doi: 10.1007/s11251-016-9382-0.

Bathgate, M., & Schunn, C. (2017). Factors that Deepen or Attenuate Decline of Science Utility Value During the Middle School Years. *Contemporary Educational Psychology, 49*, 215-225. doi: 10.1016/j.cedpsych.2017.02.005.

Bickel, R. (2007). Multilevel Analysis for Applied Research: It's Just Regression! New York, NY: The Guilford Press.

Boedeker, P. (2017). Hierarchical Linear Modeling with Maximum Likelihood, Restricted Maximum Likelihood, and Fully Bayesian Estimation. *Practial Assessment, Research, and Evaluation, 22*(2). doi: 10.7275/5vvy-8613.

Bonnette, R., Crowley, K., & Schunn, C. (2019). Falling in love and staying in love with science: ongoing informal science experiences support fascination for all children. *International Journal of Science Education, 41*(12), 1626-1643. doi: 10.1080/09500693.2019.1623431.

Burke, E. (1757). *A Philosophical Enquiry into the Origin of our Ideas of the Sublime and Beautiful*. Vol. XXIV, Part 2. The Harvard Classics. New York: P.F. Collier & Son, 1909-14. Bartleby.com, 2001. www.bartleby.com/24/2/101.html.

Byrnes, J., & Miller, D. (2007). The Relative Importance of Predictors in Math and Science Achievement: An opportunity-propensity analysis. *Contemporary Educational Psychology, 32,* 599-629. doi: 10.1016/j.cedpsych.2006.09.002.

Catsambis, S. (1995). Gender, Race, Ethnicity, and Science Education in the Middle Grades. *Journal of Research in Science Teaching, 32*(3), 243-257. doi: 10.1002/tea.3660320305.

Christidou, V. (2011). Interest, attitudes and images related to science: Combining students' voices with the voices of school science, teachers, and popular science. *International Journal of Environmental & Science Education, 6*(2), 141-159.

Chung, J., Cannady, M., Schunn, C., Dorph, R., Bathgate, M. (2016a). Measures Technical Brief: Comptency Beliefs in Science (v. 3.2). Retrieved from http://activationlab.org/wp-content/uploads/2018/03/Competency-Beliefs-Report-3.2-20160331.pdf.

Chung, J., Cannady, M., Schunn, C., Dorph, R., Bathgate, M. (2016b). Measures Technical Brief: Fascination in Science (v. 3.2). Retrieved from http://activationlab.org/wp-content/uploads/2018/03/Fascination-Report-3.2-20160331.pdf.

Chung, J., Cannady, M., Schunn, C., Dorph, R., Bathgate, M. (2016c). Measures Technical Brief: Scientific Sensemaking (v. 3.2). Retrieved from http://activationlab.org/wp-content/uploads/2018/03/Sensemaking-Report-3.2-20160331.pdf.

Chung, J., Cannady, M., Schunn, C., Dorph, R., Bathgate, M. (2017). Measures Technical Brief: Valuing Science (v. 3.3). Retrieved from http://activationlab.org/wp-content/uploads/2018/04/Values-Report-3.3-20170306.pdf.

Covington, M. (2000a). Goal Theory, Motivation, and School Achievement: An integrative review. *Annual Review of Psychology, 51,* 171-200.

Covington, M. (2000b). Intrinsic versus Extrinsic Motivation in Schools: A reconciliation. *Current Directions in Psychological Science, 9*(1), 22-25.

DeBoer, G. (2000). Scientific Literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching, 37*(6), 582-601.

Dewey, J. (1910). Science as Subject-Matter and as Method. *Science, 31*(787), 121-127. http://www.jstor.org/stable/1634781.

Dewey, J. (1913). *Interest and Effort in Education*. Boston, MA: Houghton Mifflin Co.

Dorph, R. (2016). *Session Overview: Critical Science Disposition and Skills that Evolve in the Middle School Years* [Conference presentation]. American Educational Research Association (AERA) 2016 Annual Meeting, Washington, DC, United States. http://activationlab.org/wp-content/uploads/2018/03/session-overview.pdf.

Dorph, R., Cannady, M., & Schunn, C. (2016). How Science Learning Activation Enables Success for Youth in Science Learning. *Electronic Journal of Science Education, 20*(8), 49-85.

Dorph, R., Shields, P., Tiffany-Morales, J., Hartry, A., McCaffrey, T. (2011). High hopes – few opportunities: The status of elementary science education in California. Sacramento, CA: The Center for the Future of Teaching and Learning at WestEd.

Dreschel, B., Carstensen, C., & Prenzel, M. (2011). The Role of Content and Context in PISA Interest Scales: A study of the embedded interest items in the PISA 2006 science assessment. *International Journal of Science Education, 33*(1), 73-95. doi: 10.1080/09500693.2010.518646.

Driver, R. (1985). Children's Ideas in Science. Philadelphia: Open University Press.

DuBois, W.E.B. (2013). *W.E.B. DuBois on Sociology and the Black Community*. Chicago, IL: University of Chicago Press. (Original work published 1920.)

Dumas, D., & McNeish, D. (2017). Dynamic Measurement Modeling: Using nonlinear growth models to estimate student learning capacity. *Educational Researcher, 46*(6): 284-292. doi: 10.3102/0013189X17725747.

Eccles, J., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology, 53*, 109-132.

Eckblad, G. (1981). *Scheme Theory: A conceptual framework for cognitive-motivational processes.* New York: Academic Press.

Engel, S. (2011). Children's Need to Know: Curiosity in schools. *Harvard Educational Review, 81*, 625-645.

"fascination," (n.d.). Oxford Living Dictionaries. June 2018. Oxford University Press. https://en.oxforddictionaries.com/definition/fascination (accessed June 23, 2018).

Fayer, S., Lacey, A., & Watson, A. (2017). STEM Occupations: Past, Present, and Future. United States Bureau of Labor Statistics. Retrieved from https://www.bls.gov/spotlight/2017/science-technology-engineering-and-mathematics-stem-occupations-past-present-and-future/pdf/science-technology-engineering-and-mathematics-stem-occupations-past-present-and-future.pdf.

Gardner, D. (1983). A Nation At Risk: The imperative for educational reform. Washington, DC: U.S. Government Printing Office.

Gardner, P. (1987). Comments on "Towards the development of a children's science curiosity scale". *Journal of Research in Science Teaching, 24*(2), 175-176.

Gatewood, C., & Obourn, E. (1963). Improving Science Education in the United States: A paper presented at the commonwealth conference on the teaching of science in schools. Ceylon, December 9-21, 1963. *Journal of Research in Science Teaching, 1*(4), 355-399. doi:10.1002/tea.3660010417.

Geesa, R.L., Izci, B., Song, H.S., & Chen, S. (2019). Exploring the roles of students' home resources and attitudes towards science in science achievement: a comparison of South Korea, Turkey, and the United States in TIMSS 2015. *Asia-Pacific Science Education, 5*(17). doi: 10.1186/s41029-019-0038-7.

George, R., & Kaplan, D. (1998). A Structural Model of Parent and Teacher Influences on Science Attitudes of Eighth Graders: Evidence from NELS:88. *Science Education, 82*(1), 93-109.

Gottfried, A., Fleming, J., & Gottfried, A. (2001). Continuity of Academic Intrinsic Motivation From Childhood Through Late Adolescence: A longitudinal study. *Journal of Educational Psychology, 93*(1), 3-13. doi: 10.1037//0022-0663.93.1.3.

Hadzigeorgiou, Y. (2016). *Imaginative Thinking in Science and Science Education*. In: Imaginative Science Education. Springer, Cham. doi: 10.1007/978-3-319-29526-8_1.

Harris, M., & Miller, J. (2005). Needed: Reincarnation of the National Defense Education Act of 1958. *Journal of Science Education and Technology, 14*(2), 157-171. doi: 10.1007/sl0956-005-4419-l.

Harackiewicz, J., Barron, K., Tauer, J., & Elliot, A. (2002). Predicting Success in College: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. *Journal of Educational Psychology, 94*, 562-575.

Hasni, A., & Potvin, P. (2015). Student's Interest in Science and Technology and its Relationships with Teaching Methods, Family Context and Self-Efficacy. *International Journal of Environmental & Science Education, 10*(3), 337-366. doi: 10.12973/ijese.2015.249a.

Hidi, S. (2006). Interest: A unique motivational variable. *Educational Research Review, 1,* 69-82. doi: 10.1016/j.edurev.2006.09.001.

Hidi, S., & Harackiewicz, J. (2000). Motivating the Academically Unmotivated: A critical issue for the 21st century. *Review of Educational Research, 70*(2), 151-179.

Hidi, S., & Renninger, K. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist, 41*(2), 111-127.

Hoeg, D.G., & Bencze, J.L. (2017). Values Underpinning STEM Education in the USA: An analysis of the Next Generation Science Standards. *Science Education, 101*(2): 278-301. doi: 10.1002/sce.21260.

Holden, J.E., Kelley, K., & Agarwal, R. (2008). Analyzing Change: A Primer on Multilevel Models with Applications to Nephrology. *American Journal of Nephrology, 2008*(28): 792-801. doi: 10.1159/000131102.

Holland, P. (2003). *Causation and Race* (Report RR-03-03). Educational Testing Service. https://www.ets.org/Media/Research/pdf/RR-03-03-Holland.pdf.

Huta, V. (2014). When to Use Hierarchical Linear Modeling. *The Quantitative Methods for Psychology, 10*(1), 13-28. doi: 10.20982/tqmp.10.1.p013.

"interest," (n.d.). Oxford Living Dictionaries. June 2018. Oxford University Press. https://en.oxforddictionaries.com/definition/interest (accessed May 30, 2018).

Izard, C. (1979). Emotions as Motivations: An evolutionary developmental perspective. *Nebraska Symposium on Motivation, 1978*, (pp. 163-200). Lincoln: University of Nebraska Press.

James, W. (1890). *The Principles of Psychology.* London: Macmillan.

Jacobs-Priebe, L. & Crowley, K. (2013). Women in science: Activation, barriers, and persistence. Poster presented at AERA Conference, San Francisco, CA.

Kahan, D. (2013). A Risky Science Communication Environment for Vaccines. *Science, 342*, 53-54. doi: 10.1126/science.1245724.

Kang, J., & Keinonen, T. (2018). The Effect of Student-Centered Approaches on Students' Interest and Achievement in Science: Relevant Topic-Based, Open and Guided Inquiry-Based, and Discussion-Based Approaches. *Research in Science Education, 48*. 865-885. doi: 10.1007/s11165-016-9590-2.

Kelly, G. J. (2007). Discourse in science classrooms. In S.K. Abell & N.G. Lederman (Eds.), *Handbook of Research on Science Education* (pp.443-469). Mahwah, NJ: Lawrence Erlbaum Associates.

Krapp, A. (2002). Structural and Dynamic Aspects of Interest Development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction, 12,* 383-409.

Krapp, A. (2005). Basic needs and the development of interest and intrinsic motivational orientations. *Learning and Instruction, 15*, 381-395. doi: 10.1016/j.learninstruc.2005.07.007.

Krapp, A., & Prenzel, M. (2011). Research in Interest on Science: Theories, methods, and findings. *International Journal of Science Education, 33*(1), 27-50. doi: 10.1080/09500693.2010.518645.

Lacey, T., & Wright, B. (2009). "Occupational Employment Projections to 2018," *Monthly Labor Review,* 82-123.

Langsdorf, P., Izard, C., Rayias, M., & Hembree, E. (1983). Interest expression, visual fixation, and heart rate changes in 2- to 8-month-old infants. *Developmental Psychology, 19*, 375-386.

Lin, H., Lawrenz, F., Lin, S., & Hong, Z. (2013). Relationships among affective factors and preferred engagement in science-related activities. *Public Understanding of Science, 22*(8), 941-954. doi: 10.1177/0963662511429412.

Linn, M., Lewis, C., Tsuchida, I., & Songer, N. (2000). Beyond Fourth-Grade Science: Why do U.S. and Japanese students diverge? *Educational Researcher, 29*(3), 4-14.

Litman, J., & Spielberger, C. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of Personality Assessment, 80*(1), 75-86.

Liu, X., & Whitford, M. (2011). Opportunities-to-learn at home: Profiles of students with and without reaching science proficiency. *Journal of Science Education Technology, 20,* 375-387. doi: 10:1007/s10956-010-9259-y.

Loewus, L. (2015). *When Did Science Education Become STEM?* Education Week online. Available at https://www.edweek.org/teaching-learning/when-did-science-education-become-stem/2015/04. Accessed January 24, 2021.

Lohman, D. (2006). Beliefs about Differences between Ability and Accomplishment: From folk theories to cognitive science. *Roeper Review, 29*(1), 32-40.

Lowenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*(1), 75-98.

Madaus, G., & Russell, M. (2009/2010). Paradoxes of High Stakes Testing. *The Journal of Education, 190*(1/2), 21-30. https://www.jstor.org/stable/42744178.

Maltese, A., & Tai, R. (2011). Pipeline Persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education, 95*(5), 877-907.

McNeish, D., & Dumas, D. (2017). Non-Linear Growth Models as Measurement Models: A second-order growth curve model for measuring potential. *Multivariate Behavioral Research, 52*(1), 61-85.

McNeish, D., Stapleton, L., & Silverman, R. (2017). On the Unnecessary Ubiquity of Hierarchical Linear Modeling. *Psychological Methods, 22*(1), 114-140.

Miller, E., Manz, E., Russ, R., Stroupe, D., & Berland, L. (2018). Addressing the Epistemic Elephant in the Room: Epistemic agency and the Next Generation Science Standards. *Journal of Research in Science Teaching, 2018*(00): 1-23. doi: 10.1002/tea.21459.

Miller, J. (2010). *The Challenge of Just-in-Time Science Learning for Museums.* Paper presented at the 2010 Association of Science-Technology Centers Annual Conference, Honolulu, Hawaii.

Moore, D., Bathgate, M., Chung, J., & Cannady, M. (2011). Technical Report: Measuring activation and engagement; Activation Lab, Enables Success Study. Available at: http://activationlab.org/wp-content/resources/aera2013/aera_research/AERA_2013_0_1_ALES2011_TechReport.pdf.

Morgan, P., Farkas, G., Hillemeier, M., & Maczuga, S. (2016). Science Achievement Gaps Begin Very Early, Persist, and Are Largely Explained by Modifiable Factors. *Educational Researcher, 45*(1), 18-35. doi: 10.3102/0013189X16633182.

Mullis, I.V.S., Martin, M.O., Foy, P., Kelly, D.L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science.* Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-results/.

National Academy of Engineering Committee on Technological Literacy. (2002). *Technically Speaking: Why all Americans need to know more about technology.* Washington, DC: The National Academies Press.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2010). *Rising above the gathering storm, revisited*. Washington, DC: National Academies Press.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2011). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads*. Washington, DC: National Academies Press.

National Research Council (2011). *Successful K-12 STEM Education: Identifying Effective Approaches in Science, Technology, Engineering, and Mathematics*. Committee on Highly Successful Science Programs for K-12 Science Education. Board on Science Education and Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Washington, DC: The National Academies Press. https://doi.org/10.17226/13165.

National Science Board. (2012). *Science and Engineering Indicators 2012.* Arlington, VA: National Science Foundation (NSB 12-01).

National Science Board. (2018). *Science and Engineering Indicators 2018 Digest.* NSB-2018-2. Arlington, VA: National Science Foundation. Available at https://www.nsf.gov/statistics/digest/.

National Science Foundation (2021a). *Racial Equity in STEM Education.* Retrieved from: https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505910.

National Science Foundation (2021b). Women, Minorities, and Persons with Disabilities in Science and Engineering. Retrieved from: https://ncses.nsf.gov/pubs/nsf21321/report/field-of-degree-minorities.

O'Dwyer, L., & Parker, C. (2014). *A primer for analyzing nested data: multilevel modeling in SPSS using an example form a REL study* (REL 2015-046). Washington, DC: US. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from http://ies.ed.gov/ncee/edlabs.

Olsen, R., & Lie, S. (2011). Profiles of students' interests in science issues around the world: Analysis of data from PISA 2006. *International Journal of Science Education, 33*(1), 97-120. doi: 10.1080/09500693.2010.518638.

Organisation for Economic Co-Operation and Development (OECD). (2007). *PISA 2006: Science competencies for tomorrow's world: Volume 1: Analysis.* doi:10.1787/9789264040014.

OECD (2015). "How Computers are Related to Students' Performance", in *Students, Computers and Learning: Making the Connection*, OECD Publishing, Paris. doi: 10.1787/9789264239555-9-en.

Ornstein, A. (2006). The Frequency of Hands-On Experimentation and Student Attitudes Toward Science: A Statistically Significant Relation. *Journal of Science Education and Technology, 15*. 285-297. doi: 10.1007/s10956-006-9015-5.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: a review of the literature and its implications. *International Journal of Science Education,* 25(9): 1049-1079.

Panksepp, J. Affective Consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition, 14*, 30-80. doi: 10.1016/j.concog.2004.10.004.

Potvin, P., & Hasni, A. (2014). Interest, motivation, and attitude towards science and technology at K-12 levels: A systematic review of 12 years of educational research. *Studies in Science Education*, *50*(1), 85-129. doi: 10.1080/03057267.2014.881626.

President's Council of Advisors on Science and Technology (PCAST). (2010). "Prepare and Inspire: K-12 Education in Science, Technology, Engineering, and Math (STEM) for America's Future."

Raudenbush, S., & Bryk, A. (2002). Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition. Newbury Park, CA: Sage.

Raudenbush, S., Bryk, A., Cheong, Y., Congdon, Jr., R., & du Toit, M. (2019). HLM 8: Linear and Nonlinear Modeling. Skokie, IL: Scientific Software International, Inc.

Renninger, K., & Hidi, S. (2002). Student Interest and Achievement: Developmental issues raised by a case study. In A. Wigfield & J. Eccles (Eds.), *Development of achievement motivation* (pp. 173-195). New York, NY: Academic Press.

Renninger, K., Nieswandt, M., & Hidi, S. (2015). On the Power of Interest. In K. Renninger, M. Niewswandt, & S. Hidi (Eds.), *Interest in Mathematics and Science Learning,* pp. 1-14. Washington, DC: American Educational Research Association.

Rodriguez, A.J. (2015). What about a Dimension of Engagement, Equity, and Diversity Practices? A critique of the Next Generations Science Standards. *Journal of Research in Science Teaching, 52*(7): 1031-1051. doi: 10.1002/tea.21232.

Ryan, R., & Deci, E. (2000). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist, 55*(1), 68-78. doi: 10.1037110003-066X.55.1.68.

Schiefele, H., Krapp, A., Prenzel, M., Heiland, A., & Kasten, H. (1983). Principles of an Educational Theory of Interest. Paper presented at the Seventh Meeting of the International Society for the Study of Behavioral Development, Munich, Germany.

Schiefele, H., Krapp, A., Winteler, A. (1988). *Conceptualization and Measurement of Interest.* Paper presented at the annual meeting of the American Educational Research Association (AERA), New Orleans, LA.

Schraw, G., & Lehman, S. (2001). Situational Interest: A review of the literature and directions for future research. *Educational Psychology Review, 13*(1), 23-52.

Shin, Y., & Raudenbush, S. (2011). The Causal Effect of Class Size on Academic Achievement: Multivariate instrumental variable estimators with data missing at random. *Journal of Educational and Behavioral Statistics, 36*(2), 154-185. https://www.jstor.org/stable/29789476.

Simpson, R., & Oliver, J. (1990). A Summary of Major Influences on Attitude Toward and Achievement in Science Among Adolescent Students. *Science Education, 74*(1), 1-18.

Spector, P., & Brannick, M. (2011). Methodological Urban Legends: The Misuse of Statistical Control Variables. *Organizational Research Methods, 14*(2), 287-305. doi: 10.1177/1094428110369842.

Tai, R., & Maltese, A. (2009). Eyeballs in the Fridge: Sources of early interest in science. *International Journal of Science Education, 32*(5), 1143-1145. doi: 10.1080/09500690902792385.

Thorndike, E. (1935). *Adult Interests*. New York: Macmillan.

Turner, J., Meyer, D., Cox, K., Logan, C., DiCintio, M., & Thomas, C. (1998). Creating Contexts for Involvement in Mathematics. *Journal of Educational Psychology, 90*(4). 730-745. doi: 10.1037/0022-0663.90.4.730.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. "The Nation's Report Card." http://www.nationsreportcard.gov/science_2015/. Retrieved March 4, 2018.

U.S. Department of Education, National Center for Education Statistics. (2003). *Overview and Inventory of State Education Reforms: 1990 to 2000,* NCES 2003-020, by David Hurst, Alexandra Tan, Anne Meek, and Jason Sellers. Project Officer: Edith McArthur. Washington, DC: 2003. https://nces.ed.gov/pubs2003/2003020.pdf. Retrieved March 3, 2018.

Vartuli, C. (2017). Increasing High School Student Interest in Science: An action research study (Doctoral dissertation). Retrieved from ProQuest (#10270550).

Vilorio, D. (2014). Occupational Outlook Quarterly, Spring 2014: *STEM 101: Intro to Tomorrow's Jobs*. U.S. Bureau of Labor Statistics (BLS).

von Winterfeldt, D. (2013). Bridging the gap between science and decision making. *Proceedings of the National Academy of Science (PNAS), 110*(3), 14055-14061. https://doi.org/10.1073/pnas.1213532110.

Wigfield, A., & Cambria, J. (2010). Students' Achievement Values, Goal Orientations, and Interest: Definitions, development, and relations to achievement outcomes. *Developmental Review, 30,* 1-35. doi:10.1016/j.dr.2009.12.001.

Zeidler, D., Sadler, T., Simmons, M., & Howes, E. (2005). Beyond STS: A research-based framework for socioscientific issues education. *Science Education, 89*(3), 357-377.

Zimmerman, B. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 82-91.

Appendix A: Instrumentation

Activation instruments & Technical Reports available online at:
http://activationlab.org/tools/.

A1—Fascination in Science scale

Item ID	Subdimension	Prompt	Response Options / Coding
F01	Curiosity	I wonder about how nature works:	4=Every day 3=Once a week 2=Once a month 1=Never
F02	Interest	In general, when I work on science I:	4=Love it 3=Like it 2=Don't like it 1=Hate it
F04	Interest	In general, I find science:	4=Very interesting 3=Interesting 2=Boring 1=Very boring
F05	Curiosity	After a really interesting science activity is over, I look for more information about it	4=YES! 3=yes 2=no 1=NO!
F06	Curiosity	I need to know how objects work.	4=YES! 3=yes 2=no 1=NO!
F07	Mastery	I want to read everything I can find about science.	4=YES! 3=yes 2=no 1=NO!
F08	Mastery	I want to know everything about science.	4=YES! 3=yes 2=no 1=NO!
F09	Mastery	I want to know how to do everything that scientists do.	4=YES! 3=yes 2=no 1=NO!

A2—Home Resources scale

Item ID	*Are these things available for you to use in your home?*	Response Options / Coding
HR01	Calculator	4=Always 3=Most of the time 2=Rarely 1=Never
HR02	Computer (do not count video game systems)	4=Always 3=Most of the time 2=Rarely 1=Never
HR03	Internet connection	4=Always 3=Most of the time 2=Rarely 1=Never
HR04	Dictionary	4=Always 3=Most of the time 2=Rarely 1=Never
HR05	Study or homework area	4=Always 3=Most of the time 2=Rarely 1=Never
HR06	E-reader (iPad, Kindle, Nexus, etc.)	4=Always 3=Most of the time 2=Rarely 1=Never
HR07	Books about science	4=Always 3=Most of the time 2=Rarely 1=Never

A3—Family Support for Science scale

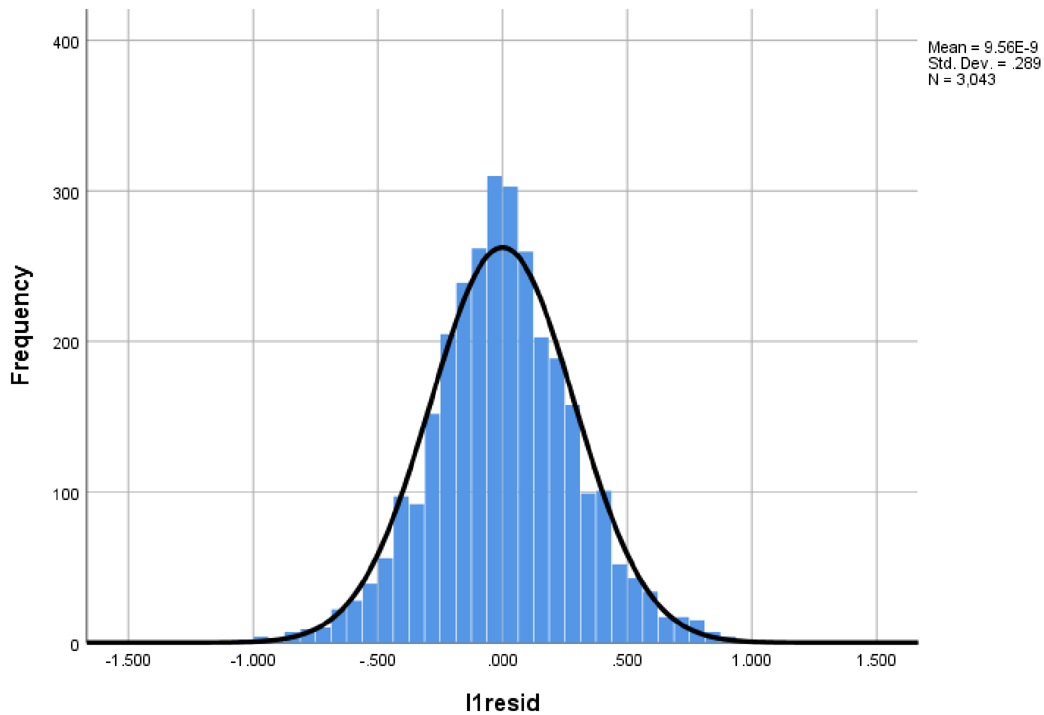
Item ID	Prompt	Response Options / Coding
FS01	My learning in school is important to someone in my family.	4=YES! 3=yes 2=no 1=NO!
FS02	When I work on homework at home, I have someone who can help me with it if I need help.	4=YES! 3=yes 2=no 1=NO!
FS03	Someone in my family is interested in teaching me things.	4=YES! 3=yes 2=no 1=NO!
FS04	Someone in my family takes me to places where I can learn new things.	4=YES! 3=yes 2=no 1=NO!
FS05	Someone in my family makes sure I finish my homework every day.	4=YES! 3=yes 2=no 1=NO!

A4—Teacher log

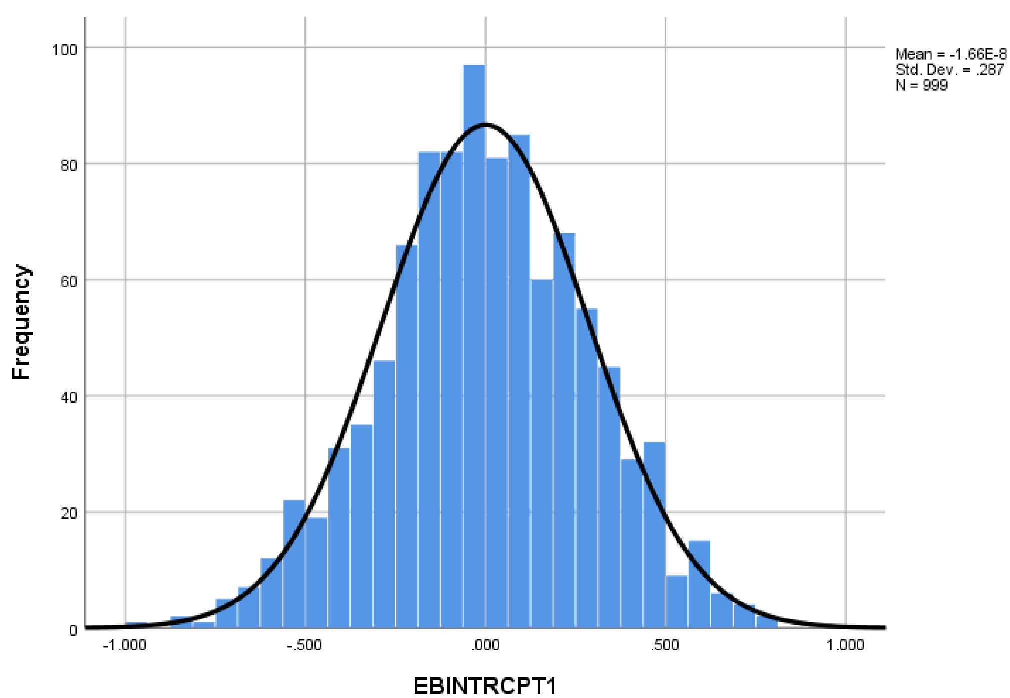
Instruction type	*Please indicate the percentage of time students spent doing each of the following in your science class over the last week. (Given that some of these items might happen at the same time, percentages do NOT need to add to 100.)*
Hands-on	Students watched a live or video-based demonstration.
Hands-on	Students did a hands-on activity.
Hands-on	Students used tools that scientists use (microscope, beakers, pipettes, etc.).
Student-centric	Students listened to a lecture or presentation.
Student-centric	Students and teacher reviewed answers to homework or classwork questions.
Student-centric	Students participated in a whole class discussion.
Student-centric	Students worked in pairs or groups.
Textbook	Students read (alone or aloud) from a book or other informational text.
Textbook	Students completed worksheets or answered questions in writing.
Textbook	Students copied notes from a book or the board.
Classroom-technology	Students used an interactive or simulation on the computer.
Classroom-technology	Students used a laptop/tablet/handheld device.

Appendix B: Residual Analyses & Model Diagnostics—Grade 6

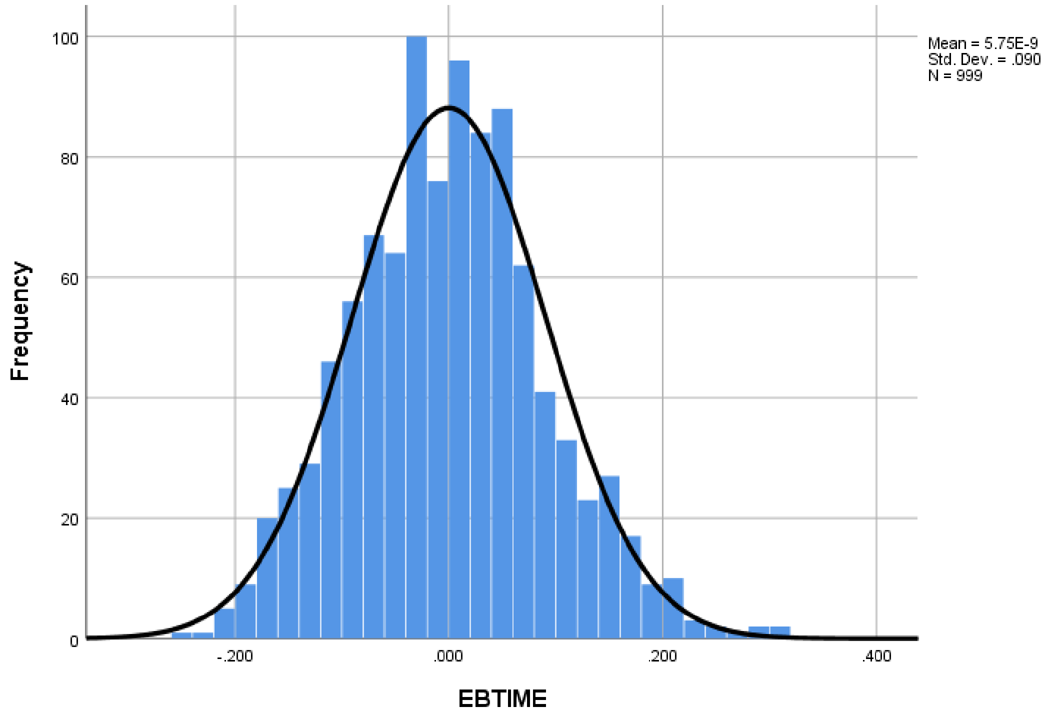
B1—Distribution of level 1 residuals (e)



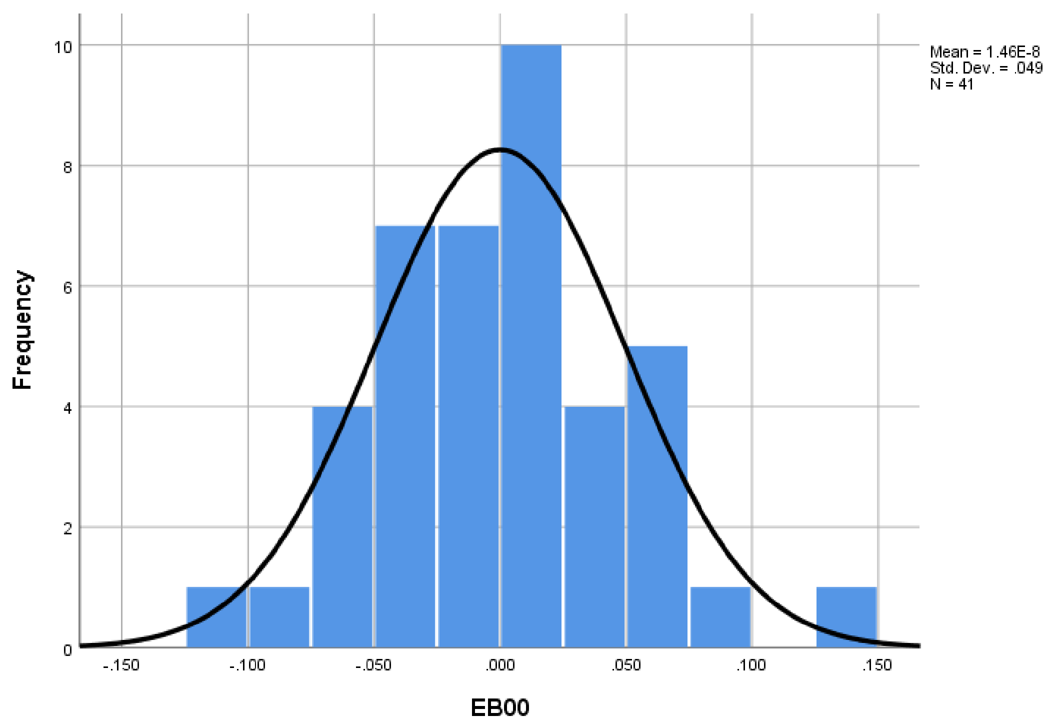
B2—Distribution of level 2 residuals – intercept (r_0)



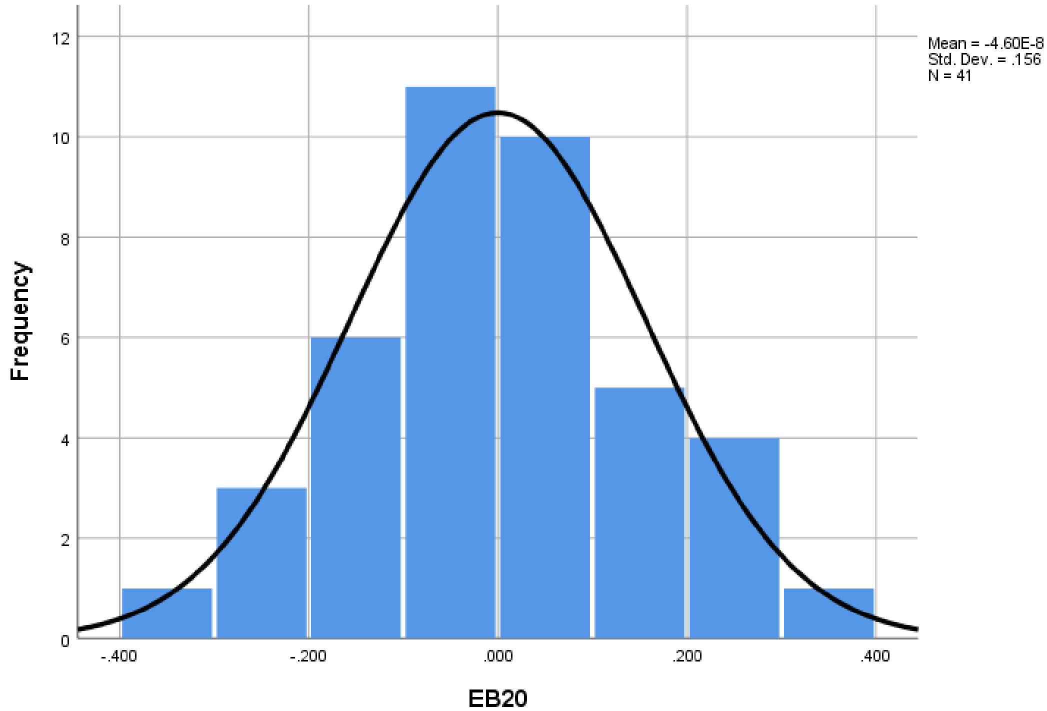
B3—Distribution of level 2 residuals – time (r_1)



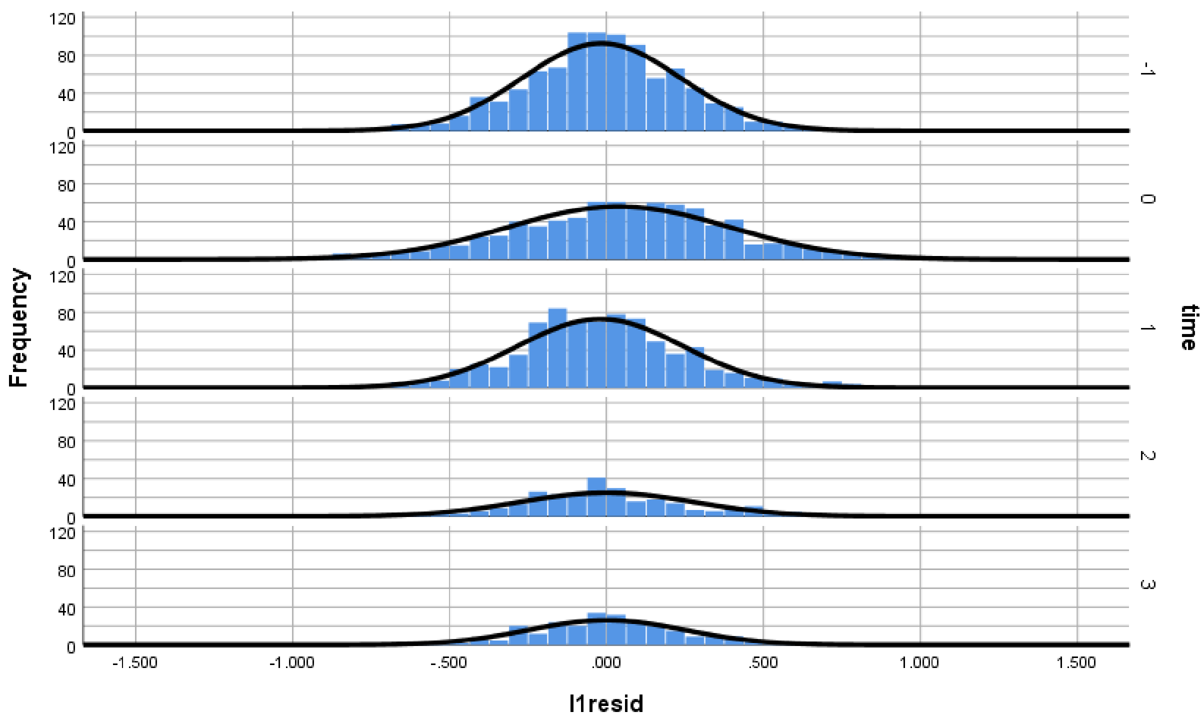
B4—Distribution of level 3 residuals – intercept (u_{00})



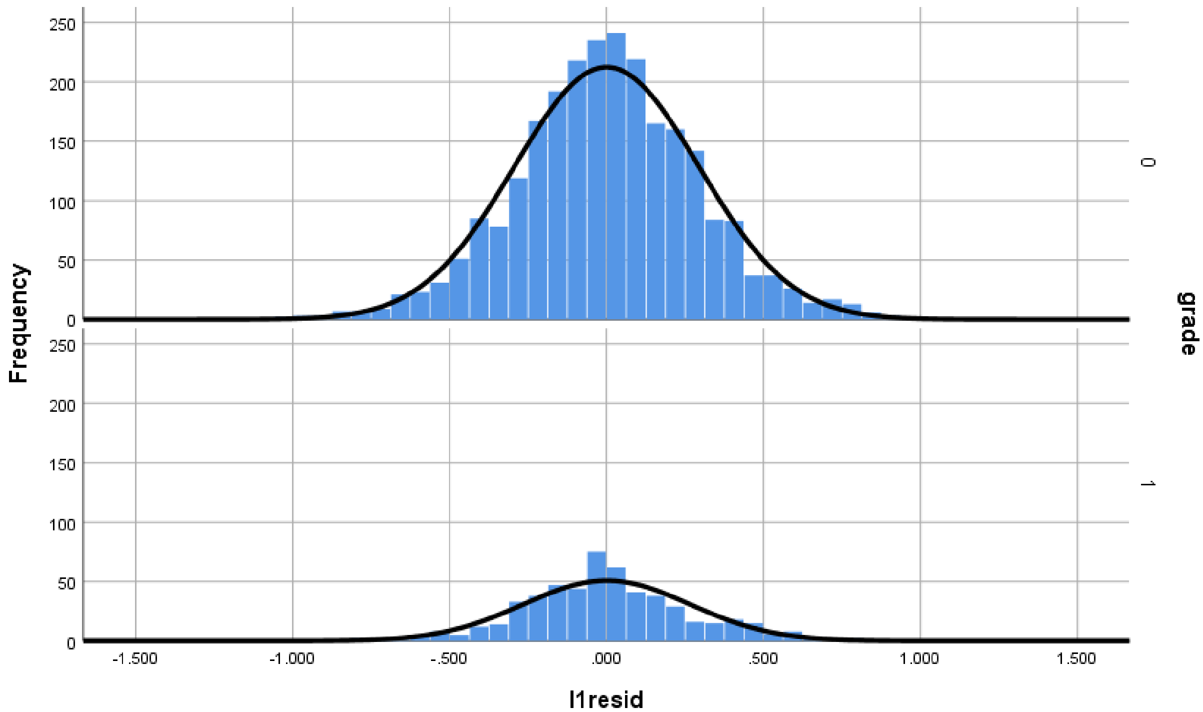
B5—Distribution of level 3 residuals – grade (u_{20})



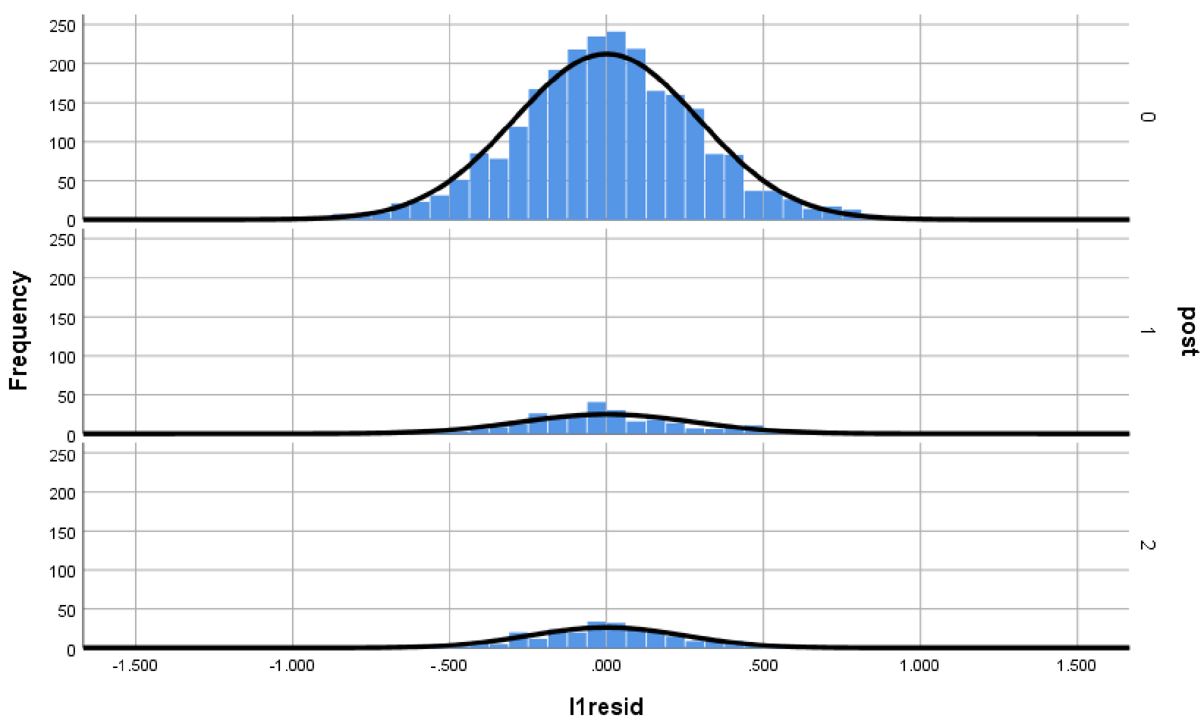
B6—Distribution of level 1 residuals by level 1 predictor, time



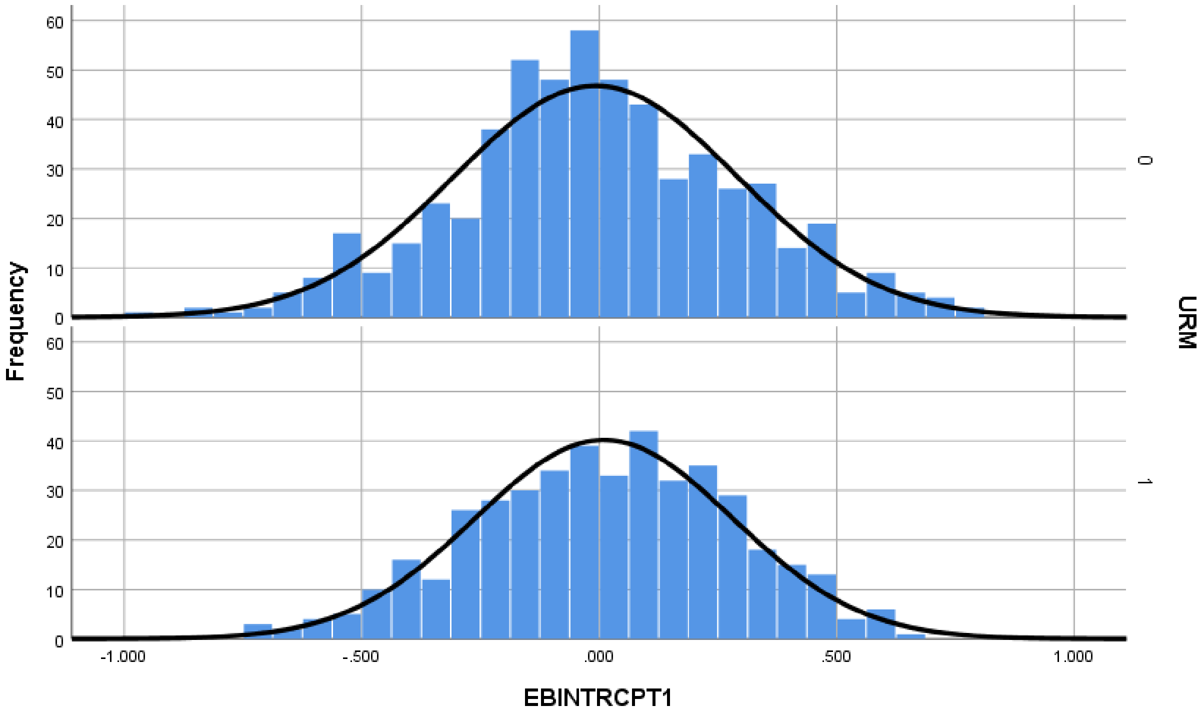
B7—Distribution of level 1 residuals by level 1 predictor, grade



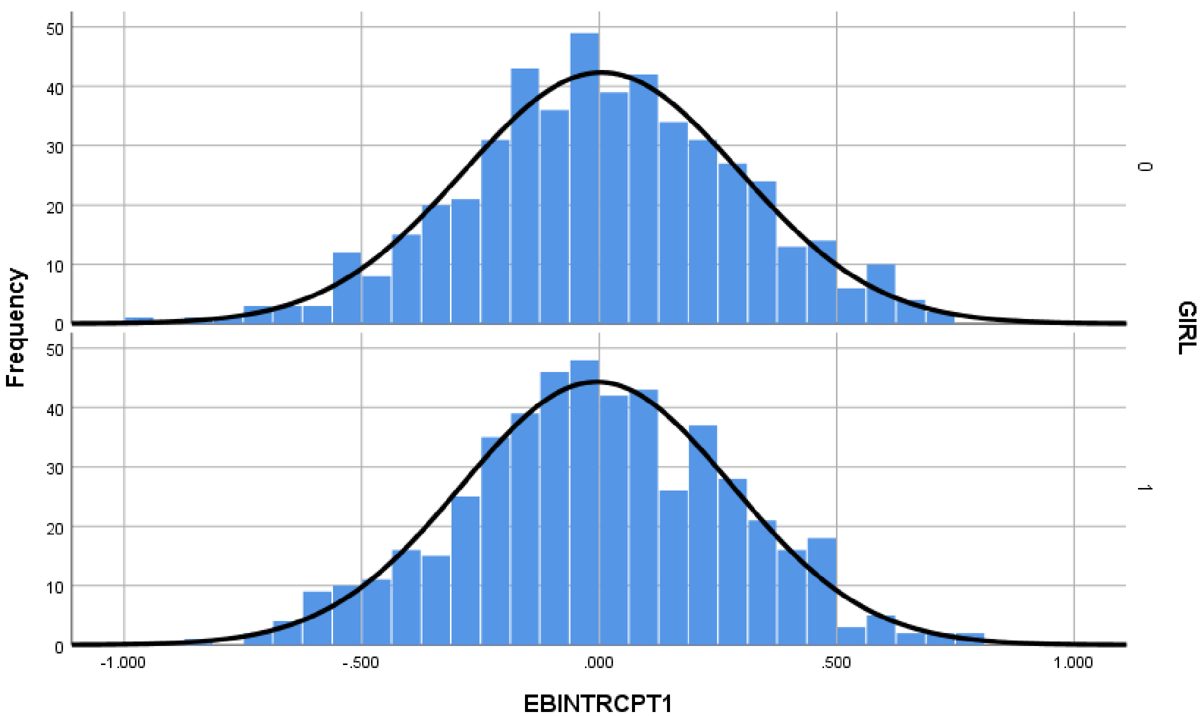
B8—Distribution of level 1 residuals by level 1 predictor, post



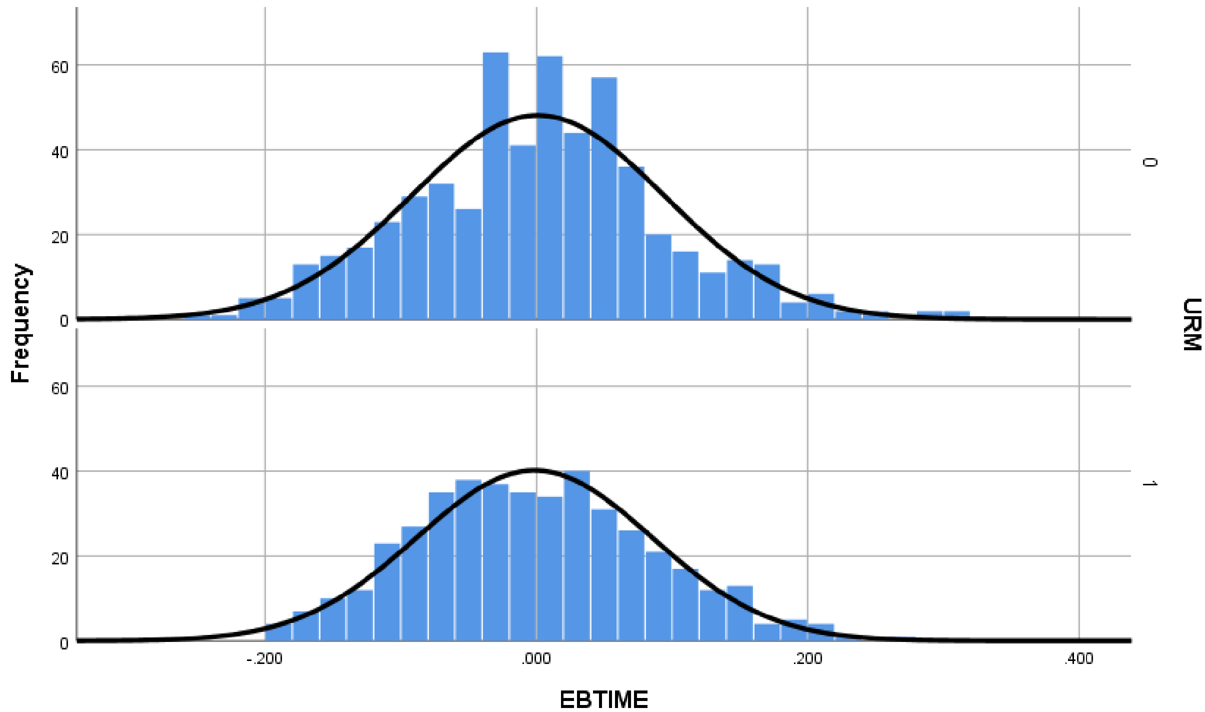
B9—Distribution of level 2 residuals (intercept, r_0) by level 2 predictor, URM



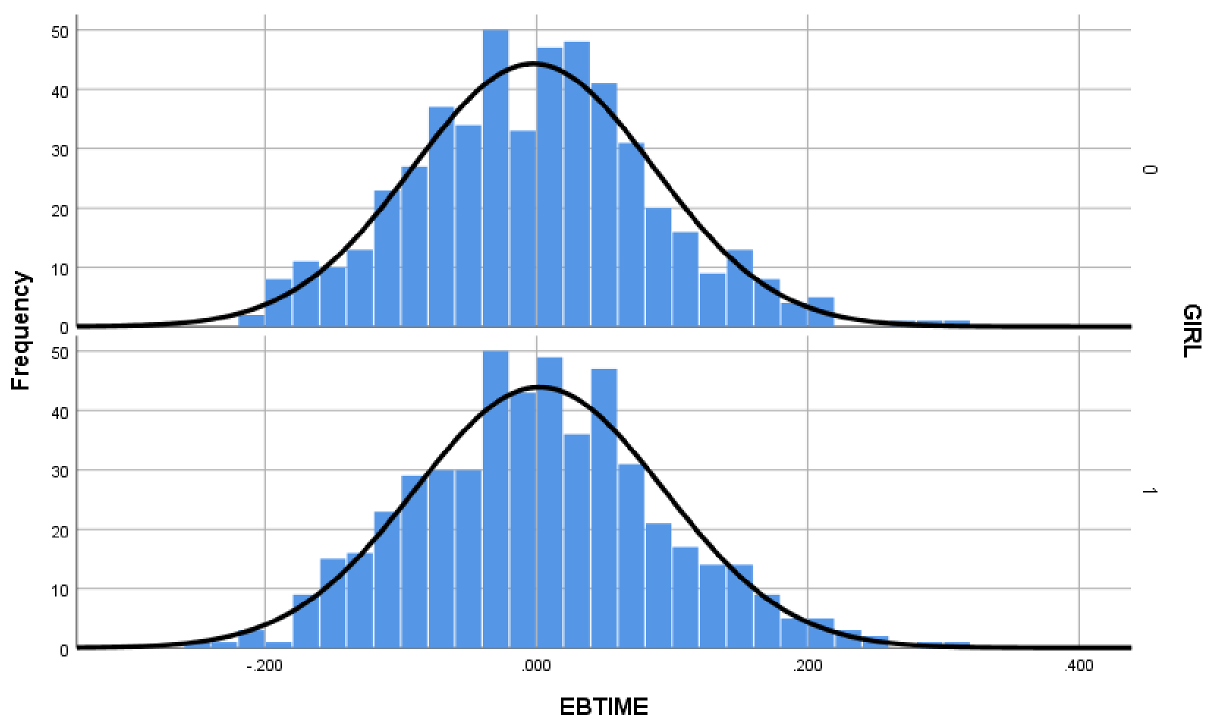
B10—Distribution of level 2 residuals (intercept, r_0) by level 2 predictor, female



B11—Distribution of level 2 residuals (time, r_1) by level 2 predictor, URM



B12—Distribution of level 2 residuals (time, r_1) by level 2 predictor, female

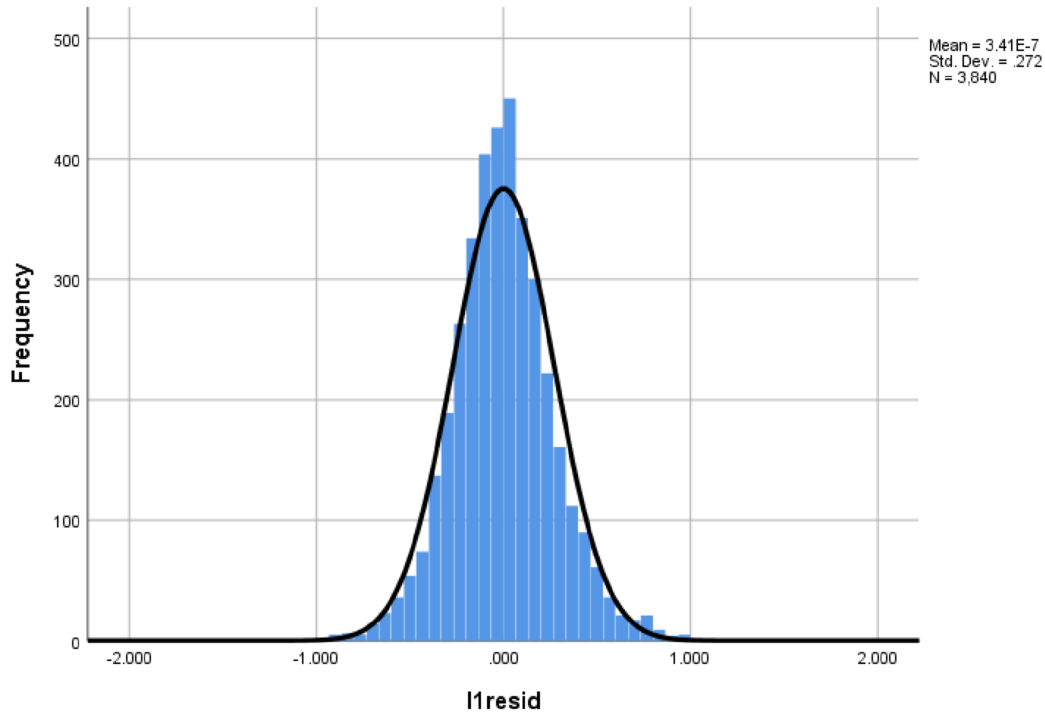


B13—Correlations of residuals and predictors

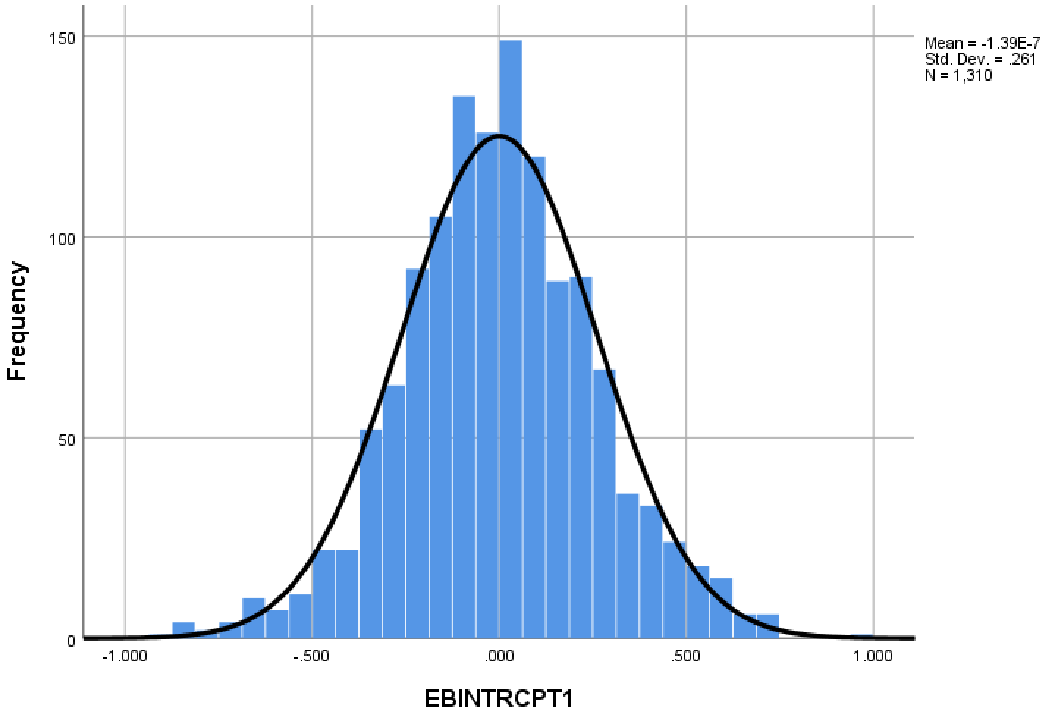
L1 pred.	L1 resid e_{ijk}	sig.	L2 resid r_0	sig.	L2 resid r_1	sig.	L3 resid u_{00}	sig.	L3 resid u_{20}	sig.
time	.00	1.00	--	--	--	--	--	--	--	--
grade	.00	1.00	--	--	--	--	--	--	--	--
post	.00	1.00	--	--	--	--	--	--	--	--
L2 pred.										
HR	.01	0.65	.03	0.30	-.03	.028	--	--	--	--
FS	.00	1.00	.00	1.00	.00	1.00	--	--	--	--
URM	.02	0.29	.03	0.34	-.01	0.67	--	--	--	--
female	.00	1.00	-.02	0.58	.03	0.42	--	--	--	--
L3 pred.										
student	.01	0.72	-.01	0.94	.01	0.67	.11	0.49	-.11	0.49
hands-on	-.01	0.90	-.02	0.57	.02	0.48	.14	0.40	-.14	0.40
class-tech	.01	0.76	.05	0.13	-.06	0.06	.19	0.23	-.19	0.23
L2 resid. (r_0)					**-.95 ****	**<0.001**				
L3 resid. (u_{00})									**-1.00 ****	**<0.001**

Appendix C: Residual Analyses & Model Diagnostics—Grade 8

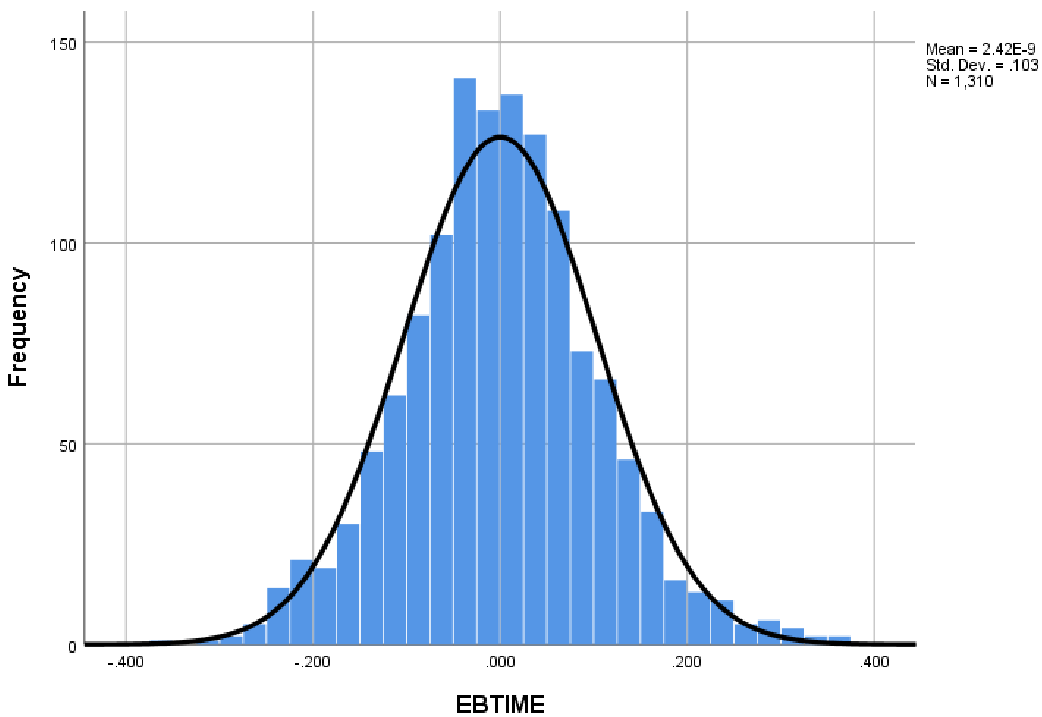
C1—Distribution of level 1 residuals (e)



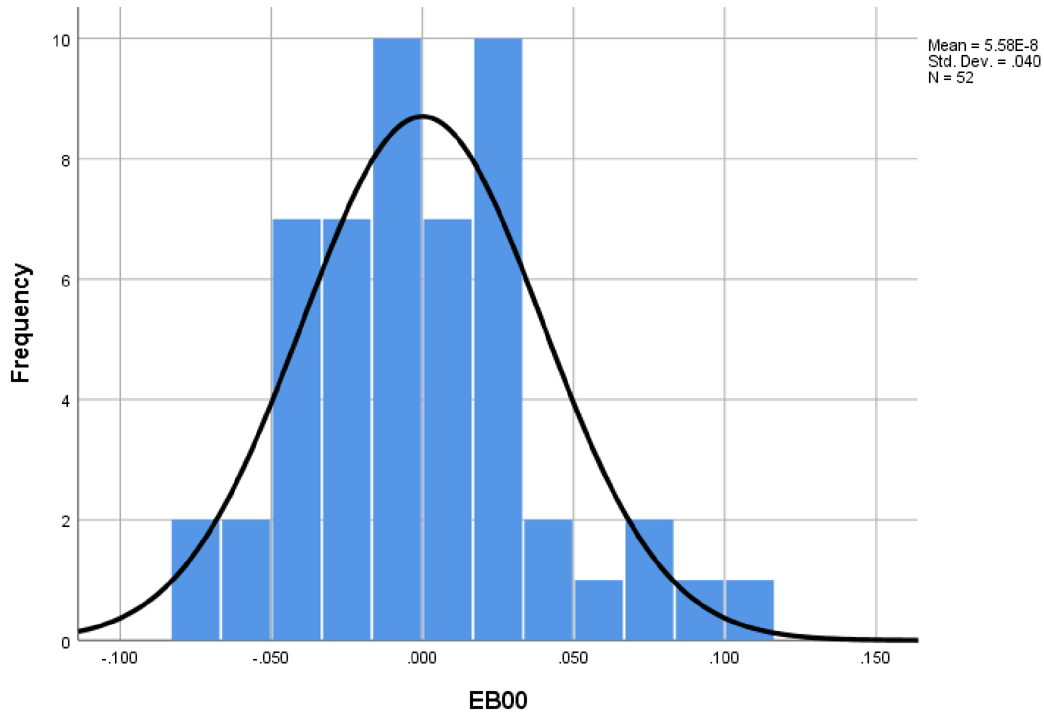
C2—Distribution of level 2 residuals – intercept (r_0)



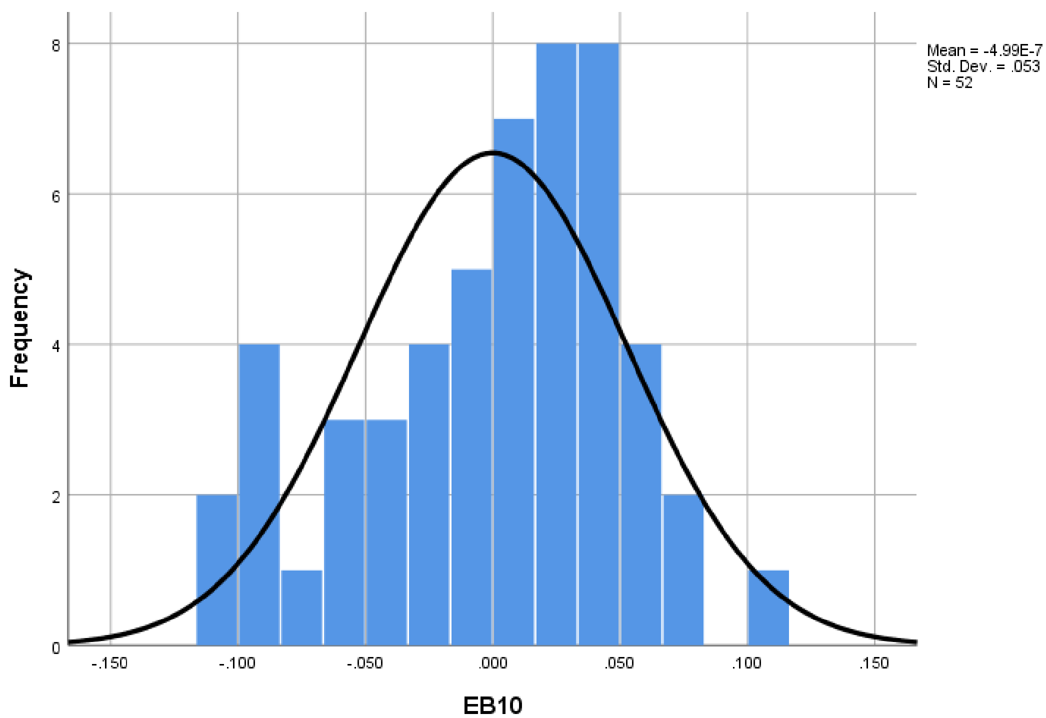
C3—Distribution of level 2 residuals – time (r_1)



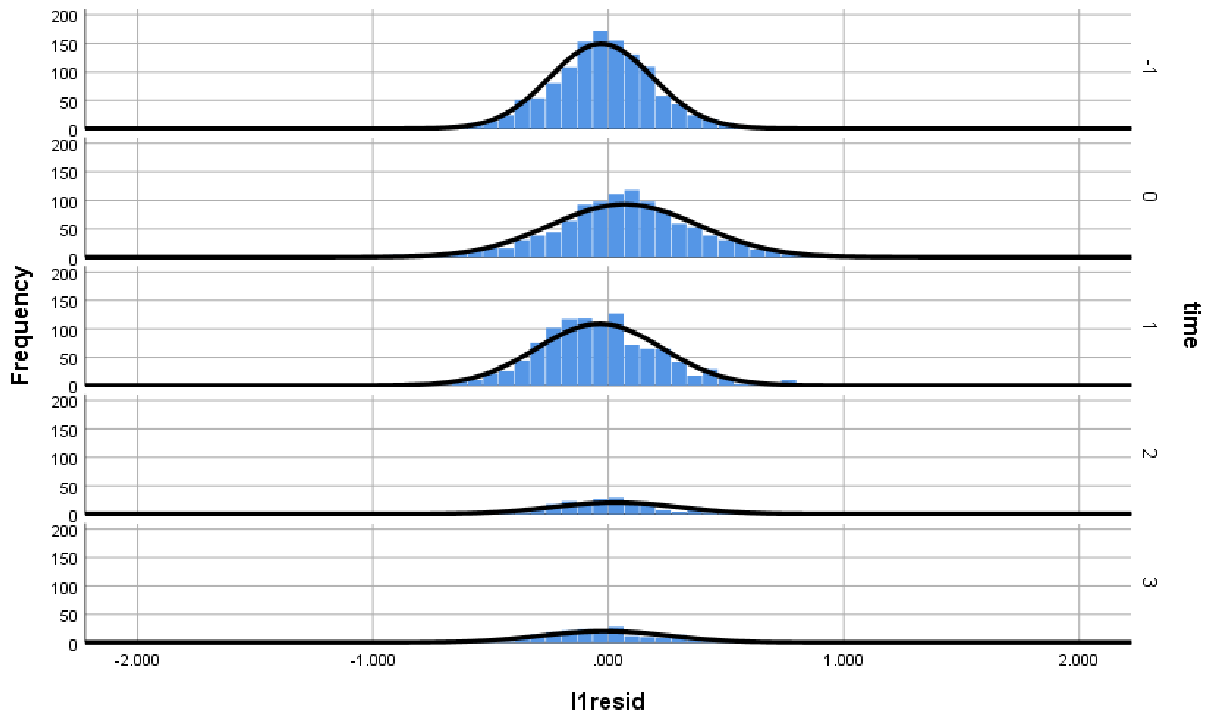
C4—Distribution of level 3 residuals – intercept (u_{00})



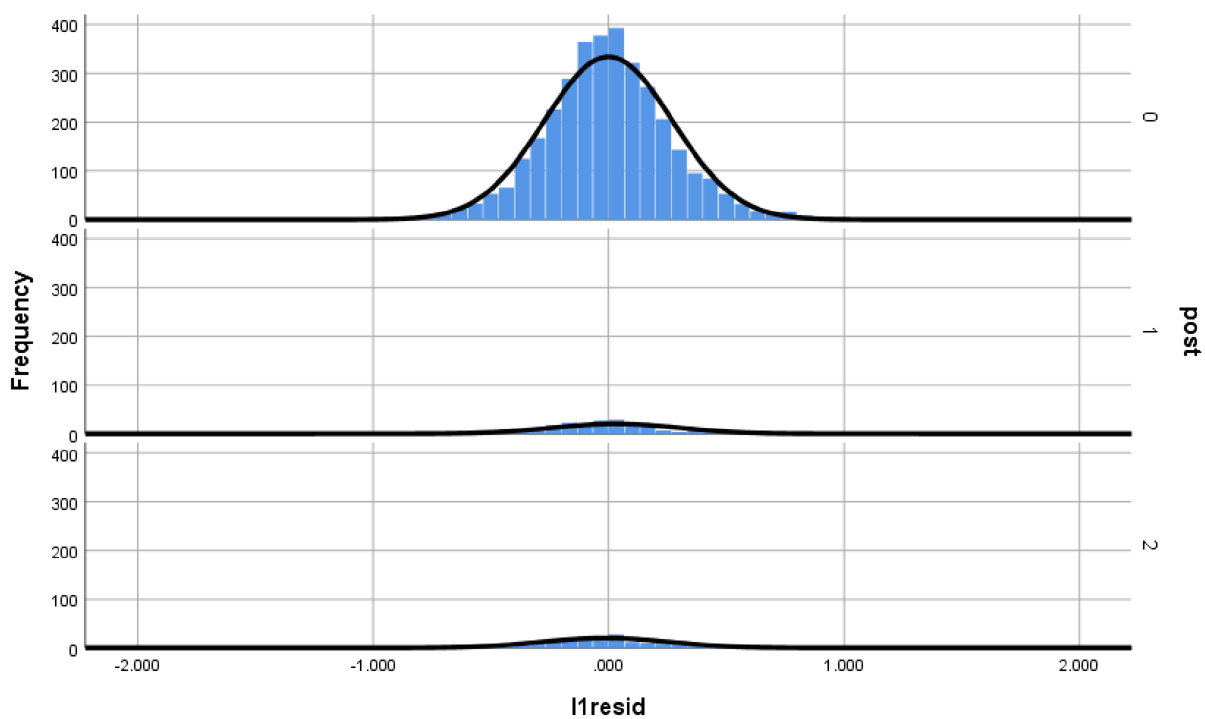
C5—Distribution of level 3 residuals – time (u_{10})



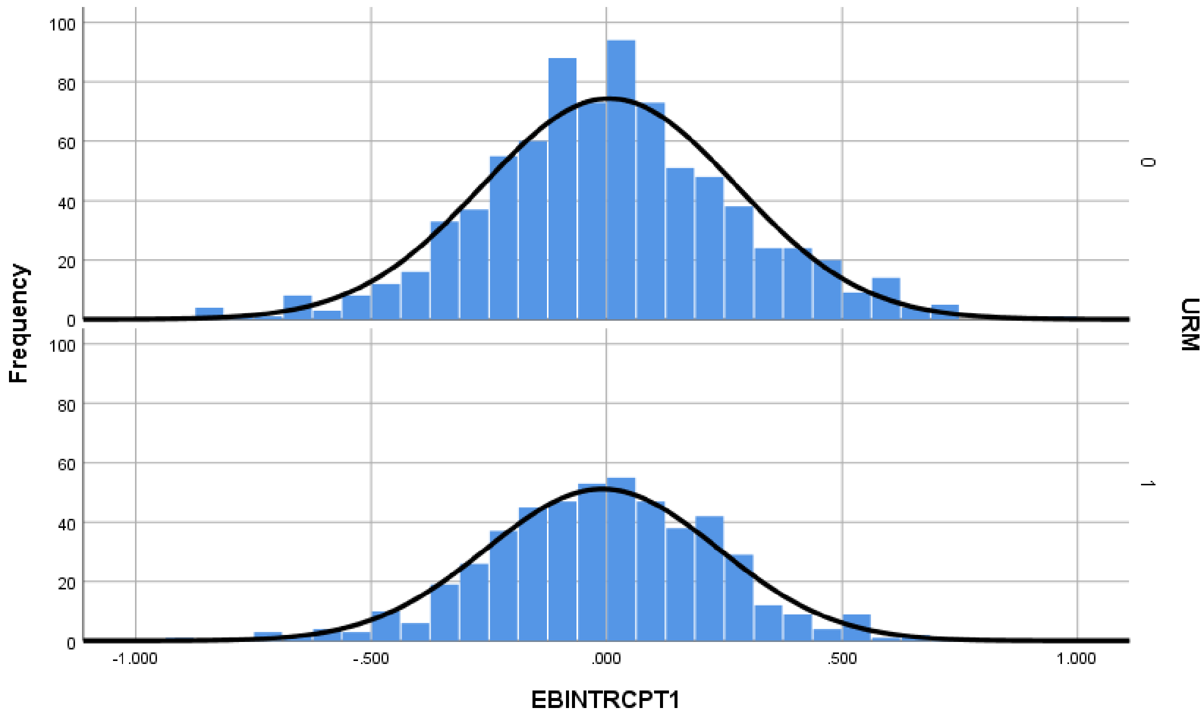
C6—Distribution of level 1 residuals by level 1 predictor, time



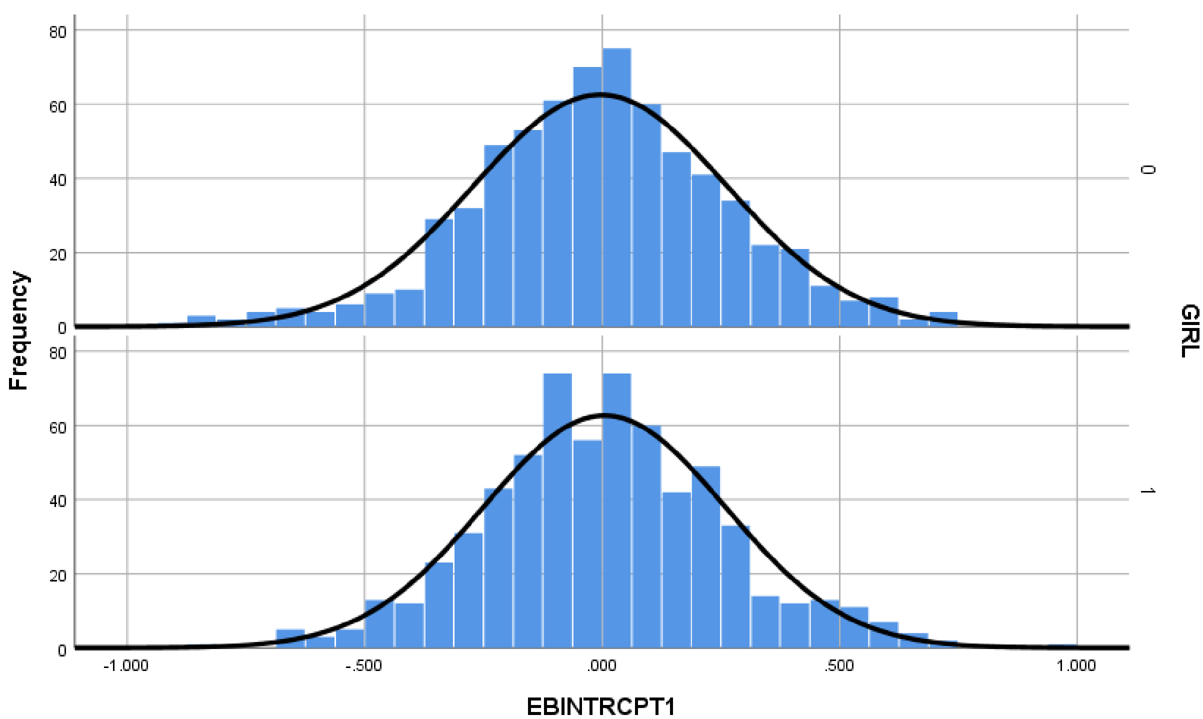
C7—Distribution of level 1 residuals by level 1 predictor, post



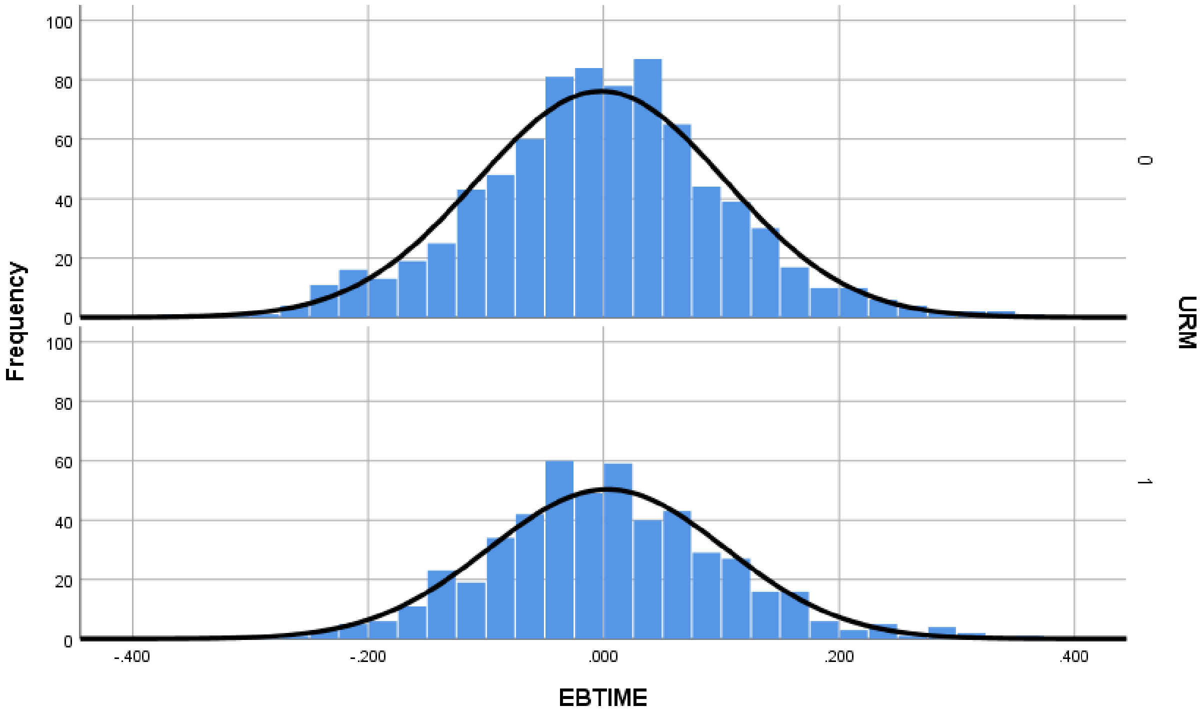
C8—Distribution of level 2 residuals (intercept, r_0) by level 2 predictor, URM



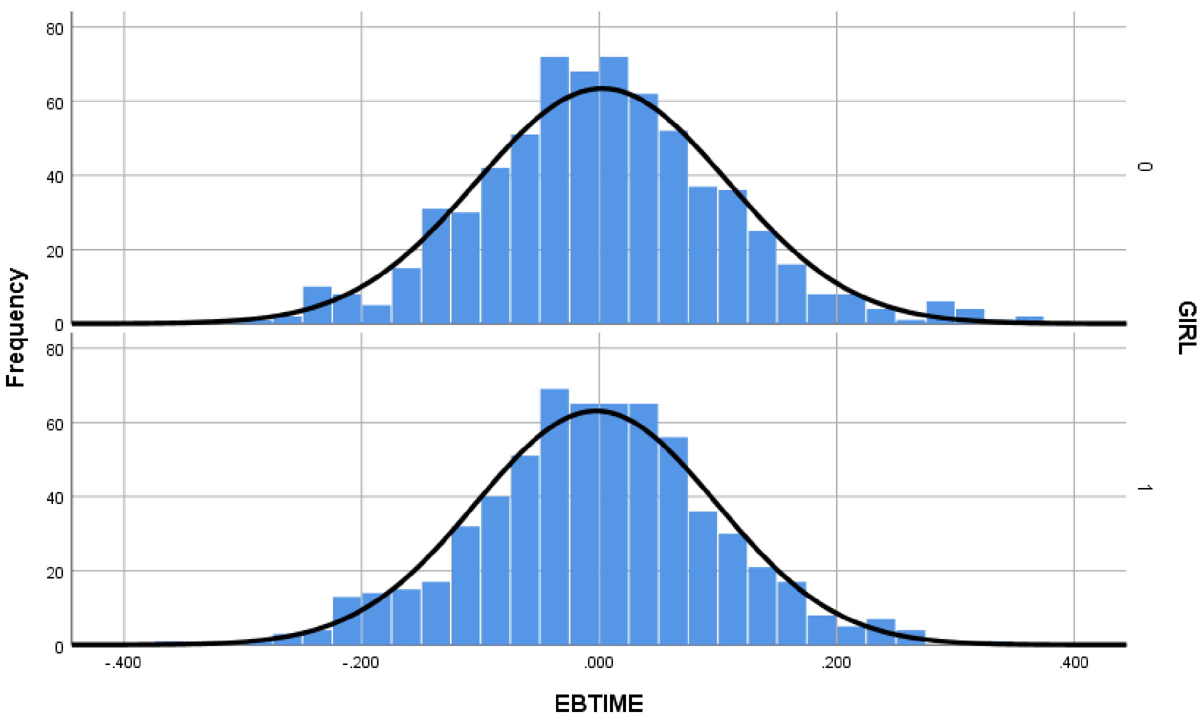
C9—Distribution of level 2 residuals (intercept, r_0) by level 2 predictor, female



C10—Distribution of level 2 residuals (time, r_1) by level 2 predictor, URM



C11—Distribution of level 2 residuals (time, r_1) by level 2 predictor, female

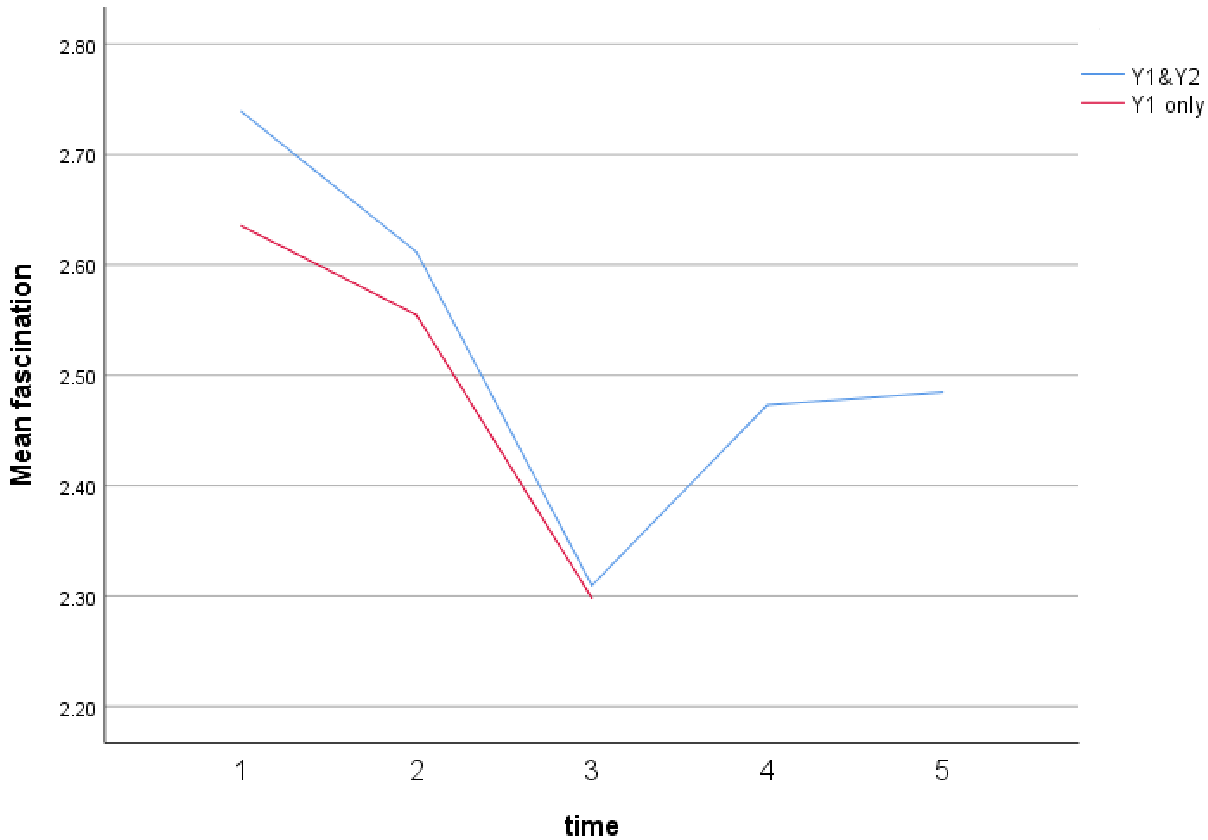


C12—Correlations of residuals and predictors

L1 pred.	**L1 resid** e_{ijk}	**sig.**	**L2 resid** r_0	**sig.**	**L2 resid** r_1	**sig.**	**L3 resid** u_{00}	**sig.**	**L3 resid** u_{10}	**sig.**
time	.00	1.00	--	--	--	--	--	--	--	--
grade	.00	1.00	--	--	--	--	--	--	--	--
post	.01	0.54	--	--	--	--	--	--	--	--
L2 pred.										
HR	.00	1.00	.00	1.00	.00	1.00	--	--	--	--
FS	.00	1.00	.00	1.00	.00	1.00	--	--	--	--
URM	-.01	0.52	-.03	0.32	.02	0.38	--	--	--	--
female	.00	1.00	.02	0.54	-.02	0.38	--	--	--	--
L3 pred.										
student	.01	0.53	.02	0.45	-.02	0.57	.09	0.52	.02	0.89
hands-on	.01	0.92	-.01	0.66	.02	0.49	-.15	0.28	.23	0.10
class-tech	-.01	0.68	.00	0.99	-.01	0.68	.12	0.41	-.27 *	0.05
L2 resid. (r_0)					**-.95 ****	**<0.001**				
L3 resid. (u_{00})									**-.72 ****	**<0.001**

Appendix D: Missing Data Analysis

D1—Comparison of fascination scores for Year 1-only students with Year 1 & 2 students



Visible above, the pattern of change in fascination scores for students who participated only in Year 1 (red line) is similar to that for students who contributed scores in both years of study. This suggests that, despite the sample attrition noted in Chapter 4, the functional form and other modeled relationships are unlikely to be affected by the shift in sample, particularly as all other included covariates at levels 2 and 3 were observed in Year 1.