INVESTIGATING STUDENT SELF-BELIEFS AND LEARNING METRICS IN ONLINE COURSEWARE: A QUANTITATIVE INQUIRY

Rachel Van Campenhout

Follow this and additional works at: https://dsc.duq.edu/etd

Part of the Educational Assessment, Evaluation, and Research Commons, Educational Psychology Commons, Educational Technology Commons, and the Higher Education Commons

Recommended Citation

This Immediate Access is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Duquesne Scholarship Collection.
INVESTIGATING STUDENT SELF-BELIEFS AND LEARNING METRICS IN ONLINE COURSEWARE: A QUANTITATIVE INQUIRY

A Dissertation

Submitted to the School of Education

Duquesne University

In partial fulfillment of the requirements for the degree of Doctor of Education

By

Rachel Van Campenhout

May, 2022
INVESTIGATING STUDENT SELF-BELIEFS AND LEARNING METRICS IN ONLINE COURSEWARE: A QUANTITATIVE INQUIRY

By

Rachel Van Campenhout

Approved February 17, 2022

Dr. Sandra Quinoñes
Associate Professor of Education
(Committee Chair)

Dr. Jason Ritter
Associate Professor of Education
(Committee Member)

Dr. Jesse Rine
Professor of Education
(Committee Member)

Dr. Gretchen Generett
Dean, School of Education

Dr. Jason Ritteer
Chair, Department of Instruction and Leadership in Education
ABSTRACT

INVESTIGATING STUDENT SELF-BELIEFS AND LEARNING METRICS IN ONLINE COURSEWARE: A QUANTITATIVE INQUIRY

By

Rachel Van Campenhout

May 2022

Dissertation supervised by Professor Sandra Quinoñes

Online courseware is an emerging educational technology that has the potential to reach students at scale. Designed with cognitive and learning science principles, courseware utilizes effective methods to maximize learning outcomes for students. Mindset (implicit theories of ability) and self-efficacy are two widely researched self-belief topics which have been shown to influence student learning outcomes and self-concepts. These self-belief theories are not specific to a domain and therefore could be
measured and analyzed in relation to student learning metrics from any subject. The
purpose of this nonexperimental correlational research study is to investigate the
relationships between student self-beliefs and key learning metrics, as collected by a
courseware platform. The data analysis process used a historical data set collected from a
natural learning context that included 1,896 students from three state higher education
institutions. This data set was analyzed using descriptive statistics and linear regression
models to answer interrelated research questions: What is the relationship between
mindset and self-efficacy for students? What is the relationship between student self-
beliefs and learning metrics in courseware? Results showed 79.7% of students selected
growth mindset and 98.6% of students selected high self-efficacy. Neither mindset nor
self-efficacy were strongly correlated to any learning metric variable. Results of the
mixed effects linear regressions model showed that mindset was significant for key
engagement metrics, though neither mindset nor self-efficacy were significant for
summative scores. The interpretation and implication of these findings are discussed in
relation to existing theory and research. Suggestions for future practice and research are
also addressed to further the application of self-beliefs in courseware environments.

Keywords
Mindset, self-efficacy, self-beliefs, courseware, engagement, learning outcomes, doer
effect, learning engineering.
DEDICATION

I dedicate this dissertation to my husband, Tom. Your steadfast support gave me the time and encouragement to complete this work. This accomplishment is a shared one, with love and gratitude.
ACKNOWLEDGEMENT

I would first and foremost like to acknowledge Dr. Sandra Quinoñes for your guidance and support throughout my doctoral experience, in addition to your service as my dissertation chair. You helped me grow as a scholarly practitioner. I am so appreciative of everything I have learned from you.

I also thank Dr. Jesse Rine and Dr. Jason Ritter for your invaluable service on this committee. Your thoughtful feedback shaped and refined this dissertation, for which I am very grateful.

Lastly, I would like to thank Dr. Benny Johnson, my colleague and mentor these many years. You have encouraged and scaffolded my acquisition of new data analysis skills, without which I could not have completed this work.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter 1: Introduction</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td>1</td>
</tr>
<tr>
<td>The Topic under Investigation</td>
<td>1</td>
</tr>
<tr>
<td>Theoretical Framework</td>
<td>3</td>
</tr>
<tr>
<td>Research Purpose Statement</td>
<td>6</td>
</tr>
<tr>
<td>Research Questions</td>
<td>6</td>
</tr>
<tr>
<td>Problem of Practice and Leadership Context</td>
<td>7</td>
</tr>
<tr>
<td>Ethics Statement</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 2: Literature Review</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td>12</td>
</tr>
<tr>
<td>Implicit Theories of Ability (Mindset)</td>
<td>12</td>
</tr>
<tr>
<td>Self-Efficacy</td>
<td>18</td>
</tr>
<tr>
<td>Courseware</td>
<td>23</td>
</tr>
<tr>
<td>Summary</td>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 3: Research Methods</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td>33</td>
</tr>
<tr>
<td>Research Design</td>
<td>33</td>
</tr>
<tr>
<td>Research Methods</td>
<td>35</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1 .................................................................................................................. 43
Table 2 .................................................................................................................. 44
Table 3 .................................................................................................................. 44
Table 4 .................................................................................................................. 59
Table 5 .................................................................................................................. 60
Table 6 .................................................................................................................. 62
Table 7 .................................................................................................................. 63
Table 8 .................................................................................................................. 64
Table 9 .................................................................................................................. 65
Table 10 ............................................................................................................... 66
Table 11 ............................................................................................................... 72
Table 12 ............................................................................................................... 74
Table 13 ............................................................................................................... 74
Table 14 ............................................................................................................... 75
Table 15 ............................................................................................................... 75
Table 16 ............................................................................................................... 77
Table 17 ............................................................................................................... 86
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>5</td>
</tr>
<tr>
<td>Figure 2</td>
<td>10</td>
</tr>
<tr>
<td>Figure 3</td>
<td>37</td>
</tr>
<tr>
<td>Figure 4</td>
<td>38</td>
</tr>
<tr>
<td>Figure 5</td>
<td>40</td>
</tr>
<tr>
<td>Figure 6</td>
<td>42</td>
</tr>
<tr>
<td>Figure 7</td>
<td>49</td>
</tr>
<tr>
<td>Figure 8</td>
<td>50</td>
</tr>
<tr>
<td>Figure 9</td>
<td>52</td>
</tr>
<tr>
<td>Figure 10</td>
<td>53</td>
</tr>
<tr>
<td>Figure 11</td>
<td>53</td>
</tr>
<tr>
<td>Figure 12</td>
<td>54</td>
</tr>
<tr>
<td>Figure 13</td>
<td>55</td>
</tr>
<tr>
<td>Figure 14</td>
<td>56</td>
</tr>
<tr>
<td>Figure 15</td>
<td>57</td>
</tr>
<tr>
<td>Figure 16</td>
<td>59</td>
</tr>
<tr>
<td>Figure 17</td>
<td>61</td>
</tr>
<tr>
<td>Figure 18</td>
<td>67</td>
</tr>
<tr>
<td>Figure 19</td>
<td>67</td>
</tr>
<tr>
<td>Figure 20</td>
<td>68</td>
</tr>
<tr>
<td>Figure 21</td>
<td>68</td>
</tr>
</tbody>
</table>
Figure 22 ....................................................................................................................... 68
Figure 23 ....................................................................................................................... 69
Figure 24 ....................................................................................................................... 69
Figure 25 ....................................................................................................................... 69
Figure 26 ....................................................................................................................... 80
Figure 27 ....................................................................................................................... 81
Chapter 1: Introduction

Overview

The purpose of this chapter is to provide an introduction to the topics being investigated in this research study. To contextualize these topics within this study, the theoretical frameworks are explained, and the research questions outlined. This chapter also discusses this research study as a problem of practice situated within my professional experience at a research-based educational technology company. Last, but not least, ethical considerations that guide both this research study and my professional purpose are highlighted.

The Topic under Investigation

Courseware is an emerging educational technology that combines content, media, formative practice, and assessments into a single comprehensive online learning environment (Van Campenhout et al., 2020). The learning science principles and methodologies on which the courseware is based are designed to be non-domain specific to be effective no matter the content subject area. Many computer-directed learning environments are designed to be context agnostic and implemented across widely varying situations (Kessler et al., 2019), and the courseware used in this study was designed with this intention. Improving learning outcomes (as measured by assessment scores) for students using non domain specific approaches is necessary for developing scalable and cost-effective courseware. One approach that has not yet been explored in this technology is how student self-beliefs could be leveraged to help increase learning outcomes. Self-beliefs, as a broad term for specific self-concepts, have long been researched as a causal agent in academic success, and a meta-analysis of longitudinal self-belief studies found small, positive effect sizes on achievement (Valentine et al., 2004).
Two types of self-beliefs which may be valuable to measure and influence are student beliefs about their intelligence and ability. As noted by Dweck (2006), students who believe their intelligence is fixed hold entity theories, or fixed mindset, while students who believe their intelligence can grow hold incremental theories, or growth mindset. Previous research has correlated those mindsets with trends in learning outcomes, self-efficacy, persistence, and other important factors to learning and student success (Dweck, 2006; Blackwell et al., 2007; Claro et al., 2016), and also showed that interventions can shift mindsets from fixed to growth, which enables students to produce better learning outcomes (Dweck 2008; Aronson et al., 2002). Self-efficacy is the degree to which a student believes he or she is able to accomplish a task (Bandura, 1986). Self-efficacy has been shown to relate to student persistence, effort, and outcomes (Tseng et al., 2020; Zimmerman, 2000). Thus, incorporating self-belief evaluations and interventions could be a valuable domain independent tool to add to online courseware learning environments.

Before interventions for self-beliefs can be developed and deployed through courseware, it is first necessary to investigate whether these self-belief variables are related to key learning metrics in the courseware—such as engagement and outcomes. As part of an educational grant, several state institutions used the same Probability and Statistics courseware as the primary learning resource for students taking an introductory course. This courseware contained explanatory content, formative practice, statistics tutors, graded checkpoints and quizzes, and adaptive practice (Van Campenhout et al., 2020). Before the core content was a short module of content on learning strategies, which included an introduction to the courseware, explanation of features, navigation tutorials, etc. Part of this module included validated surveys on mindset and self-efficacy. The data collected by the courseware platform are used to investigate the relationships between student self-beliefs and the various learning metrics available.
There are several ways this research is significant. First, the body of research on self-efficacy and growth mindset is mostly done in more traditional, face-to-face learning environments. This research contributes to this literature by show whether or not similar results can be found using data from student engagement and learning in courseware. One benefit of online learning environments is the wealth of high-quality data (Goldstein & Katz, 2004) that can provide insights into student behavior and learning. Evaluating student self-beliefs using this type of data could reveal new insights not previously seen when using pre- and post-tests, surveys, etc. from traditional face-to-face classroom settings. Second, if student mindset and self-efficacy do have a relationship with engagement and learning outcomes within the courseware, this could lead to new intervention techniques to help students learning with digital resources such as courseware. An online environment allows for new and different intervention strategies that could be employed to shift student mindsets and efficacy toward those which are more productive for learning outcomes.

Future uses of these research findings will ultimately benefit students. My goal as a professional in the educational technology field is to put students first. This research is timely, as the findings could influence how self-beliefs are incorporated into educational technology learning environments. Investigating mindsets of real students in comparison to their engagement and learning outcomes provides insights which could help future students who enroll in courses that use courseware.

**Theoretical Framework**

This research draws upon theoretical frameworks related to self beliefs about learning, as well as emerging theoretical frameworks related to both the learning science principles used in courseware as well as how this technology is developed. There is a significant body of research
on implicit theories of ability, and how they can influence learning outcomes and personal characteristics (such as self-esteem) (Dweck, 2006; Claro et al., 2016). With the relationship between ability theory and outcomes established, researchers have gone on to identify intervention methods that help students shift their mindset about ability for more beneficial outcomes (Aronson et al., 2002; Blackwell et al., 2007; Paunesku et al., 2015). For example, interventions ranged from watching videos and receiving instructional material on mindset, to peer mentor activities. This research has been done extensively in traditional classrooms.

Similarly, self-efficacy has been studied as one’s belief in one’s ability to accomplish a task (Bandura, 1986), which has been shown to correlate to specific outcomes such as student persistence, effort, and outcomes (Tseng et al., 2020; Zimmerman, 2000). Research by Chen et al. (2001) into general self-efficacy found that it acted as a motivational trait and acted as a main effect predictor variable for other outcomes of interest. In this research, these self-belief theoretical frameworks will be used to identify if these same beliefs about intelligence and ability are apparent in a courseware environment, and if they relate to the same types of outcomes found in the existing body of research.

This research study also employs frameworks on emerging technology and processes. The doer effect is a learning science principle identified and studied at Carnegie Mellon University (Koedinger et al., 2015; 2016; 2018). This research used a courseware environment and found that doing formative practice at frequent intervals while reading content was causal to better outcomes (Koedinger et al, 2016; 2018). The concept of the doer effect is not particularly new; most educators are aware that doing practice is good for learning. However, investigating the doer effect with online learning environments produces large volumes of data that allows for educational data mining and scaled learning analytics to evaluate learning science principles in
naturally occurring contexts (Baker 2016; Koedinger et al., 2016). Therefore, the doer effect framework is central to the courseware learning methodology used in this study.

The final theoretical framework is the process of an emerging professional and academic field: learning engineering. Learning engineering as a professional role was spread through the leadership of Carnegie Mellon University, which has a master’s program in learning engineering and employs learning engineers to develop learning environments at the university (Simon, 1967; Lieberman, 2018). This role transitioned to Acrobatiq when it formed from Carnegie Mellon’s Open Learning Initiative (OLI) and I was hired at Acrobatiq as a learning engineer in 2015. While there is a role called learning engineer, it also refers to a process which can be engaged in by a single person or an entire team (Goodell et al., 2020). The IEEE standards committee IC Industrial Consortium on Learning Engineering (ICICLE) is a volunteer professional organization which formed to develop definitions and standards for learning engineering for professional and academic use (ICICLE, 2020). An ICICLE special interest group, led by Aaron Kessler, developed a Learning Engineering Process model (Figure 1) based on the experiences of those doing this work in a variety of settings (Kessler et al., 2020). Learning engineering as a profession and process serves as a framework for both the courseware being studied and how this research is situated in its own learning engineering process.

**Figure 1**

*The Learning Engineering Process (Kessler et al., 2020).*
Research Purpose Statement

The purpose of this study is to investigate if the learning benefits found in the mindset and self-efficacy literature can also be found in an online courseware environment. The research questions and methods will help to determine if students who identified as having growth/fixed mindset or high/low self-efficacy also correlated to specific patterns of engagement or learning outcomes. This research will determine if developing interventions for shifting self-beliefs (as described in Chapter 2) could be a worthwhile project in the future. Thus, this research acts as an input into the learning engineering process to help determine whether future design and development changes should be made in courseware for self-belief interventions.

Research Questions

There are two higher level research questions that encompass the overall research goals of this study:

1. What is the relationship between mindset and self-efficacy for students?
2. What is the relationship between student self-beliefs and learning metrics in courseware?

Each research question has finer-grained sub-questions outlined in Chapter 3. These research questions group students according to their mindset and self-efficacy beliefs in order to investigate if these groups have differences in engagement behaviors or learning outcomes. The mindset and self-efficacy groups are also compared to determine if they correlate or behave independently. These analyses help to conclude if further research and development is warranted in this area.
Problem of Practice and Leadership Context

I was originally hired at Acrobatiq as a learning engineer, and a primary responsibility for this role was to advocate for the student by using learning science-based methods and research to influence the product and oversee the content for courseware. As a learner-centered role, the learning engineers kept the best interests of the student at the heart of what we created. The mission of Acrobatiq was to develop courseware based on learning science to create a more effective online learning environment for students, and in doing so, make learning more equitable and accessible.

The Carnegie Project on the Educational Doctorate (CPED) defines a problem of practice as: “a persistent, contextualized, and specific issue embedded in the work of a professional practitioner the addressing of which has the potential to result in improved understanding, experience, and outcomes” (CPED, n.d.). This research project developed from a problem of practice that I observed through my professional experience. The primary focus of the work I did on the courseware environment was on learning methods, such as formative practice design and adaptivity, and driving student engagement with these features. I then began thinking about other approaches that we could take to help students learn in this type of environment, and metacognitive approaches such as mindset and self-efficacy became an interest that I focused on in my doctoral studies. If the distinctions identified in the literature on the differences between incremental and entity theory mindsets as well as high/low self-efficacy hold true for engagement and learning outcomes in courseware, we could potentially leverage intervention treatments directly in the platform. Research identified how mindset intervention treatments could benefit under-performing students and buffer the effects of poverty on learning outcomes and racial stereotypes (Claro et al., 2016; Paunesku et al., 2015; Aronson et al., 2002). If these
findings could be replicated in a courseware setting, this could lead a shift in technology solutions to address learning from a more holistic perspective.

**Ethics Statement**

_Ethical Standards_. Professional standards are incorporated into this doctoral program, both for coursework as well as doctoral requirements such as the digital portfolio process. The Association for Educational Communication and Technology (AECT) provides ethics indicators for each professional standard (AECT, 2012), which inform this dissertation-in-practice. The research standard ethics indicator states, “Candidates conduct research and practice using accepted professional and institutional guidelines and procedures” (AECT, 2012). In addition to completing research procedures required by Duquesne University, such as the Internal Review Board (IRB) process, it is also necessary to understand how ethics are incorporated into the procedure and methodology of this research study itself. For instance, the use of a historical data set eliminates direct contact with students, eliminating the concern for ethical treatment of human subjects in this study. As this existing data set was collected from a real course in a natural learning environment, there were no experimental controls or treatments, so all students received the same learning environment benefits.

Within the context of this research project, special attention should be paid to the AECT (2012) ethics indicator for the standard on learning environments, that states learning environments should be fostered through ethics which guide health, safety, and best practices. The learning environment used in this research is a technology-based environment that uses real-time student data for predictive modeling and adaptation (Van Campenhout et al., 2020), with increasing evolution toward the incorporation of artificial intelligence (Van Campenhout, Dittel, et al., 2021). Research bodies such as the European Commission have presented reports on how
the “the recent advances in artificial intelligence and machine learning will have profound impacts on future labour markets, competence requirements, as well as in learning and teaching practices” (Tuomi, 2018, p. 2). While this report identifies the ability of artificial intelligence to create new methods for teaching and learning, it also identifies the potential for automating bias in education. Mayfield et al. (2019) identified the ways in which machine learning in educational technologies could enact widespread algorithmic bias, and suggest an agenda to combat this. One simple recommendation was to define boundaries; not all problems should be solved by machine learning.

This research project will not evaluate student demographic of any kind. This data does not exist, as it is not collected by the platform. This serves a legal purpose, as it reduces the locations in which student personal identifying information is stored. But the lack of demographic information serves more than just a legal purpose; demographics are not included as part of the courseware technology to avoid potential algorithmic bias. While there is much value in studying the relationships between demographic characteristics of race or gender on learning, that is not a relationship this technology solution should attempt to impact. The use of demographic data in predictive analytics in educational technology is controversial and less actionable than learning behavior (Baker, 2016). Simply put, demographic characteristics are not features that artificial intelligence or machine learning in educational technology should tackle. This is further supported by research from Western Governors University that showed learning data to be the most valuable and standalone predictor of student success, over all other readiness assessment and demographic characteristics (Olsen and Shackelford, 2021). They suggest adaptive systems rely on activity data alone, as it is readily available, does not require gathering legally protected data, and is the most indicative variable for student success.
An Ethical Framework. In addition to the specific ethical standards developed by AECT, my experience as a learning engineer informed the development and participation in a learning engineering ethical framework (Van Campenhout, 2021a). This ethical framework has shaped my professional work and identity, and therefore shapes this research project as well. In this framework (Figure 2), I outlined how professional purpose can develop an ethical voice. This ethical voice is used to engage in a dialogic ethic with both team members and the students for whom the learning environment is being created. The learning engineering process can then be engaged in as a reflective and ethical process. Engaging in this ethical framework helps to maintain a learner-centered approach to research and the learning engineering process (Van Campenhout, 2021b)

Figure 2

The Learning Engineering Ethical Framework (Van Campenhout, 2021a).

This dissertation is guided by my professional purpose—both as a doctoral student and a learning science specialist—and acts as the first steps of the learning engineering process. This
process requires input from the learning and cognitive science research (Goodell et al., 2020; Kessler et al., 2020, Van Campenhout, 2021a). This research study provides insights into the relationship between student self-beliefs and learning behaviors in courseware, and this research could then be used for data-driven decisions for new technology development. The design and decision making for educational technology should be learner-centered and informed by research, and in doing so, engage an ethical practice.
Chapter 2: Literature Review

Overview

This literature review serves to situate this research study in the relevant contexts for the topics investigated. The self-belief theories of intelligence and ability are defined and select research from the substantial bodies for each reviewed. It is necessary to understand the relevant research outcomes for each self-belief independent variable, as this helps to illuminate the potential outcomes of this study. The learning environment for this study shapes the learning metric outcome variables being investigated, so it is also necessary to review the literature on the courseware and the learning science principles it is based on. The literature in this chapter is referenced again in context to the results and discussion in Chapters 4 and 5.

Implicit Theories of Ability (Mindset)

For several decades, Carol Dweck and colleagues have been studying implicit theories of ability (also called mindset, which will be used as the variable term in this study), and how those affect student behavior and outcomes. Dweck (1991) studied children who were presented with a series of eight questions they could successfully answer followed by four questions which were too difficult. The children were asked to think aloud as they worked, and researchers were able to group children into two groups based on their reactions to the difficult problems: helpless or mastery oriented. The helpless children began to describe themselves as failing the task, became pessimistic, and their problem-solving strategies became less sophisticated; they defined themselves as having limited ability. However, the master-oriented students increased concentration, increased self-talk of instructions and problem-solving strategies, and spoke positively of being able to master the difficult problems. Dweck (1991) proposed these two
different responses and behaviors from children were due to their conceptions of ability. Students with an entity theory believe ability is a static entity and cannot be controlled while those with an incremental theory believe ability can be changed incrementally with effort.

Blackwell et al. (2007) outlined a framework of characteristics for these mindset beliefs. Those with an incremental theory of ability tend to have learning goals to increase their ability while those with an entity theory of ability tend to have performance goals that document their ability. Those with incremental theories believe that effort is useful and those with entity theories believe it is futile. When confronted with failure, those with incremental theories will attribute it to low-effort mastery-oriented causes. Those with entity theories attribute failure to low-ability and develop helpless attitudes when confronted with challenges. When faced with difficulties, those with incremental theories will display mastery-oriented strategies—like increasing effort or changing strategies—while those with entity theories will display helpless strategies like decreasing effort. Blackwell et al. (2007) found learning outcomes were related to student mindsets in a longitudinal study of middle school math grades. Students who held entity theory beliefs showed drops in grade point averages and self-esteem, while those with incremental theory beliefs maintained or increased in those areas.

Dweck and other researchers continued to study and expand upon these implicit theories of ability with regard to self-esteem and self-worth. Nussbaum and Dweck (2008) researched how mindset related to how students deal with difficulties, repair deficiencies, and deal with self-esteem. Only half of students with entity theory beliefs chose to remediate their deficiencies compared to 90% of students with incremental theory beliefs. Mindset also correlated to what types of social comparison students sought; students with entity theory beliefs compared themselves to students who did worse as a means of repairing self-esteem while students with
incremental theory beliefs compared themselves to those who did better. Niiya et al. (2004) investigated how ability theories and academic contingent self-worth were independent variables that can interact to shape how students react to failure. They found that incremental mindset could act as a self-esteem buffer when confronted with failure, especially among vulnerable students who tie self-worth to academic success.

In 2016, Claro et al. published a study using a nationwide dataset of high school students in Chile to identify how mindset impacted achievement. All students in Chile completed a standardized test, which included a growth mindset survey (n = 169,203 and n = 168,553). The two largest predictors of achievement were socioeconomic status and mindset. Students from lower-income families were less likely to have an incremental theory mindset compared to wealthier peers. However, those low-income students who did hold an incremental theory mindset were far less affected by the effect of poverty on achievement. The scale of this study is promising for future research and intervention treatments which could specifically target disadvantaged students to mitigate the impact of socioeconomic predictors.

Mindset Interventions. With the behaviors and correlations of learning outcomes to ability mindsets established, researchers began to experiment with intervention treatments that aimed to shift mindsets from entity theory to incremental theory. Dweck (2008) argued that self-theories (beliefs about ability) have important causal relationships to challenge-seeking, resilience, and self-regulation, and that changing these self-theories can have meaningful impact for people. Dweck believes that modest intervention efforts can yield large changes for students. Blackwell et al. (2007) continued their research on mindset by repeating the same longitudinal study on middle school math students, but also included a treatment for mindset for half of the students. Students participated in eight undergraduate-led workshops which delivered content to
describe the brain as malleable and demonstrate that intelligence can be increased. The intervention treatment was shown to shift student mindsets about their ability, and students who had an entity theory prior to the treatment and shifted to incremental theory halted their decrease in scores in the months following the intervention.

In a study by Aronson et al. (2002), African American college students were given a mindset intervention treatment which researchers hypothesized would help to reverse the negative ability stereotypes which adversely affect this group of college students. Students watched a film about how the brain can grow and develop with challenges, and were asked to write letters to younger students explaining this concept. At the end of the semester these students had a greater value of academics and a higher grade-point average compared to the control groups. Another mindset intervention study focused on rural adolescent girls—who often have lower academic attainment than their suburban counterparts (Burnette et al., 2018). Their randomized experiment created a treatment group which engaged in a highly structured 45-minute lesson on mindset theory. Modules were focused on intelligence mindsets, person mindsets, and self-regulation mindsets, and each had a four-part structure that included research findings, messaging, a peer role-model with tips, and a saying-is-believing exercise. Girls who completed the mindset intervention treatment reported stronger growth mindsets at the end of the treatment and at a 4-month follow-up. They also reported greater learning motivation, efficacy, and higher grades (though this was only an indirect effect on grades via a shift in mindset). Researchers suggest the mindset treatment may not have been strong enough to shift academic attitudes and grades, and suggest future improvements based on other studies. Paunesku et al. (2015) delivered online mindset intervention modules to 1,594 students across 13 diverse high schools. These treatments were more beneficial for low-performing students; for the third of the
sample at-risk for dropping out of high school, the intervention increased the rate at which students passed core courses by 6.4 percent.

**Unsuccessful Mindset Replications.** Mindset interventions would be a popular option for any educational setting if it were true that with minimal effort, interventions could create significant improvements in student learning. However, not all studies have found positive results. In a recent study (McCabe et al., 2020), freshman students at a small university were given an intervention which involved a mindset survey followed by a video on growth mindset and an infographic. Midway through the semester students received a reminder email with an invitation to review the material. At the end of the semester students answered the survey questions again. Analysis showed no measurable benefits to student grades or retention for the intervention group over the control group.

Li and Bates (2019) attempted to replicate the Mueller and Dweck findings (1998) with a group of 624 9–13-year-old Chinese children. In one study they closely replicated the previous research study, the growth mindset manipulation was associated with performance on one post-failure test, but not on the other 8 motivation and attribution metrics Mueller and Dweck used. In two additional studies designed to distinguish the effects of mindset versus other manipulations, they found no similar results to the original study. Glerum et al. (2020) followed up previous unsuccessful mindset research to further investigate if specific replication of the types of feedback associated with entity or incremental theory mindsets could reproduce the mindset outcomes previously found by Dweck. Their study using 108 vocational students provided students with feedback praising intelligence or praising effort, but no difference was found on mindset or student task choices. This study claims to have closely replicated the Mueller and Dweck (1998) study and still did not find similar results.
In Response to Unsuccessful Mindset Replications. Glerum et al. (2020) characterized Dweck’s responses to non-results with three main arguments. First, mindset is often misinterpreted and oversimplified, resulting in false assumptions on how growth mindset is passed on or encouraged. Second, intervention treatments cannot simply be copied from one situation to another—what works for fifth graders may not work with undergraduate students. Third, developing growth mindset is more beneficial for underperforming students or with difficult content, and more research is required to understand what is not working for students and how to improve it. While the number of unsuccessful replications is increasing, it is also still unclear if it is a failure of the study or the theory as a whole. With regards to average or non-results, Yeager et al. (2019) noted:

Without clear evidence about why average effect sizes differ in later-conducted studies—evidence that could be acquired from a systematic investigation of effect heterogeneity—researchers may prematurely discard interventions that yield low average effects but could provide meaningful and replicable benefits at scale for targeted groups.

Interventions may not always be effective on the first try. This is an important note to keep in mind as research is conducted, including the research and outcomes of this study. Iterative improvement may be necessary to change intervention content or approaches and generate better outcomes for students (which is part of the learning engineering framework used in this study).

Yeager et al. (2019) put forth a comprehensive national study with independent data collection of thousands of high school students from a representative sample of the national population, with pre-registration of analysis and confirmation of results from a blinded Bayesian analysis. An intervention which took less than an hour was found to shift mindsets and improve
grades, and more so for lower-performing students. As the largest study on mindset interventions, rigorously tested and confirmed, these results suggest guidelines for implementation and analysis for future intervention treatments.

Self-Efficacy

Albert Bandura has been studying self-efficacy for decades across age groups, disciplines, and contexts. Bandura (1977) explains the relationship between outcomes and efficacy as: “An outcome expectancy is defined as a person's estimate that a given behavior will lead to certain outcomes. An efficacy expectation is the conviction that one can successfully execute the behavior required to produce the outcomes” (p. 193). A student might know that studying will help them pass a test, but if they doubt their ability to pass the test, their efficacy beliefs can hinder their outcomes. Bandura continues to explain that a person’s conviction in their own efficacy can affect whether they attempt to cope with a situation. Bandura also states that, “Efficacy expectations determine how much effort people will expend and how long they will persist in the face of obstacles and aversive experiences” (p. 194). According to Bandura (1986), self-efficacy is the degree to which a student believes he or she can accomplish a task. Self-efficacy is highly influential for academic confidence and can influence whether people think optimistically or pessimistically, which can be either self-enhancing or self-hindering. Bandura (2001) argued that people’s beliefs about their efficacy determine the challenges they choose and their level of engagement. This is one reason self-efficacy is used as a predictor variable in this research study; engagement in the learning environment is a measure of great interest.

The self-efficacy Bandura describes has situational demands. According to social cognitive theory (Bandura, 1986; 1997), self-efficacy can vary based on three different
dimensions: level or magnitude (of task difficulty), strength (or certainty of successfully performing a particular task), and generality of the magnitude and strength across tasks. Regarding the generality of self-efficacy, Bandura (1997) states:

Powerful mastery experiences that provide striking testimony to one’s capacity to effect personal changes can also produce a transformational restructuring of efficacy beliefs that is manifested across diverse realms of functioning. Such personal triumphs serve as transforming experiences. What generalizes is the belief that one can mobilize whatever effort it takes to succeed in different undertakings. (p. 53)

According to Chen et al. (2001), these situational requirements have created a narrow focus on the magnitude and strength dimensions, studying specific self-efficacy (SSE). Research into the generality trait of self-efficacy differs in scope and is therefore described as general self-efficacy (Chen et al., 2001). Researchers (Eden, 1988; Gardner & Pierce, 1998; Judge et al., 1997) have suggested that SSE is a motivational state and GSE is a motivational trait, and Eden (1988) suggests “GSE is much more resistant to ephemeral influences than SSE” (Chen et al., 2001, p. 63). The results of several field research studies confirmed:

…that the impact of experimental treatments on motivation and performance was greater among participants with low GSE than among those whose GSE was high (Eden & Aviram, 1993; Eden & Kinnar, 1991; Eden & Zuk, 1995). Thus, as predicted based on sound theory, internally and externally valid experimentation has shown that GSE acts as both a main effect predictor variable and as a moderator of motivational processes of major interest to organizational scholars. (Chen et al., 2001, p. 64).
Self-efficacy and Achievement. Bandura has contributed to many research studies, including a study in which Caprara et al. (2008) examined perceived self-efficacy on 412 Italian students to determine effects on academic achievement and drop-out rates. In this longitudinal research study they found that self-efficacy declined from junior to senior high school. The smaller this decline was, the higher student grades were in senior high school and the less likely students were to drop out.

Self-efficacy perceptions are distinctive from other self-concepts and motivational constructs because of their specificity to performance tasks (Zimmerman, 2000). Self-efficacy is a valid predictor of student motivation and learning, including activity choice, effort, persistence, and emotional reaction. When it comes to analysis and specificity of self-efficacy, Yeo and Neal (2006) investigated the relationship of self-efficacy when evaluated at the within-person and between-person levels. After evaluating self-efficacy at repeated intervals, they found that within-person and between-person self-efficacy differed and the level at which self-efficacy is being conceptualized needs to be considered. This is an important note when considering evaluations of different students and how self-efficacy correlates to engagement and outcomes, compared to intervention treatments and evaluating individual students over time.

Self-Efficacy and Online Learning. With the rise of online learning in the past decades, self-efficacy has been investigated in relation to student engagement and outcomes in this new and evolving environment. A meta-analysis of research (Tsai et al., 2011) looked at 46 papers on self-efficacy and Internet-based learning (IBL) between 1999 and 2009. This meta-analysis reported that in general, student’s self-efficacy had a positive influence on their processes and outcomes in IBL. While this finding is congruent with self-efficacy research in traditional
settings, there are important notes to be made given the year of this analysis. First, IBL covers a wide range of technologies and implementations of that technology. This would certainly influence how self-efficacy is studied and how generalizable those results would be. Second, this body of research often measures computer self-efficacy, internet self-efficacy, or internet-based learning self-efficacy. For studies conducted more than a decade ago, this would have been logical and insightful as IBL was emerging. However, for current and upcoming college students, self-efficacy for using a computer or the internet are much less of a concern, as this generation of students are considered digital natives (Prensky, 2001).

Spence and Usher (2007) studied the self-efficacy of students in online versus traditional sections of remedial algebra, where courseware (MyMathLab) was used as a resource. Students selected a traditional in-person course, or an online course which was self-directed. Researchers measured a variety of variables including course settings, gender, computer self-efficacy, self-efficacy for regulated learning, computer playfulness, engagement, mathematics grade self-efficacy, and mathematics achievement. The traditional sections outperformed the online sections in achievement; however, this difference disappeared when controlling for math grade self-efficacy, which also was the primary predictor for student achievement. Self-efficacy for self-regulation was also a strong predictor of success. This study found that engagement with courseware was not found to be a predictor for mathematics achievement, however, there are several likely reasons for this. The courseware was not required, and was only one resource of many. The courseware itself was also not a comprehensive learning environment, but rather a compilation of resources including drill-style practice which the researchers note have been less impactful for learning. Student engagement was also a self-reported measure and not a complete dataset gathered by the courseware itself. While the courseware resource used and the
methodologies of this study are not analogous to this research project, key findings that mathematics self-efficacy was the strongest predictor of success, and self-efficacy of self-regulation was more predictive of computer self-efficacy, shows the importance of self-belief constructs. Student beliefs about their own ability to monitor their learning process, and achieve success in the content, predict achievement.

Tseng et al. (2020) studied first-time online undergraduate and graduate students to evaluate how mindset and self-efficacy were related to engagement in online courses. Their findings overall stated that mindset and self-efficacy had a positive relation to online learning. Both growth mindset and self-efficacy were predictive of student engagement. There were some interesting findings which were contrary to previous research, however. In the three dimensions of flexible learning measured, “open-mindedness” and “adapting to new learning situations” were highly correlated to self-efficacy. However, the third dimension “learning technology acceptance” had no significance to self-efficacy, indicating that those students did not perceive technology as a factor of their learning ability. Contrary to other research findings, growth mindset did not influence self-efficacy; a student’s belief in their mindset ability did not influence their feelings of self-efficacy toward success in the online course. This study did show that students with incremental mindset theory showed higher engagement in learning. While this very recent study engages a likely digital native population of students and is revealing in its correlations of mindset, self-efficacy, and engagement, there are some considerations to make for their direct comparability to this study. First, there was no specific description of what the online courses were, so it’s possible these were online courses with traditional learning resources managed via a learning management system. Second, engagement was once again self-reported from students via survey and not as a dataset tracking all engagement habits. It should also be
noted that these results are from only 23% of the identified first-time online learners who self-selected to return the survey.

Courseware

**Courseware and Learning Theory.** The learning environment described as courseware is an emerging technology with specific requirements and components. Courseware will have an operational definition as a comprehensive learning environment that is comprised of short, topically-aligned content lessons with integrated formative practice, adaptive assessment, and summative assessment (Van Campenhout et al., 2020). Courseware provides all of the content a student needs in order to complete a course. Lessons are single pages of content that align to a central learning objective. The content, media, and formative practice all align to that learning objective, and questions gather data against that objective. Formative practice has immediate feedback, does not produce a grade, and can be answered repeatedly. Adaptive activities are placed at the end of a module of lessons and provide scaffolded question sets tailored to an individual student’s needs. Quizzes are placed at the end of the module and produce a score for the gradebook. This courseware provides students with a single place to learn, practice, and complete assessments while providing instructors with gradebook scores and data-driven dashboards to monitor student learning in real time.

As with all instruction, the courseware is influenced by the learning theories which have shaped educational practice for decades, and the concepts of knowledge acquisition which have been in existence for centuries. Empiricism, beginning with Aristotle, is the view that knowledge is constructed through experience and external stimuli, while rationalism, beginning with Plato, is the view that knowledge originates in the mind separate from the senses (Schunk, 1991). These concepts may not be at the forefront of the design of educational technology, but they
influence the learning theories which in turn impacts educational technology and learning resources. Ertmer and Newby (2013) described behaviorism, cognitivism, and constructivism in relation to the design of instructional contexts, so I will use to describe and apply these theories in the context of the courseware design—though focusing at this time only on behaviorism and cognitivism. The behaviorist approach associates learning with observable performances which can be reinforced through stimuli and responses. The learner is considered reactive and there is no attempt to evaluate the student’s structure of knowledge or mental processes. Behaviorist approaches to design are beneficial for building and strengthening stimuli-response relationships using reinforcement and practice (Winn, 1990)—especially effective for discriminations (recalling facts) and generalizations (defining concepts) (Ertmer & Newby, 2013). These behaviorist views manifest in common learning design elements such as incorporating observable measures of success, preassessment for learner evaluation, sequencing of instruction based on complexity, and reinforcement such as feedback or rewards. Cognitivism focuses on the acquisition of knowledge and internal structures and processes. The learner is considered an active participant in the learning process who must encode, store, and retrieve information, and cognitivists focus on learning strategies to help students do so (Ertmer & Newby, 2013). Cognitivist approaches are generally beneficial for reasoning, critical thinking, problem-solving, and information processing (Schunk, 1991). Cognitivist views manifest in instructional techniques such as encouraging the use of metacognitive strategies, focus on structuring and organizing content for easier processing, and facilitating connections with prerequisite content (Ertmer & Newby, 2013).

Much of the design of the courseware was created based on cognitivism and the importance of information processing and the role of the learner as an active participant. The
purpose of the learning strategies unit is to teach students metacognitive learning strategies and help them understand the strategies and goal setting that will help them be successful. Both the platform and content were created to reduce extraneous cognitive load, with a critical evaluation of working memory capacity. Lessons are organized around student-centered learning objectives that help organize and process content. The approach to formative practice was to scaffold students as they read to incrementally increase their knowledge and actively engage them in learning. These are all cognitivist design approaches, as Ertmer and Newby (2013) describe.

It is certainly true that certain features can be interpreted differently depending on whether one holds a behavioralist, cognitivist, or constructivist viewpoint. Consider varying views on the purpose of feedback, for example. As Ertmer and Newby (2013) state, “A behaviorist uses feedback (reinforcement) to modify behavior in the desired direction, while cognitivists make use of feedback (knowledge of results) to guide and support accurate mental connections (Thompson et al., 1992)” (p. 53). In the design of the formative practice and feedback in this courseware, the feedback was elaborative with the intention of helping students correct misconceptions, integrate new content into mental models, and inform their next answer choice—a cognitivist approach. And yet, this same feature can also be interpreted from a behaviorist perspective as a reinforcement technique. One widely socialized problem of practice I have discussed with instructors who use courseware is how to change student behavior so that they engage with the practice. Students do not get the learning benefit of the design if they do not do it. A behavioralist would call that a behavior challenge to change how students respond to stimuli, while a cognitivist might call that a change in cognition from passive to active learning. Both are looking to achieve the same goal. It is likely the best approaches would be to use methods from both approaches.
The courseware was designed and developed based on learning research with the singular goal of helping students learn. As a complex learning environment, different features draw upon methodology. Snelbecker (1983) proposed that those who work to address learning needs cannot afford the luxury of choosing just one educational theory, but rather must evaluate all theories and select the principles and applications which would best suit this particular learning context.

**Learn by Doing and the Doer Effect.** The Acrobatiq courseware used in this research study was originally developed at Carnegie Mellon University’s Open Learning Initiative (OLI). In a study by Lovett et al. (2008), the OLI Statistics courseware was observed to accelerate student learning (an 8-week course compared to a traditional 15-week course) while students also gained an average of 18-points higher over the course of the semester compared to the traditional course. Analysis showed that when used as a stand-alone asynchronous resource with no instructor present, students performed as well as a traditional instructor-led course. When the Statistics courseware was used in a hybrid form, students learned more effectively and efficiently with higher learning outcomes compared to the traditional course. After Acrobatiq was launched from OLI in 2013, this Statistics course was recreated on a new platform with new content, adaptivity, assessments, and homework to build upon and extend the learning benefits initially identified at OLI.

The primary learning science methodology behind the design of the courseware being studied is *learn by doing*, which is achieved through the engagement with formative practice (Van Campenhout, Johnson, & Olsen, 2021a). Each lesson page has formative practice questions interspersed with the content. After students read small chunks of content, they are presented with questions to answer on that topic. These formative questions allow students to practice at
the point of learning. Questions have immediate, targeted feedback and students may continue to answer until they arrive at the correct response.

Learning by doing has been used to describe different kinds of learning engagement so it is important to clarify how learning by doing is applied in this courseware. Learning by doing has been used to describe a range of minimally guided learning experiences, such as problem-based learning or constructivist learning (Kirschner et al., 2006). It has also been used to describe—quite literally—learning by doing via virtual reality for hands-on STEM content (Chen et al., 2020). In this context, learning by doing is a method of engaging the learner in the learning process by providing formative practice at frequent intervals. Formative practice engages students in learning by doing and increases learning gains for students of all ages and in diverse subjects; and while this method benefits all students, it can benefit low-performing students most of all (Black & William, 2010).

The learning by doing methodology is central to the “doer effect.” The doer effect is the learning science principle that asserts the amount of interactive practice questions a student does is more predictive of learning than reading text or watching video alone (Koedinger et al., 2016; Van Campenhout, Johnson, & Olsen, 2021; 2022). Five courses developed at OLI were used in research studies to investigate the relationship between doing formative practice and learning outcomes (Koedinger et al., 2015; 2016; 2018). In the first study (Koedinger et al., 2015), researchers investigated the relationships between reading text, watching video, and doing formative practice by comparing students who worked through a psychology Coursera MOOC accompanied by the OLI course. The critical finding was that doing practice had six times the effect size on learning outcomes than reading text or watching video alone. In the second study,
(Koedinger et al., 2016) used a statistical model controlled for the amount of reading, watching, and doing within units to reading, watching, and doing outside of the unit.

If the causal explanation is correct, then the amount of doing a student chooses to engage in during a unit should be predictive of their performance on that unit assessment above beyond any effect of the amount of doing outside that unit. In contrast, if some general trait is an explanation then the amount of doing outside a unit should be equally predictive (or more because there is more data outside a unit) of that unit’s assessment results as the amount of doing within that unit.

Statistically, a regression model should reveal no within-unit effect above and beyond the outside-unit effect. (p. 389)

The within-unit doing remained a large and significant predictor when controlling for other choices, consistent with a causal interpretation (Koedinger et al., 2016). These results showed that the doer effect could not be entirely attributed to a third variable, such as student motivation. These causal results were replicated among other OLI courses using causal data mining algorithms (Koedinger et al., 2018). Causality is critical to identify in order to put forth a learning methodology for wider use with a high degree of confidence.

The replication of doer effect research is also imperative to recommend this method at scale with confidence. Replication of causal findings is also important to increase the validity of the method and generalizability of results. Replication itself is critical, as a large proportion of published research in the social sciences has not been replicated, and studies that cannot be replicated are cited more frequently than those that can (Serra-Garcia & Gneezy, 2021). In an analysis of macroeconomics courseware used at a large online university, the same statistical design was used to replicate the causal doer effect findings (Van Campenhout, Johnson, &
Olsen, 2021; 2022). The correlational and causal doer effect models were also successfully replicated using unit tests from courseware as well as the final exam in a follow-up paper, and this study also found the doer effect was not accounted for by student demographics (Van Campenhout, Johnson, & Olsen, 2021b). This research is directly related to this study, as the same type of courseware is being analyzed.

**Formative Practice.** The formative practice questions integrated with the content in this courseware act as no-stakes practice testing. The technique of practice testing is specifically defined as follows: “Note that we use the term *practice testing* here (a) to distinguish testing that is completed as a low-stakes or no-stakes practice or learning activity outside of class from summative assessments that are administered by an instructor in class, and (b) to encompass any form of practice testing that students would be able to engage in on their own,” (Dunlosky et al., 2013, p. 29). The types of activities that fall within this category can range from flashcards to practice problem sets to actual practice tests. The formative practice are questions which provide no-stakes practice that students can complete on their own—and are separate from high-stakes summative assessments—are a type of practice testing.

While there are many studies involving practice testing in a variety of settings, differing materials, testing intervals, etc. (for examples, see Dunlosky et al., 2013), there are studies which look specifically at the relationship between practice testing and restudy on learning outcomes. Karpicke and Roedinger (2007) provided undergraduate students with a passage of text to study. They followed up with one group completing a free-recall practice test and a second group having a second study session of the passage. After a week, those students completed a free recall assessment and the group which had the recall practice test scored higher (56%) than those who did the restudy session (42%). In another study to compare practice testing with restudying,
Karpicke and Roedinger (2008) compared undergraduate performance on practice recall followed by either additional practice testing or restudying. After a week, the group who continued practice tests performed better (80%) than those who only followed initial practice with additional study (36%). These studies show that practice testing outperforms rereading content, and that continuing to use practice testing will also continue to benefit students. In a review of the literature on practice testing, Dunlosky et al. (2013) found that practice testing has a high utility and has been demonstrated across a wide range of materials, formats, learner ages, and retention intervals.

**Feedback.** Feedback for formative practice is important for the effectiveness of this technique. In the Acrobatiq courseware, feedback is provided for each answer option that explains why that choice is correct or incorrect to provide additional guidance and another opportunity for learning. Dunlosky et al. (2013) found that in studies on practice testing which did not find the same benefits of practice tests over restudying, no feedback was used; practice testing with feedback outperforms practice testing alone. Incorporating feedback has been shown to have other specific advantages for students. In an early intelligent tutoring system (LISP) at Carnegie Mellon University, feedback was found to minimize the time it took students to learn content (Anderson et al., 1989). In an analysis of Statistics courseware developed at Carnegie Mellon, immediate targeted feedback was also shown to reduce the time it took students to reach a desired outcome (Lovett et al., 2008). Regarding the immediacy of feedback, Anderson et al. (1989) argued that immediate feedback reduced the amount of time students spent attempting to correct mistakes, and increased the comprehension of the correct answer. Immediate feedback was also found to increase student satisfaction (Schaeffer et al., 2016).
The type of feedback used for formative questions matters. Feedback is commonly categorized into three distinct types: knowledge of response (KR), knowledge of correct response (KCR) and elaborative feedback (Shute, 2008; Schaeffer et al., 2016; Huang et al., 2015). Only providing KR feedback is the least effective method as the learner is provided no information on how to improve, though research shows KCR is only slightly more effective (Schaeffer et al., 2016; Van der Kleij et al., 2015). Elaborative feedback typically includes KR or KCR in addition to explanations, hints, strategies, etc. and therefore becomes an additional form of instruction which is more effective than KR or KCR alone (Schaeffer et al., 2016). Anderson et al. (1989) found that repeat errors were reduced using explanatory feedback (37%) over no feedback (60%). Within a web-based multimedia module, students who received corrective feedback had higher achievement and also perceived less cognitive load (Huang et al., 2015). The learning by doing methodology that incorporates formative practice with immediate elaborative feedback was foundational to the courseware developed at Carnegie Mellon’s Open Learning Initiative (OLI), and this method was found to increase learning gains and do so in less time than traditional course materials (Lovett et al., 2008).

Summary

The literature reviewed in this chapter—while not exhaustive to the entire body of research on each topic—outlines the context for the goals of this research study and provides comparisons for the results of Chapter 4. Both mindset and self-efficacy have been found to relate to many positive outcomes for students, however, these findings have not always been successfully replicated or identified in the research. This literature on self-beliefs—especially the mixed findings—highlights the relevance of the question of whether the relationship of self-beliefs and learning outcomes is evident in a courseware environment. The use of courseware as
the learning resource in this investigation is also a key feature for the research design, as the courseware utilizes learning methods that need to be considered in conjunction with the self-beliefs. The recent research on courseware as a learning resource guided this research and the decisions made for analysis in the Chapter 3.
Chapter 3: Research Methods

Overview

There is considerable research on self-beliefs, as sampled in Chapter 2, but little research on whether the relationships these beliefs have with their described learning benefits is evident in emerging educational technologies such as courseware. This chapter outlines the research design and methods used in this study, with the goal of answering the following research questions:

- What is the relationship between mindset and self-efficacy?
- What is the relationship between student self-beliefs and learning metrics in courseware?

Research Design

The quantitative research design used in this research study is a nonexperimental correlational design (Creswell & Creswell, 2017). This study is nonexperimental as no controls were used on the population. The correlational design is used to measure the relationship between two or more variables (Creswell & Creswell, 2017). Additionally, the research design is influenced by the type and scale of data collected by the courseware. The emergence of online learning has provided large volumes of high-quality data (Goldstein & Katz, 2004). As Koedinger et al. (2016) states, “the increasing availability of process and outcome data from online courses makes it possible to investigate the generalizability of associations between learning method and outcomes” (p. 1). This type of data has inspired research communities focused on educational data mining and learning analytics, which “are concerned with exploring the increasing amounts of data now becoming available on learners, toward providing better information to instructors and better support to learners” (Baker & Inventado, 2016, p. 84). Specific research designs have emerged within EDM (Baker, 2015), including correlation mining which “is the area of data mining that attempts to find simple linear relationships
between pairs of variables in a data set” (Baker & Inventado, 2016, p. 89). While these data were not the result of a controlled experiment which would provide high internal validity, data sets from natural learning contexts, such as this one, are valuable tools for learning analytics to provide external validity (Koedinger et al., 2016). By analyzing a large dataset gathered by the courseware platform, this research evaluates the mindset and self-efficacy variables as they occur naturally in a student’s use of their learning resource.

There are two types of correlational design being used in this study. The first is explanatory design, in which the researcher attempts “to explain the association between or among variables” (Creswell, 2012, p. 341). All participants are evaluated as a group without any experimental controls, and a score is determined per participant for each variable. While the procedures of this research design allow the researcher to draw conclusions from the results, these “conclusions do not establish a probable cause-and-effect (or causal inference) relationship” (p. 341). While this research seeks to determine what relationships exist between the variables selected, it does not propose any causal relationship between those variables. In this study, an explanatory design is used to identify the relationships between self beliefs and learning metrics using relatively simple correlational and descriptive statistics.

The second form of correlational design used in this study is a prediction research design, in which predictor variables are used to anticipate outcomes (Creswell, 2012, p. 342). For example, “in this form of research, the researcher uses one or more predictor variables and a criterion (or outcome) variable. A prediction permits us to forecast future performance, such as whether a student’s GPA in college can be predicted from his or her high school performance” (p. 356). In prediction design, the predictor variables and outcome variables are often measured at different points in time. There are more advanced types of analyses that can predict
correlational relationships between predictor variables and outcome variables, such as mixed-effects linear regression models (also called hierarchical linear models) (p. 256). This study uses regression models to predict the likelihood of outcome variables, given mindset and self-efficacy.

**Research Methods**

**Population.** The population for this study consisted of students enrolled at three different statewide institutions. The first was a public land-grant research university, the second a public research university, and the third a public community college—all accredited by the Middle States Commission on Higher Education. These institutions shared a research grant and a common goal of standardizing specific courses to guarantee students that credit could transfer between institutions. The data for this study were collected from a Probability and Statistics courseware run at all three institutions. As the common learning resource, all instructors across the three institutions had the same training and policy guidance. This similar context for the courseware usage provided a level of consistency matched by the courseware being the exact same across all sections.

**Sample.** As a non-experimental study using a historical data set, the sample of students in this data set is a convenience sample, a type of non-probability sampling. As the courseware was the learning resource for the introductory probability and statistics course, the data set included all students who were enrolled in this course as part of their program of study. There were a total of 4,055 students, with varying numbers at each institution, included in this convenience sample. There were no demographic or other student characteristics included in the data set. However, this sample provided a wide range of student characteristics because it included *all* undergraduate students across several degree programs who took the introduction to probability
and statistics course at these three institutions. It is also notable that the population of students from which this convenience sample was drawn includes students from both community colleges and universities, increasing the diversity of the total population, compared to a sample only drawn from a single institution.

**Data Collection.** The method of data collection for this historical data set was the Acrobatiq courseware platform. The data are owned by Acrobatiq and written approval to use this data was obtained from the Director of Research and Development (Appendix A). This was a primary data analysis, as the raw data had not been previously analyzed (Brewer, 2011). The raw data stored in the database was primarily clickstream data from student interactions with the courseware, including login events, survey answers, page visits, formative question responses, adaptive question responses, summative assessment responses, and learning estimate events.

The data collected were only associated with a numerical identifier per student to maintain anonymity. No details were known about the gender, age, race, etc. of the study participants. This demographic information was not gathered by the platform and so could not be used as variables for analysis (for more details, please refer to the Ethics Statement in Chapter 1).

The Probability and Statistics courseware included the same elements described in Chapter 2. There was a unit on learning strategies followed by four content units. Each content unit was separated into topical modules, and each module was made up of a series of lesson pages. After the lessons in the module was a formative adaptive activity and a graded summative assessment, as seen in Figure 3.
The lesson pages themselves consisted of content—text, video, interactive simulations—as well as formative practice opportunities. The formative practice was interspersed throughout the page, as seen in Figure 4. Students were able to answer the question as many times as they liked, and received immediate, elaborative feedback each time. The formative practice was tagged with the relevant learning objective, and data for each question was used in the predictive learning model, drive the adaptive features, and report progress in the instructor dashboard (Van Campenhout et al., 2020).
Figure 4

A lesson page from Probability and Statistics showing a formative practice question in between sections of content.

Measures. This correlational study design investigates the relationships between variables. As this study is a correlational and not an experimental design, the variables are
described as predictor and outcome variables. According to Creswell & Creswell (2017), “Predictor variables are similar to independent variables in that they are hypothesized to affect outcomes in a study, but dissimilar because the researcher is not able to systematically manipulate a predictor variable” (p. 51). The outcome variables are similar to dependent variables in that they are believed to be outcomes or results from the predictor variables (Creswell & Creswell, 2017). In this study, student self-beliefs are the predictor variables, as they are not something which can be assigned to students, but rather are naturally held beliefs. The two self-belief predictor variables are mindset and self-efficacy. The outcome variables are the learning metrics captured by the platform as students interact with the courseware. Learning metrics include multiple variable measures, as it is suggested to measure multiple outcome variables (Creswell & Creswell, 2017). The learning metrics include number of study sessions, the amount of formative practice attempted, first-attempt accuracy of formative practice, and summative assessment scores. These learning metrics are categorized into engagement metrics and performance metrics: engagement meaning how students used the courseware resource, and performance meaning how well students did on formative and summative questions, as shown in Figure 5. These measures are interpreted from the raw data collected by the platform during the analysis phase.
Figure 5

A diagram of the variables in this correlational study design.

For this analysis, the student self-beliefs were collected as part of the courseware content in the first unit on learning strategies. The mindset survey used was the PERTS growth mindset survey developed at Stanford University (PERTS, 2015). This survey consisted of three items and has been shown to have acceptable validity and reliability (Hanson, 2017; Farrington et al., 2012). The self-efficacy survey was the New General Self-Efficacy Scale (NGSE) consisting of eight items, developed by Chen et al., (2001). Items for both surveys are included in the Appendix.

**Research Questions.** The literature on growth mindset and self-efficacy noted that those with growth mindset and those with high self-efficacy had higher persistence, effort, and learning outcomes (Tseng et al., 2020; Zimmerman, 2000; Paunesku et al., 2015; Aronson et al., 2002; Blackwell et al., 2007). The purpose of this study is to identify if similar relationships can
be seen between these self-beliefs and learning metrics in a courseware environment. As noted in Chapter 1, there are several broad research questions which address the relationships between the predictor variables and outcome variables. Here, I include additional finer-grained sub-research questions that help to guide the analysis.

1. What is the relationship between mindset and self-efficacy?
   a. How do students’ level of self-efficacy correlate to their mindset?
   b. How does each mindset/self-efficacy combination group relate to engagement and learning metrics?

2. What is the relationship between student self-beliefs and learning metrics in courseware?
   a. How do students’ mindsets relate to engagement and learning metrics?
   b. How do students’ self-efficacy relate to engagement and learning metrics?

Data Analysis. The data analysis took place using tools and software I have become familiar with from my professional experience. The analysis was done in Jupyter Notebook, using Python 3.0 and common packages for R, pandas, numpy, and matplotlib and others. These tools are free and open access and are commonly used among researchers in this field.
The data was not ready for analysis when I retrieved it; there were several steps to complete before the data was ready for analysis, called data wrangling (Poulson, 2019). The first step of the preparation was the collection; I retrieved the data from the database, stored it on my computer locally, and imported it into the notebook so I could begin to work with it. Similarly, I installed the Python and R packages needed to do the analysis and ensured they were all able to run.

**Preliminary Data Analysis.** Before I could begin to explore the research questions through statistical analysis, the survey responses for each student needed to be scored and grouped to determine the predictor variable groups. This process included: collecting all student responses to the survey items; removing students who did not complete all survey items, or selected multiple responses for any items; assigning a score to each response in the Likert scale; and grouping students according to their overall score on the survey instruments.
Once students had been grouped for the predictor variables—mindset and self-efficacy—I computed Cronbach’s alpha, a coefficient of reliability. Cronbach’s alpha determines how closely related the set of items are as a group. In preparation for the first research question, I also used Pearson’s correlation coefficient to determine the correlation of the predictor variable raw scores. In addition, I grouped students by both self-belief categories, which had four combinations as shown in Table 1. I ran the Jaccard Similarity coefficient to compare the predictor variables as categorical variables (Karabiber, 2021). These preliminary analyses illuminated how mindset and self-efficacy are related, e.g., is growth mindset highly correlated with high self-efficacy, or is there little to no correlation?

<table>
<thead>
<tr>
<th></th>
<th>Low Self-Efficacy</th>
<th>High Self-Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Mindset</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth Mindset</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Data Analysis.** The preliminary data analysis helped to answer the first research question on how mindset and self-efficacy were related. To answer the second question on how self-beliefs are related to learning metrics, I used descriptive and correlational statistics common to the explanatory research design described previously. With students grouped according to their self-beliefs, for each predictor variable I analyzed descriptive statistics for each outcome variable. For instance, I calculated the mean number of formative practice questions students with growth mindset completed, and similarly with students with fixed mindset. Basic
descriptive statistics and visualizations helped determine a baseline understanding of what the data were like for each group of students on each learning metric.

To identify if there were statistically significant differences between groups of students for each predictor variable, I planned to use a MANOVA test. Because there are several outcome variables, a MANOVA would evaluate all outcome variables for a predictor variable at once (Tables 2 and 3), which would produce more rigorous results than performing a t-test for each outcome variable separately. Prior to performing the MANOVA, I used the Shapiro-Wilk test for multivariate normality, as this is an assumption of the MANOVA.

Table 2

*The MANOVA for predictor variable mindset and outcome variables.*

<table>
<thead>
<tr>
<th>Study Sessions</th>
<th>Amount of Practice</th>
<th>Practice Accuracy</th>
<th>Summative Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth Mindset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed Mindset</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3

*The MANOVA for predictor variable self-efficacy and outcome variables.*

<table>
<thead>
<tr>
<th>Study Sessions</th>
<th>Amount of Practice</th>
<th>Practice Accuracy</th>
<th>Summative Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Self-Efficacy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Self-Efficacy</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The MANOVA would analyze the predictor variable as a categorical variable; students were grouped into a category no matter their score in the appropriate range. This would be beneficial for comparing the main effect of the predictor variables. However, students who selected “somewhat agree,” generating a mid-range score, may behave differently than students who selected “strongly agree,” generating a high score on the survey instruments. A regression model could account for student scores as a continuous variable and show the changes in outcome for changes in scores. Furthermore, there were many individual course sections included in the data, and instructor implementation policies for courseware can have a meaningful impact on student engagement (Van Campenhout & Kimball, 2021). The course section introduced a hierarchy in the data and could be treated as a random variable in a regression model to control for the differences in course section. Therefore, the next step was to do a set of mixed methods linear regressions with random effects. These models could account for both the individual student’s score on the survey instruments and course section when considering the outcome variables. The mixed-effects linear regressions were part of the predictive correlational research design, as they explored a predictive relationship between the predictor and outcome variables.

The outcome variables in the data set were not entirely independent. For example, the amount of formative practice a student chose to do could influence summative assessment scores. Another mixed effects linear regression model was used with the summative assessment score as the outcome of the model to identify how the other outcome variables, as well as the predictor variables, related to student scores.

It is worth noting that the significance level was set at .10, which is higher than a 95% confidence interval. This decision was influenced by the ongoing discussion of changing
research practices in the social science regarding significance levels. In an observational study, Pritchet et al., (2016) reported that in recent decades, research has increasingly reported marginal significance in psychological and social sciences. Regarding the range of marginal significance: “Sampling the first p value labeled as marginally significant from each article (n = 459) revealed that the vast majority (92.6%) of marginal p values fall between .05 and .10” (p. 1038). Pritchet et al. conclude that the increasing use of marginal significance in recent decades indicates a change in research practices among psychologists. This increase in the acceptance of marginal significance may be “representative of a graded, Fisherian interpretation of p values, according to which hard cutoffs are thought to be arbitrary” (p. 1041). Pritschet et al. also note that these practices could also be indicative of questionable research practices such as insufficient power, manipulation of data, and unreproducible results.

As the use of marginal significance continues to evoke discussion and debate, I chose to use the conventional significance level of .10. While there are limitations to this analysis (such as using a convenience sample with student self-selection to answer surveys), there are thousands of students included in the data set, so statistical power is not a factor in the analysis. This study values replicable research, as the analysis seeks to identify if similar outcomes are evident in courseware as in previous research. Furthermore, the doer effect principle that is part of the contributing theoretical framework for this study was reproduced in similar courseware, and these doer effect studies both reported marginal significance (Koedinger et al., 2016, Van Campenhout, Johnson, & Olsen, 2021). In this analysis, the $p$-values were reported in their entirety, and significance levels noted in the discussion.
Summary

The purpose of these data analyses is to answer the research questions proposed: what is the relationship between mindset and self-efficacy, and what is the relationship between student self-beliefs and learning metrics in courseware? Using a historical data set for an explanatory and predictive correlational design, this study aims to use multiple types of statistical analyses to determine the relationship between self-beliefs and learning metrics. These analyses provide insight into what relationships exist between the predictor variables and outcome variables, from basic descriptive statistics to the regression models designed to account for the complexities of the data.
Chapter 4: Results

Overview

This chapter performs the investigation of the relationship between student self-beliefs and learning metrics. The sections to follow provide a step-by-step guide for the data analysis run—from general preparation, to tasks necessary for analysis, to final statistical analyses. Throughout this chapter, the exact code to complete tasks is often provided, as well as screenshots of the output in the Jupyter Notebook, to make the analysis process easier to follow.

The data preparation section discusses how the predictor variables were determined from the survey responses, as well as how the outcome variables were combined into a single data frame. The data preparation section also includes preliminary data analysis for the predictor variables.

The data analysis section is organized into three sections: descriptive statistics, MANOVA, and regression models. These three sections of analysis each work toward answering the research questions put forth in Chapter 3.

Data Preparation

The first step in the analysis was to prepare the data. The analysis was done in Jupyter Notebook, using python and pandas as the primary languages for tasks such as organizing and describing the data. In addition, I installed several commonly used third party python packages needed to conduct the analysis. R was also installed to be used for running statistical analysis. The data files were stored as .json files and were retrieved from a local data directory on my computer. The preparation required for my computer and data analysis environment required a sizable portion of the total effort needed for analysis.
**Predictor Variables.** An aggregated data frame was assembled of survey data from all students in all sections for the mindset survey (mindset_qpa) and self-efficacy (self_efficacy_qpa). This was done so that all data can be analyzed at once, while maintaining the course section as a column in the data frame. The data frame in Figure 7 shows the course key, student, survey question ID, attempt, timestamp, response, and score.

**Figure 7**

*The mindset_qpa data frame.*

<table>
<thead>
<tr>
<th>student</th>
<th>question_part</th>
<th>attempt_number</th>
<th>timestamp</th>
<th>response</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>5b860a9800000097f83f3738</td>
<td>1</td>
<td>2018-08-29 11:17:37.799000-04:00</td>
<td>E</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>5b860a9800000097f83f3738</td>
<td>1</td>
<td>2018-08-29 11:17:44.527000-04:00</td>
<td>E</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>5b860a9800000097f83f3738</td>
<td>1</td>
<td>2018-08-29 11:17:49.219000-04:00</td>
<td>E</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>5b82b59f600000097f7f40902a</td>
<td>1</td>
<td>2018-08-29 13:41:09.897000-04:00</td>
<td>F</td>
<td>6</td>
</tr>
<tr>
<td>18</td>
<td>5b82b59f600000097f7f40902a</td>
<td>1</td>
<td>2018-08-29 13:41:17.564000-04:00</td>
<td>F</td>
<td>6</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>45584</td>
<td>5b8f052be16a3a11d3a3e3f9a</td>
<td>1</td>
<td>2018-09-10 23:09:11.225000-04:00</td>
<td>C</td>
<td>3</td>
</tr>
<tr>
<td>45585</td>
<td>5b8f052be16a3a11d3a3e3f9a</td>
<td>1</td>
<td>2018-09-10 23:09:38.195000-04:00</td>
<td>B</td>
<td>2</td>
</tr>
<tr>
<td>45615</td>
<td>5b91303a15f51a1796e529a2</td>
<td>1</td>
<td>2018-09-11 08:23:46.783000-04:00</td>
<td>C</td>
<td>3</td>
</tr>
<tr>
<td>45616</td>
<td>5b91303a15f51a1796e529a2</td>
<td>1</td>
<td>2018-09-11 08:23:54.996000-04:00</td>
<td>C</td>
<td>3</td>
</tr>
<tr>
<td>45617</td>
<td>5b91303a15f51a1796e529a2</td>
<td>1</td>
<td>2018-09-11 08:23:58.743000-04:00</td>
<td>C</td>
<td>3</td>
</tr>
</tbody>
</table>

5688 rows x 6 columns

There were 75 individual course sections containing 4,055 students included in the data set. This student total was simply the number of students who were enrolled in those course sections, so it was expected that the final number of students would be smaller once selection criteria have been applied further on. The survey items each have a unique question ID, and as the version of the courseware was identical across all institutions, these question IDs were also identical. The question IDs were grouped and labeled as the survey instrument questions, as seen in Figure 8.
The next step was to apply the data selection criteria to the data frame. The first selection criterion was to remove any students who appear in multiple course sections, as I was unable to determine the reason for multiple attempts at the course and how that would influence responses.

The next step was to only include students who properly completed each survey. At this point there were 3,750 students who attempted the mindset questions and 3,681 students who attempted the self-efficacy questions. Given that the constraints of the platform question types available made these questions be implemented as a multiple choice multiple select, students could select more than one option. The survey was not intended to allow for multiple answers per question, so those students were removed. Students who did not answer all questions in the survey were also removed. This left 3,429 students who answered the mindset survey and 3,050 students who answered the self-efficacy survey. Finally, only students who answered both surveys according to these constraints were included for a total of 2,855 students.

There was another type of selection criteria to apply related to the number of summative assessments students completed. Given that this was a natural learning context and there were no experimental controls, every student who completed the surveys were included to this point. However, students may have not completed any or all of the courseware for various reasons (i.e., dropped the course for personal reasons). Many studies on mindset have a group of students who
all take the survey and then complete a final assessment (for example, Claro et al., 2016). So for this study, I compared students who had similar completion of summative assessments to avoid conflating self-beliefs with other reasons for not completing the courseware. By adding another selection criteria that students must complete 9 of 10 quizzes, the final number of students was 1,896.

In order to group students according to the predictor variables, the survey Likert scale options were scored. The PERTS mindset survey (PERTS, 2015; Hanson, 2017) had three items “fixed-worded” questions (e.g., ‘You can learn new things but you can’t really change your intelligence’). As was done in other research studies using this instrument (Burnette et al., 2018; Yeager et al., 2019), each answer option was assigned a score from 1 to 6. As the answer options in this survey were from strongly agree (1) to strongly disagree (6), lower scores are associated with a fixed mindset and higher scores are associated with a growth mindset. The NGSE instrument had positively worded questions (e.g., I will be able to achieve most of the goals I have set for myself’) and the Likert scale responses range from strongly agree to strongly disagree. If the responses were scored from 1 to 6 as the mindset responses were, then lower scores would identify higher self-efficacy and higher scores would identify lower self-efficacy. As this was counter to the mindset variable scoring, I reverse-scored these items (strongly agree = 6, strongly disagree = 1) so that lower scores identify lower self-efficacy and higher scores identify higher self-efficacy.

After the items were scored, the predictor variable groups could be assigned. The responses for each survey were a six-option Likert scale, so the responses were on a symmetrical scale, evenly split between agreement and disagreement. Because it was a requirement that all students answer all questions, the scores could only start at 3 for mindset and 6 for self-efficacy.
The mindset survey had three answer options, and so the midpoint of scores was an impossible score. However, the self-efficacy instrument had an even number of items, so the midpoint was possible (e.g., a combination of 3, 3, 3, 3, 4, 4, 4, 4 scores). While it was unlikely students would split their responses between agree and disagree in that way, if it did happen, those students would be included in the low self-efficacy category. The final predictor variable groups were defined as follows:

- Growth mindset: students with mindset scores between 10.5 and 18
- Fixed mindset: students with mindset scores between 3 and 10.5
- High self-efficacy: students with self-efficacy scores between 28 and 48
- Low self-efficacy: students with self-efficacy scores between 8 and 28

**Figure 9**

*The code defining the predictor variable groups based on survey scores.*

```
growth = mindset_scores > 10.5  
fixed = mindset_scores < 10.5  
high_se = self_efficacy_scores > 28  
low_se = self_efficacy_scores <= 28
```

Figure 9 shows how these predictor variable groups were defined in the notebook. Note that this code shows the predictor variables in two different ways. `Mindset_scores` and `self_efficacy_scores` were the self-beliefs recorded as a continuous variable using raw scores. The categorical variables ‘growth’ or ‘fixed’ were defined from the continuous variables. After the predictor variable categories were defined, the students could be categorized according to their combined scores on the survey items. Figure 10 shows the code used to combine scores for each student and the abridged output.
Figure 10

The code to calculate the mindset scores.

```
In [40]:
score_mindset = lambda question_scores: question_scores.sum()

mindset_scores = mindset_qpa.groupby('student').score.apply(score_mindset)
mindset_scores

Out[40]:
       student
5890ba0597a6a343653d3865  17
59a6c6a4c68a96c6fc4d5b  18
59a9987284f3fc5e1b1f80ac  17
59ac4ee73829ab5d4bbbdaf5   7
59ac9cbd4c68a904e0e349e5  18
   ...
6022ea7c0000091560cf3063  15
6024695ae16a3a66840c4530  14
6026a0b51e99fb4b4597db28  15
6033e431e16a3a07966afc8c  13
604579e61e99fb63800d8365  18
Name: score, Length: 2855, dtype: int64
```

With the scores for mindset and self-efficacy calculated for each student, we can plot the distribution of scores as a histogram for both self-beliefs (Figures 11 and 12).

Figure 11

*A histogram of mindset scores.*
In total, there were 1,511 students with a growth mindset and 385 students with a fixed mindset. Roughly one in five students had a fixed mindset. These results are consistent with the literature, where there was a higher proportion of students with a growth mindset. Blackwell et al. (2007) reported a mean of 4.45 on a six item average, indicating a mean growth mindset for the 373 students included in the study. Similarly, McCabe et al. (2020) showed a mean mindset score of 4.45 on the same six question scale for the 229 students in their study. However, these results are contrary to the ratio of fixed and growth mindset found by Claro et al. (2016) when evaluating all tenth grade students in Chile. In that study, more students held a fixed mindset for all but the top two deciles of family income wherein the groups were roughly equivalent. Other cited studies from Chapter 2 did not report the number or ratio of students in each group (Paunesku et al., 2015; Burnette et al., 2028).

Surprisingly, there were 1,869 students with high self-efficacy and 27 students with low self-efficacy. These extremely unbalanced groups were unexpected. Nearly all students (98.6%) interpreted themselves as having positive general self-efficacy to some degree.

**Cronbach’s Alpha.** As a preliminary data analysis check on the reliability of the survey instruments, I determined Cronbach’s alpha on each instrument. This coefficient alpha is used to
test for internal consistency when items are scored as continuous variables (e.g. strongly agree to
strongly disagree) (Creswell, 2012). In order to do this test, I needed to construct a new data
frame. The data frame mindset_qpa (Figure 7) had a row for every student-question
interaction. Cronbach’s alpha measures how correlated student responses are on the survey
items, so I had to have a data frame with each response grouped by student. To do this, I:

- created a new data frame named mindset_cronbach
- set the index to be the student
- grouped the question parts
- collected the score series as the new columns of the data frame

**Figure 13**

*The pandas code and partial output used to create the mindset_cronbach data frame.*

```python
In [66]: mindset_cronbach = pd.DataFrame(index=mindset_qpa.student.unique())

In [67]: for qp, g in mindset_qpa.groupby('question_part'):
   ...:     g = g.set_index('student')
   ...:     print(qp, len(g))
   ...:     display(g.score.head())
   ...:     mindset_cronbach[qp] = g.score

5b84f82b807e2b466e5774a5 1896

student
5b860a9800000097ff83f3738 5 5b82b5960000097ff740902a 6 5b8621f315f61a5c0f927dbf 6 5b874e37e16a3a459511bdc6 6 5b84d28115f61a5c0db25aa7 5
Name: score, dtype: int64
```

The final mindset_cronbach data frame had each of the 1,896 students as rows and
their responses to each of the three items in the mindset survey as the columns. This gave me the
correct format to run the Cronbach’s alpha. I repeated these same steps to create a self-efficacy data frame.

**Figure 14**

The *mindset_cronbach* data frame.

<table>
<thead>
<tr>
<th>In [68]:</th>
<th>mindset_cronbach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Out[68]:</td>
<td><img src="image" alt="Table" /></td>
</tr>
</tbody>
</table>

With the data frame in the proper format to be able to run the Cronbach’s alpha test, I loaded the R package with the Cronbach’s alpha test (`library(ltm)`). I also converted the python CSV file so that R could read the files, as python and R do not use the same file formats. Figure 15 shows the code for this task, which was also done for any new data frame which needs to be used for a test in R.
The code used to read the CSV files into R.

```r
mindset_cronbach.to.csv( '/tmp/to_r.csv', index=False )
```

```r
mindset_cronbach <- read.csv( '/tmp/to_r.csv' )
```

```r
str( mindset_cronbach )
'data.frame': 1896 obs. of 3 variables:
$ X5b84f02b807e2b466e5774a5: int 5 6 6 6 5 4 2 5 3 1 ...
$ X5b84f02b807e2b466e577513: int 5 6 6 6 5 5 5 3 3 3 ...
$ X5b84f02b807e2b466e577581: int 5 6 6 5 5 5 5 3 3 3 ...
```

The code to run the test was `cronbach.alpha(mindset_cronbach, CI=TRUE)`, which specifies the data frame created and to calculate the confidence interval. The PERTS mindset survey had three items, a sample of 1,896 students, and an alpha of .834. The bootstrap confidence interval (set at a sample of 1,000), had 95% confidence that the alpha was between .834 and .863. This alpha was not only highly correlated, but there was a high degree of confidence that the alpha was between a very small range. While different texts claim differing ranges, the ideal coefficient is between .70 and .90, as an alpha value too low indicates poor internal interrelatedness between items and an alpha too high may indicate items are redundant (Tavakol & Dennick, 2011). As a check for validity, I compared this value to other similar studies, such as Burnette et al. (2018), which reported an alpha of .86. Paunesku et al. (2015) used only two items from this set and had an alpha of .84. Blackwell et al. (2007) used an extended six item survey (which includes the three used in this study) and had an internal reliability score of .78.

The same code was used to run Cronbach’s alpha on the self-efficacy data frame (se_cronbach). The NGSE had eight items, a sample of 1,896 students, and an alpha of .905. This alpha was even higher than the mindset results, meaning student responses were even more
highly correlated. The 95% confidence interval was between .898 and .912. This alpha was at the
top end of the preferred range, showing high internal interrelatedness. For comparison, Chen et al. (2001) found an alpha of .85 and .86.

**Correlation of Predictor Variables.** With the predictor variables scored and grouped for students, descriptive and correlational statistics can help answer the research question: What is the relationship between mindset and self-efficacy? The student mindset and self-efficacy raw scores can be correlated to each other. I used a simple correlational command that uses the Pearson method. The correlation result was .1464 \( (p = 1.507 \times 10^{-10}) \). According to Cohen et al. (2007), correlations between 0 and .2 show no relationships, .2 to .35 only very slight relationships, .35 to .65 have a moderate relationship with limited use, and .65 to .85 have strong relationships useful for some prediction. This result of .146 falls in a range of *no relationship.*

This effectively answers the sub-question 1a, that there is no relationship on the correlation between mindset and self-efficacy.

For a visualization, I created a scatterplot of the mindset and self-efficacy scores (Figure 16), using the code: `student_data.plot('mindset_score', 'formative_mean_score', kind = 'scatter').` This scatterplot similarly provided a visual display that there was no discernable linear correlation between these predictor variables.
Figure 16

*A scatterplot of mindset and self efficacy scores.*

**Predictor Variable Quadrants.** Research question 1b was to group students by both self-beliefs (so now in four groups) and investigate how those groups related to learning metrics. At this point, each predictor variable group had been defined, so students could be grouped according to both self-beliefs. Using `pd.crosstab(growth, high_se)` I determined how many students were in each quadrant, as shown in Table 4.

**Table 4**

*The number of students in each predictor variable cross-group.*

<table>
<thead>
<tr>
<th></th>
<th>Low Self-Efficacy</th>
<th>High Self-Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Mindset</td>
<td>11</td>
<td>374</td>
</tr>
<tr>
<td>Growth Mindset</td>
<td>16</td>
<td>1495</td>
</tr>
</tbody>
</table>

As so few students identified having low self-efficacy, there was an unbalanced grouping of students across categories. The majority of students (1,495) were in the growth mindset/high
self-efficacy group. The next largest group (374) was the fixed mindset/high self-efficacy group. The low self-efficacy group was split into groups of 11 students for fixed mindset and 16 for growth mindset. Because the self-efficacy and mindset groups were so imbalanced, analyzing the low self-efficacy/mindset groups would not produce reliable results with so few students. Therefore, there was no analysis of outcome variables for these groups, as was my initial plan for my sub-research question on how each mindset/self-efficacy combination group related to engagement and learning metrics.

The predictor variable groups were converted from raw scores to categorical variables. The growth_mindset category, for example, was a boolean value, either true (1) or false (0). To evaluate the relationship of these categorical variables, I used the Jaccard Similarity coefficient (or Jaccard index). The Jaccard Similarity is a common data science test that can be used to find the similarity between two sets of binary data (Karabiber, 2021). The results are between 0 (no similarity) and 1 (perfect similarity). I ran a test for each of the four groups, for example, \texttt{jaccard\_score( high\_se, growth, )}, with results listed in Table 5. Given that all students in the data set had either 0 or 1 for each predictor variable, the resulting values are similar to the overall group percentage of the total student population.

\textbf{Table 5}

\begin{tabular}{|l|c|c|}
\hline
 & Low Self-Efficacy & High Self-Efficacy \\
\hline
Fixed Mindset & .027 & .199 \\
\hline
Growth Mindset & .011 & .793 \\
\hline
\end{tabular}
Outcome Variables. The next phase of this data analysis was to analyze the outcome variables. To this point, I had only run analyses on the predictor variables, looking specifically at the student responses to the mindset and self-efficacy surveys. In order to work on the outcome variables I needed to gather all the required data into a single data frame. From the individual course section data files, I pulled the data for: student, course key, question part, attempt number, timestamp, response, and score.

This resulted in a data set of 7,469,066 data points. However, there were data that needed to be excluded. For example, I did not need the responses for the surveys, students that were excluded from the analysis, or any attempt at the formatives or summatives after the first attempt (recall that formative questions can be answered repeatedly and every answer is recorded). I also excluded activity types which were not formatives or summatives, such as adaptive activities. Once these filters were set, the resulting data set had 1,215,456 data points.

Figure 17

The student_data data frame showing a subset of columns.

<table>
<thead>
<tr>
<th>self_efficacy_score</th>
<th>growth_mindset</th>
<th>high_self_efficacy</th>
<th>n_sessions</th>
<th>n_formative_questions</th>
<th>formative_mean_score</th>
<th>n_summative_questions</th>
<th>summative_mean_score</th>
</tr>
</thead>
<tbody>
<tr>
<td>34.0</td>
<td>True</td>
<td>True</td>
<td>52</td>
<td>92</td>
<td>0.70522</td>
<td>441</td>
<td>0.616780</td>
</tr>
<tr>
<td>31.0</td>
<td>True</td>
<td>True</td>
<td>39</td>
<td>214</td>
<td>0.739645</td>
<td>49</td>
<td>0.754694</td>
</tr>
<tr>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>23</td>
<td>151</td>
<td>0.701987</td>
<td>270</td>
<td>0.591398</td>
</tr>
<tr>
<td>39.0</td>
<td>True</td>
<td>True</td>
<td>24</td>
<td>601</td>
<td>0.752080</td>
<td>108</td>
<td>0.814815</td>
</tr>
<tr>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>99</td>
<td>589</td>
<td>0.609349</td>
<td>184</td>
<td>0.639024</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>48</td>
<td>150</td>
<td>0.533333</td>
<td>31</td>
<td>0.451613</td>
</tr>
<tr>
<td>38.0</td>
<td>True</td>
<td>True</td>
<td>59</td>
<td>241</td>
<td>0.688797</td>
<td>98</td>
<td>0.725588</td>
</tr>
<tr>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>18</td>
<td>301</td>
<td>0.697674</td>
<td>26</td>
<td>0.576923</td>
</tr>
<tr>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>1</td>
<td>11</td>
<td>0.630064</td>
<td>0</td>
<td>NaN</td>
</tr>
<tr>
<td>NaN</td>
<td>NaN</td>
<td>NaN</td>
<td>6</td>
<td>72</td>
<td>0.611111</td>
<td>5</td>
<td>0.800000</td>
</tr>
</tbody>
</table>

The data frame created for the remainder of the analysis was named student_data, and it included the predictor variables as both a categorical and continuous (boolean value and raw score) as well as the data for the outcome variables.
Data Analysis

**Descriptive Statistics.** The descriptive statistics for this data provided insights into the general nature of the data. These statistics were the first step in understanding my second research question: What is the relationship between student self-beliefs and learning metrics in courseware?

To start, because students self-selected on whether or not to complete the survey, I used the un-filtered data set to identify mean outcome variables for students who completed the surveys, had incomplete surveys, and did not attempt the surveys, as shown in Table 6. This was to determine if there were dramatic differences in mean values between these groups, as self-selection was a limiting factor of this historical data set. I did this by using the original `student_data` data frame, grouped by instrument, and calculating the mean `(student_data.groupby( 'instruments' ).mean())`.

**Table 6**

*Mean outcome variables for students who completed, had incomplete, or did not complete the surveys.*

<table>
<thead>
<tr>
<th>Survey Status</th>
<th>Students</th>
<th>Mean Number of Sessions</th>
<th>Mean Number of Formatives</th>
<th>Mean Formative Score</th>
<th>Mean Number of Summatives</th>
<th>Mean Summative Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>2855</td>
<td>53.817</td>
<td>414.697</td>
<td>.670</td>
<td>242.756</td>
<td>.690</td>
</tr>
<tr>
<td>Incomplete</td>
<td>227</td>
<td>45.581</td>
<td>337.026</td>
<td>.645</td>
<td>217.264</td>
<td>.645</td>
</tr>
<tr>
<td>No Attempt</td>
<td>842</td>
<td>38.868</td>
<td>212.461</td>
<td>.652</td>
<td>177.167</td>
<td>.650</td>
</tr>
</tbody>
</table>

Students who completed the surveys had the highest mean sessions, number of formative and summative questions answered, and highest mean scores on formatives and summatives.
Students who started but did not complete the survey had the next highest mean number of formatives and summatives attempted, but had slightly lower mean scores than the group who did not attempt the surveys. The spread of mean scores for both formatives and summatives was within a few hundredths across groups, indicating there were not dramatic mean differences in groups who selected not to complete the survey (e.g. students who did not complete or attempt the survey did not have mean scores around .30 by comparison).

**Predictor Variables as Categorical Variables.** The next descriptive statistics to review were for the students who completed the survey. I grouped students based on their survey scores to create the predictor variable groups. For each group, I calculated the mean values for the outcome variables (for example:

```python
student_data.groupby('growth_mindset').n_sessions.mean()
```

Table 7 shows the mean values for mindset.

**Table 7**

*Mean values for the outcome variables for the mindset predictor variable.*

<table>
<thead>
<tr>
<th></th>
<th>Number of Sessions</th>
<th>Number of Formatives</th>
<th>Formative Score</th>
<th>Summative Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Mindset</td>
<td>61.377</td>
<td>479.727</td>
<td>.671</td>
<td>.723</td>
</tr>
<tr>
<td>Growth Mindset</td>
<td>62.503</td>
<td>498.432</td>
<td>.675</td>
<td>.712</td>
</tr>
</tbody>
</table>

Generally, the mean values for the outcome variables were very close between fixed and growth mindset groups. The students who identified as having a growth mindset had barely higher mean study sessions (about 1 more), completed more formative questions (about 20), and more summative questions (about 2). While students with a growth mindset technically had
higher mean formative scores, they had slightly lower mean summative scores. Overall, the mean values were very close between the mindset groups.

When we consider self-efficacy, the mean values for the outcome variables were similarly close (Table 8). Like with the mindset variable, students with high self-efficacy had slightly higher mean values on the outcome variables, with the exception of mean formative question scores. It is also noteworthy that these values were very close to those seen for the mindset predictor variable. However, because the low self-efficacy group was so small (27) compared to the high self-efficacy group (1,869), these mean values are not a reliable measure of difference between groups.

Table 8

*Mean values for the outcome variables for the self-efficacy predictor variable.*

<table>
<thead>
<tr>
<th></th>
<th>Number of Sessions</th>
<th>Number of Formatives</th>
<th>Formative Score</th>
<th>Summative Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Self-Efficacy</td>
<td>53.259</td>
<td>447.592</td>
<td>.678</td>
<td>.681</td>
</tr>
<tr>
<td>High Self-Efficacy</td>
<td>62.404</td>
<td>495.313</td>
<td>.675</td>
<td>.715</td>
</tr>
</tbody>
</table>

When I classified students according to their survey responses as having high or low self-efficacy and growth or fixed mindset, I made these predictor variables categorical variables. When looking at the mean values for the outcome variables, there were only slight differences between these groups. Reviewing descriptive statistics for these categorical variables may have obscured more nuanced findings that could be gained from looking at the predictor variables as continuous variables.
*Predictor Variables as Continuous Variables.* In order to analyze self-beliefs as continuous variables, I used the raw scores for students on the surveys which provided a range of values. Mindset had a possible range from 3 to 18, and self-efficacy had a range from 8 to 48. I used the continuous values to view the outcome variables at quartiles using the code `student_data.describe()`. Results in Table 9 show what students did at multiple points in the range for each variable, not just the overall mean.

**Table 9**

The quartile descriptive statistics for students who completed the surveys.

<table>
<thead>
<tr>
<th></th>
<th>Mindset Score</th>
<th>Self-Efficacy Score</th>
<th>N Sessions</th>
<th>N Formative Questions</th>
<th>Mean Formative Score</th>
<th>Mean Summative Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>1896</td>
<td>1896</td>
<td>1896</td>
<td>1896</td>
<td>1896</td>
<td>1896</td>
</tr>
<tr>
<td>Mean</td>
<td>13.172</td>
<td>39.795</td>
<td>62.274</td>
<td>494.633</td>
<td>.675</td>
<td>.714</td>
</tr>
<tr>
<td>Std</td>
<td>3.218</td>
<td>4.848</td>
<td>33.737</td>
<td>241.873</td>
<td>.107</td>
<td>.141</td>
</tr>
<tr>
<td>Min</td>
<td>3.0</td>
<td>20.0</td>
<td>6.0</td>
<td>5.0</td>
<td>.312</td>
<td>.172</td>
</tr>
<tr>
<td>25%</td>
<td>11.0</td>
<td>37.0</td>
<td>41.0</td>
<td>291.0</td>
<td>.600</td>
<td>.626</td>
</tr>
<tr>
<td>50%</td>
<td>14.0</td>
<td>40.0</td>
<td>55.0</td>
<td>544.5</td>
<td>.684</td>
<td>.732</td>
</tr>
<tr>
<td>75%</td>
<td>15.0</td>
<td>43.0</td>
<td>76.0</td>
<td>727.0</td>
<td>.752</td>
<td>.823</td>
</tr>
<tr>
<td>Max</td>
<td>18.0</td>
<td>48.0</td>
<td>302.0</td>
<td>844.0</td>
<td>.955</td>
<td>.994</td>
</tr>
</tbody>
</table>

The quartiles show an expected range for each outcome variable, from the minimum to maximum. The mindset scores started at the minimum score of 3, but then jumped to 11 at the 25th percentile—just above the divide for growth mindset. The 50th and 75th percentiles were close at 14 and 15, with the maximum at the maximum mindset score of 18. The minimum self-efficacy score was 20, with each percentile increasing toward the maximum score. The range for
the number of formative questions answered showed a surprising minimum of only 5 questions, while the 25th percentile was 291 questions and each subsequent quartile increased by over a hundred questions. The mean formative scores began at a minimum of .31 and increased to a maximum of .95.

The quartile breakdown showed a range of values for each variable that vary dramatically from lows to highs. This was a different view of the variables than the predictor variable group means showed. However, it did not indicate if there were any patterns between these values (i.e., did the low mindset scores correlate to the low number of formative questions attempted?). Therefore, I created a correlation matrix (Table 10) using the student_data data frame.

### Table 10

A correlation matrix of all predictor and outcome variables.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mindset Score</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Self-Efficacy Score</td>
<td>.146***</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. N Sessions</td>
<td>.053*</td>
<td>.039.</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. N Formative Questions</td>
<td>.036</td>
<td>.01</td>
<td>.374***</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Mean Formative Score</td>
<td>.022</td>
<td>-.082***</td>
<td>.124***</td>
<td>-.027</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>6. Mean Summative Score</td>
<td>-.048*</td>
<td>-.041.</td>
<td>.153***</td>
<td>.167***</td>
<td>.475***</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Significance codes: *** p < .001, ** p < .01, * p < .05, . p < .10

The values for the correlation coefficients range from -1.0 to 1.0 with 0 indicating no correlation at all (Cohen et al., 2007). This correlation matrix produced the same correlation coefficient of mindset and self-efficacy (.146) performed with the Pearson’s correlation previously. The remaining correlations coefficients for both predictor variables on all outcome
variables were very small (under .06) and three combinations were even negative: self-efficacy with mean formative and summative score and mindset with mean summative score. The number of formative questions and mean formative score was also slightly negative. The largest coefficients were for number of sessions and number of formative questions (.374), and mean formative score and mean summative score (.475).

I used the student_data data frame to create scatterplots to look at the relationship between predictor and outcome variables. Figure 18 shows the correlation between mindset score and the number of formative questions completed. There is no discernable line in the graph, but rather a wide range of questions completed at each mindset score. Each scatterplot for the predictor variables and outcome variables show similar results (Figures 19–25).

**Figure 18**
*A scatterplot of mindset scores and number of sessions.*

**Figure 19**
*A scatterplot of mindset scores and number of formative questions.*
Figure 20

A scatterplot of mindset scores and formative mean scores.

Figure 21

A scatterplot of mindset scores and summative mean scores.

Figure 22

A scatterplot of self-efficacy scores and number of sessions.
Figure 23
A scatterplot of self-efficacy scores and number of formative questions.

Figure 24
A scatterplot of self-efficacy scores and formative mean scores.

Figure 25
A scatterplot of self-efficacy scores and summative mean scores.
These graphs showed that there was no clear pattern between the predictor variable and outcome variables. For both mindset and self-efficacy, there were fewer low scores that tended to have a smaller vertical spread. Overall, these visualizations showed that most self-belief scores had a full range of values for the outcome variables. These visual representations provided insight into the nature of the data, which was further scrutinized using statistical tests in the following sections of Chapter 4.

**MANOVA.** The goal of the MANOVA test was to compare the predictor variables on all outcome variables at once to determine if there were statistically significant differences between the predictor variable groups. In order to perform the MANOVA, there are several assumptions about the data which must be satisfied (Data Novia, 2021). One of those assumptions is that the groups are normally distributed, which is tested with the Shapiro-Wilk test. To run the Shapiro-Wilk test, I used the python code: `stats.shapiro( student_data.mindset_score.dropna()`. The mindset variable had a statistic of .958 at a $p$-value of 7.08e-23. The self-efficacy variable had a statistic of .971 at a $p$-value of 3.948e-19. The results of the Shapiro-Wilk test found that both mindset and self-efficacy variables were not normally distributed.

Another assumption of the MANOVA is for the groups to be homogenous, or equal. The Box’s M test is a parametric test used to compare variation in multivariate samples (Data Novia, 2021). However, this test can only be run if the data is normally distributed, which it is not. While I did not use a statistical test, the total students in the self-belief quadrants shown in Table 4 indicated that the groups were extremely unbalanced. Because the Shapiro-Wilk test found both predictor variables to be non-normally distributed, and the groups are observationally unbalanced, the assumptions of the MANOVA were not supported and the results would not be reliable. Therefore, the MANOVA test could not be used for this analysis. While there are non-
parametric MANOVA approaches, I was unable to identify a test that could be run given the nature of the data and the tools I used in this analysis.

**Regression Models.** The final type of analysis to run on the data were regression models. The regression models provide another way to understand the second research question—how are self-beliefs related to learning metrics? Unlike the MANOVA which treated the self-belief predictor variables as categorical variables, the regression models use the self-belief scores for students, treating the predictor variables as continuous variables (Keith, 2019). Whereas the MANOVA requires balanced groups, the regression models do not (Keith, 2019). Furthermore, Keith (2019) argues that multiple regression is more appropriate than tests such as ANOVA for nonexperimental research where the predictor variables cannot be manipulated (p. 18).

I needed to complete a few steps in order to do the regression models. First, I converted the `student_data` dataframe that I created using pandas into a format that can be read by R. This was the same process used to convert the data frames I created for the Cronbach alpha tests. Then I installed and loaded the R library package `lme4`. With these steps complete, I could begin creating the regression models.

**Ordinary Linear Regression.** I began with the simplest linear regression model relevant to gain a baseline understanding of the relationship between self-beliefs and formative practice. Before I look into outcome variables such as mean scores, I wanted to simply investigate if there was a relationship between self-beliefs and how much formative practice students attempted. I used an ordinary fixed effects linear regression model that used the mindset and self-efficacy predictor variables as inputs in the model and the number of formative questions as the output. The R formula for this model was as follows:
model <- lm( n_formative_questions ~ mindset_score + self_efficacy_score, 
              data=student_data )

The output from R included significance levels from .001 to .10 (any \( p \)-value higher than .10 is not considered significant). The results of this model (Table 11) showed that neither mindset nor self-efficacy were significant variables for the amount of formative practice students did. However, this model did not take the random effect of the course section into account.

**Table 11**

*The regression results for number of formative questions.*

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
<th>T-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>450.369</td>
<td>48.724</td>
<td>9.243</td>
<td>&lt;2e-16 ***</td>
</tr>
<tr>
<td>Mindset Score</td>
<td>2.677</td>
<td>1.745</td>
<td>1.534</td>
<td>.125</td>
</tr>
<tr>
<td>Self-Efficacy Score</td>
<td>0.226</td>
<td>1.158</td>
<td>0.195</td>
<td>.845</td>
</tr>
</tbody>
</table>

Significance codes: *** \( p < .001 \), ** \( p < .01 \), * \( p < .05 \), . \( p < .1 \)

*Mixed Effects Linear Regression Models.* The `student_data` data frame was aggregated from many different course sections, and I knew from previous research that different implementation practices can impact how students use the courseware (Van Campenhout & Kimball, 2021). Therefore, the course section could be used as a random effect to account for those differences by course section in the model.

*Number of Formative Questions.* I repeated the first linear regression model I did to look at how adding the course section as a random variable might impact the outcomes. In this mixed model, I used the mindset score, self-efficacy score, and course section as input variables to look at the number of formative practice students completed. The R formula for this test was:
model <- lmer( n_formative_questions ~ mindset_score + self_efficacy_score
               + (1|course_key),
               data=student_data )

An additional statistical test was needed to determine significance—a Likelihood Ratio Test. As described by Winter (2013):

Likelihood is the probability of seeing the data you collected given your model.
The logic of the likelihood ratio test is to compare the likelihood of two models with each other. First, the model without the factor that you’re interested in (the null model), then the model with the factor that you’re interested in…Now you have two models to compare with each other – one with the effect in question, one without the effect in question. We perform the likelihood ratio test using the anova() function. (pp. 11–12).

For this analysis, I created another null model which removed an input variable and did the Liklihood Ratio Test to compare those models. This determined if the variable removed was significant to the model. In this instance, I did this twice—removing either mindset or self-efficacy. The R formula to do one of the follow-up test, for example, was:

lme.null <- lmer( n_formative_questions ~ self_efficacy_score
               + (1|course_key),
               data=student_data )
anova( lme.null, model )
The results for this model (Table 12) show that the mindset score predictor variable was significant for the number of formative practice questions done. Self-efficacy score was not significant, and had a negative estimate. While mindset was significant at \( p < .1 \), it was interesting to note that in this model—when accounting for the course section as a random effect—the values changed and mindset score went from not significant to significant.

**Table 12**

*Regression model results for the number of formative practice completed.*

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
<th>T-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>496.447</td>
<td>49.590</td>
<td>10.011</td>
<td>2.2e-16 ***</td>
</tr>
<tr>
<td>Mindset Score</td>
<td>2.979</td>
<td>1.700</td>
<td>1.752</td>
<td>.079 .</td>
</tr>
<tr>
<td>Self-Efficacy Score</td>
<td>-0.098</td>
<td>1.125</td>
<td>-0.087</td>
<td>.932</td>
</tr>
</tbody>
</table>

Significance codes: *** \( p < .001 \), ** \( p < .01 \), * \( p < .05 \), . \( p < .10 \)

I repeated this same model for each outcome variable.

*Number of Sessions.* Using the number of sessions as the outcome, the results of this model show mindset was significant at \( p < .05 \) while self-efficacy was not significant.

**Table 13**

*Regression model results for the number of sessions.*

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
<th>T-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>60.923</td>
<td>6.858</td>
<td>8.883</td>
<td>2.2e-16 ***</td>
</tr>
<tr>
<td>Mindset Score</td>
<td>0.505</td>
<td>0.220</td>
<td>2.295</td>
<td>.027 *</td>
</tr>
<tr>
<td>Self-Efficacy Score</td>
<td>0.153</td>
<td>0.146</td>
<td>1.052</td>
<td>.292</td>
</tr>
</tbody>
</table>

Significance codes: *** \( p < .001 \), ** \( p < .01 \), * \( p < .05 \), . \( p < .10 \)
**Mean Formative Score.** Using the mean formative score as the outcome, the results of this model show mindset was significant at $p < .10$ while self-efficacy was negative and significant at $p < .001$.

**Table 14**

*Regression model results for the mean formative scores.*

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
<th>T-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.703</td>
<td>0.022</td>
<td>31.888</td>
<td>2.2e-16 ***</td>
</tr>
<tr>
<td>Mindset Score</td>
<td>0.001</td>
<td>0.001</td>
<td>1.877</td>
<td>.062</td>
</tr>
<tr>
<td>Self-Efficacy Score</td>
<td>-0.002</td>
<td>0.001</td>
<td>-3.297</td>
<td>.0009 ***</td>
</tr>
</tbody>
</table>

Significance codes: *** $p < .001$, ** $p < .01$, * $p < .05$, . $p < .10$

**Mean Summative Scores.** Using the mean summative scores as the outcome, the results of this model show neither mindset nor self-efficacy were significant.

**Table 15**

*Regression model results for mean summative scores.*

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
<th>T-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.706</td>
<td>0.029</td>
<td>24.604</td>
<td>2.2e-16 ***</td>
</tr>
<tr>
<td>Mindset Score</td>
<td>-0.001</td>
<td>0.001</td>
<td>-0.859</td>
<td>0.388</td>
</tr>
<tr>
<td>Self-Efficacy Score</td>
<td>-0.001</td>
<td>0.001</td>
<td>-0.837</td>
<td>0.401</td>
</tr>
</tbody>
</table>

Significance codes: *** $p < .001$, ** $p < .01$, * $p < .05$, . $p < .10$

Mindset studies have had mixed results on significance for summative assessments. Blackwell et al. (2007) found mindset to be a significant predictor of middle school students’ math scores. Claro et al. (2016) found significant relationships between mindset and standardized test scores for all tenth graders in Chile, after controlling for other contributory variables. Burnette et al. (2018) did not evaluate summative assessments for a single course, but rather
grade point average for the tenth graders included in the study; they did not find a direct effect of mindset treatment on grades between control and intervention groups, but did identify a significant mediated effect on grades. Paunesku et al. (2015) found that the mindset intervention had a 6.4% increase in pass rate for the 367 students in the at-risk group, which was significant. Yeager et al. (2019) studied 6,320 lower-achieving students and found mindset interventions successful and the students in the intervention achieved higher GPAs in core classes.

**Summative Scores with All Variables.** The mean summative scores outcome variable required additional consideration. One of the primary learning science methods in the courseware was learn by doing, intended to elicit the doer effect. The Koedinger et al. (2016) causal doer effect results were replicated in this type of courseware (Van Campenhout, Johnson, Olsen, 2021; 2022). Therefore, we know there is a relationship between the outcome variables of formative practice and summative scores. In order to take into account the related variables, I created a model that uses all other predictor and outcome variables while still controlling for the course section, as shown below.

```r
model <- lmer( summative_mean_score ~ mindset_score + self_efficacy_score + n_sessions + n_formative_questions + formative_mean_score + n_summative_questions + (1|course_key), 
               data=student_data )
```

Like the mixed effects models above, for the Liklihood Ratio Test, an additional model and ANOVA test were created to determine the significance of each variable. The results in Table 16 show a mixture of outcomes. Mindset was significant at $p < .05$, but is negative. Self-
efficacy was not significant, as in the previous model. The number of sessions was not significant either. However, the number of formative questions and mean formative scores were significant at \( p = 2.2e-16 \). This shows that the amount of formative questions and their accuracy were highly significant to mean student summative scores. These findings on the formative outcome variables were in line with expectations.

**Table 16**

*Expanded regression model results for the mean summative scores.*

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard Error</th>
<th>T-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.336e-01</td>
<td>3.524e-02</td>
<td>9.466</td>
<td>2.2e-16 ***</td>
</tr>
<tr>
<td>Mindset Score</td>
<td>-2.085e-03</td>
<td>8.221e-04</td>
<td>-2.537</td>
<td>0.011 *</td>
</tr>
<tr>
<td>Self-Efficacy Score</td>
<td>3.739e-04</td>
<td>5.446e-04</td>
<td>0.687</td>
<td>0.492</td>
</tr>
<tr>
<td>N_Sessions</td>
<td>2.465e-04</td>
<td>9.270e-05</td>
<td>2.659</td>
<td>0.492</td>
</tr>
<tr>
<td>N Formative Questions</td>
<td>1.233e-04</td>
<td>1.193e-05</td>
<td>10.330</td>
<td>2.2e-16 ***</td>
</tr>
<tr>
<td>Mean Formative Score</td>
<td>5.671e-01</td>
<td>2.543e-02</td>
<td>22.302</td>
<td>2.2e-16 ***</td>
</tr>
</tbody>
</table>

Significance codes: *** \( p < .001 \), ** \( p < .01 \), * \( p < .05 \), . \( p < .10 \)

**Summary**

This results chapter outlined the data analysis process, from preparation to results. This research project was consistent with the data science adage that the majority of data analysis is in the preparation. Not only did my computer need to be set up to do the analysis using the tools selected, but there was preparation needed in order to analyze the variables in this study.

The data analysis included three main components: descriptive statistics, a MANOVA, and regression models. The descriptive statistics provided insight into the nature of the data, looking at the predictor variables both as categorical and continuous variables. The MANOVA could not be run due to violations of assumptions, as detected with the Shapiro-Wilk test. The
regression models provided insight into the relationships between the predictor variables and outcome variables, and identified when variables were significant.
Chapter 5: Summary

Overview

The purpose of this quantitative study is to investigate the relationships between student self-beliefs and learning metrics using student data from online courseware. Many researchers have found that specific self-beliefs of mindset and self-efficacy have had positive relationships to learning outcomes (Blackwell et al., 2007; Claro et al., 2016; Paunesku et al., 2015; Yeager et al., 2019). There has also been research identifying null results (McCabe et al., 2020; Li & Bates, 2019). This study examines the relationship between these self-beliefs as well as their relationship to learning metrics in courseware, expanding the existing research to a courseware learning environment.

This chapter includes an interpretation of the major findings from the data analysis of Chapter 4, as well as the implications of these within the theoretical frameworks of this study. For the purpose of assisting data visualization for practitioners and researchers, I have included infographics of key findings throughout the chapter. This chapter also includes a discussion on the limitations of this study, the implications of these findings for future theory, research, and practice and a brief conclusion.

Interpretation of the Findings

**Predictor Variables: Self-Beliefs.** My first research question is: What is the relationship between mindset and self-efficacy? The results of Chapter 4 provided some interesting—and surprising—results starting from simply identifying students’ self-beliefs. For mindset, there were 1,511 students with a growth mindset and 385 with a fixed mindset, which is roughly one of five, or 20% having fixed mindsets. Figure 26 provides a visualization of this ratio. Two other studies reported mean mindset scores at 4.45 (Blackwell et al., 2007; McCabe et al., 2020).
These mean scores are the total score divided by the number of questions, so to compare directly to this study, this mean mindset calculation would be $13.1729 / 3 = 4.39$. These are similar mean scores given the wide variation in student ages and contexts across these studies, indicating that this mindset ratio is consistent with other research. It is also notable that of the two studies used to compare mean mindset scores, Blackwell et al. (2007) found positive relationships between growth mindset and outcomes, while McCabe et al. (2020) did not.

**Figure 26**

*The ratio of students with a growth or fixed mindset.*

The results of the self-efficacy survey showed very unexpected results. There were 1,869 students who identified as having high self-efficacy and 27 students who selected low having self-efficacy, as represented by Figure 27. These are very imbalanced groups, with only 1.4% of the students in the low self-efficacy group. These results are contrary to examples from the literature on self-efficacy (Chen et al., 2001; Yeo & Neal, 2006). However, in a recent national report by the Center for Community College Student Engagement (CCCSE), tens of thousands of community college students answered a set of self-efficacy questions (CCCSE, 2019). While the
question set was not the NGSE instrument used here, a similar question (“When facing difficult tasks, I am certain I will accomplish them”) showed similar trends in responses. Out of 78,010 responses, 77% agreed, 19% were neutral, and only 4% disagreed. The 2018 CCCSE data from this study showed very few students select disagree for positively written self-efficacy statements, but a portion did select neutral, which was not an option in the NGSE items presented in this study. Given the large volume of community college student responses, the CCCSE study provides a similar benchmark for the small percentage of students in this study who identified as having low self-efficacy.

**Figure 27**

*The ratio of students with high and low self-efficacy.*

These unexpected results caused me to wonder, was there something that could have impacted student self-selection for the self-efficacy results? The scoring of the survey items for the NGSE survey cannot be manipulated, as they are positively written statements with three “agree” and
three “disagree” options. Nearly all students interpreted themselves as having positive general self-efficacy to some degree. I also noted that not only are the survey item statements positively written (e.g. “I will be able to successfully overcome many challenges.”), but the “agree” options were first in the order of answer choices. Could it be possible that even if students did have low self-efficacy that they read the responses and did not want to select the disagree options? In future studies it would be worthwhile to reverse the order of answer choices to determine if the order of answer options shifts responses.

However, I would not hypothesize that the order of agree to disagree answer options is causing students to answer untruthfully en masse. It is important to note that we are looking at the students who self-selected to answer, so there could have been additional students with low self-efficacy not included in this analysis. It could also be that students do, in fact, have high general self-efficacy. Future research could include specific self-efficacy to focus on the specific learning task at hand, including neutral options as used by the Center for Community College Student Engagement (2019). As noted by Yeo and Neal (2006), self-efficacy can be conceptualized at varying levels and it may not exert the same effect at all levels.

For this research question I asked a sub-question to investigate the correlation of these predictor variables. It is not the case that students who have fixed mindsets also had low self-efficacy, and vice versa. In reality, nearly all students had high self-efficacy, no matter their mindset. Essentially, students felt confident in their abilities no matter what they believed about their minds. The correlation coefficient for mindset and self-efficacy was 0.146, which falls in the range of no correlation (Creswell, 2012). Overall, I found no relationship between the mindset and self-efficacy variables.

These unbalanced predictor groups limited the ability to answer my second research sub-question: How does each mindset/self-efficacy combination group relate to engagement and learning metrics? Once the low self-efficacy group of 27 students was divided between growth and fixed
mindset, those quadrants were 11 and 16 students, respectively. The standard minimum number of students for reliable analysis is 30 (Creswell, 2012). Therefore, my second sub-question to investigate these paired predictor variable groups was not possible, as the low self-efficacy groups were too small to produce reliable or meaningful results.

**Descriptive Statistics.** The second research question in this investigation is: What is the relationship between student self-beliefs and learning metrics in courseware? The first step to answering this research question was through descriptive statistics. The descriptive statistics for the predictor and outcome variables did not provide any glaring evidence of relationships between the self-belief groups and learning metrics. There were no wide differences in mean values for the outcome variables. Instead, the mean values for the predictor groups showed very close values on nearly all outcome variables, with growth mindset and high self-efficacy having slightly higher values than fixed mindset and low self-efficacy the majority of the time.

The correlations of the variables also did not reveal meaningful relationships. When looking at the visualizations, there were no compact diagonal lines on the scatterplots of predictor and outcome variables. The scatterplots of the relationship between both self-beliefs as continuous variables showed that for each outcome variable, each score for the predictor variable had high and low values.

The correlation matrix of all variables verified that there were no strong correlations between any variables. There were several very small negative coefficients that—while still in the no relationship range—are interesting to consider. Both predictor variables had negative coefficients (-0.046 and -0.041) for mean summative score, meaning that higher mindset and self-efficacy scores had slightly lower scores. This was the same case with self-efficacy and mean formative scores. Considering that nearly all students identified having high self-efficacy, this is a plausible result considering the variation within a group of 1,869 students. There were
three correlation coefficients between .12 and .17: number of sessions with mean formative and summative scores, and number of formative questions with mean summative scores. Again, while not classified as a meaningful relationship—Creswell (2012) identified a minimum of .2 for a slight correlation—these coefficients are larger than most others and positive. These relationships also made sense to be positive and larger than the others; more sessions had a positive coefficient with mean scores and more formative practice had a positive coefficient with higher summative scores.

There were only two coefficients that were in the range to be useful for limited prediction and classified as having a moderate relationship (Cohen et al., 2007; Cohen, 1988). The number of sessions and number of formative questions had a coefficient of .37. This matches intuition that to a certain extent, as the number of sessions increases, so do the number of formative questions students do. The highest coefficient was between mean formative score and mean summative score at .48. As mean formative scores increased, so did mean summative scores. The strongest correlation coefficient of the matrix was between the practice and assessment, not the self-beliefs and assessment.

Both the mean values for the outcome variables as well as the correlations showed no strong indication of a relationship between self-beliefs and learning metrics. And yet, would we expect drastically different mean values for predictor variable groups or clear diagonal scatterplots? No, we would not. If there were dramatic differences or clear trends, that would be fantastic, but in this case, not realistic. It is hard to imagine that mindset and self-efficacy could produce such results, given how many other factors likely contribute to what students do in the courseware and how they perform on formative and summative questions. It would be unreasonable to expect that simply having a fixed mindset equals minimal work and low scores.
while having a growth mindset equals a lot of work and high scores. So the results of the descriptive statistics may not seem meaningful at a glance, but they are likely realistic given that there are many other external variables that contribute to what students do.

**The Regression Models.** While the descriptive statistics helped to characterize the data for the explanatory correlation design, the regression models primarily gave insights into the predictive correlational design for these variables. The regression models use the predictor variables as continuous variables to determine which variables are significant for a particular outcome variable. In this study, the unbalanced nature of the predictor variables as categorical variables meant the data did not meet the assumptions of the MANOVA test. The regression models are not limited in the same way, and can also account for random effects, such as course section.

**Engagement Metrics.** I did both an ordinary linear regression in addition to a mixed effects linear regression using the number of formative questions as the outcome variable to get a better understanding of the influence of the random variable course section. As I knew from previous research (Van Campenhout & Kimball, 2021), engagement in the courseware could vary a great deal based on instructor choices and course policies. Therefore, I created one model to look at the effect of mindset and self-efficacy on the number of formative questions and another model to do the same thing, but also account for the course section as a random variable. The results are directly compared in Table 17.
Table 17

A comparison of significance for number of formatives using fixed vs mixed effect model.

<table>
<thead>
<tr>
<th></th>
<th>Fixed Effects Model Significance</th>
<th>Mixed Effects Model Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mindset</td>
<td>.125</td>
<td>.08</td>
</tr>
<tr>
<td>Self-Efficacy</td>
<td>.845</td>
<td>.932</td>
</tr>
</tbody>
</table>

Significance codes: *** $p < .001$, ** $p < .01$, * $p < .05$, . $p < .10$

The fixed effect model shows neither mindset nor self-efficacy had a significant effect on the number of formative questions students attempted. The mixed effects model—which was the same except for the addition of the course section random variable—showed different results. The self-efficacy variable actually became even closer to 1.0 than it was previously. The mindset value became smaller by about .045 and was significant at $p < .10$. The change is notable for two reasons: (a) the significance of the mindset variable could be meaningful to pursue in further research and (b) the controlling for the course section does matter to the results.

The other engagement metric that is a meaningful outcome variable is the number of sessions. How many times a student logs in to work and how many formative questions they choose to do are both engagement metrics. The mixed effects regression model showed that mindset was a significant variable ($p = .022$), while self-efficacy was not. This is interesting because mindset was significant for the number of formative questions students did ($p = .08$), making it an intriguing predictor variable for both major engagement metric outcome variables.

**Performance Metrics.** There are two outcome variables that are performance metrics: how well students did on formative and summative questions. For mean formative scores, the mindset variable was significant ($p = .062$), which is notable given the previous significance for
number of sessions and significance for number of formatives. The combination of these results provide a compelling case for continued future research.

For mean formative scores, the self-efficacy variable was significant ($p < .001$). However, it was negative. What does that mean? The self-efficacy score had a negative effect on mean formative scores. Yet this is not as meaningful as it seems. First, when we look at the mean scores for each group, the low self-efficacy group had a mean of .678 and the high self-efficacy group had a mean of .675. The low self-efficacy group had higher mean scores by a very narrow margin. Second, the low self-efficacy group is only 27 of 1,896 students—all of 1.4% of the total students in this study. Given how large the high self-efficacy group is and how small the actual effect is, we could be observing the small differences among high self-efficacy students. While the result was significant, I do not interpret this as being meaningful in a practical way.

Summative scores are an important performance metric, as these are the assessments that generate a grade and can be used as part of a student’s course grade. Neither mindset nor self-efficacy were significant variables in the simple regression model. However, mean summative scores are a more complex outcome variable because they are not entirely independent from the other outcome variables. Koedinger et al. (2016) found that doing formative practice while reading was causal to learning outcomes, and this research on the doer effect was replicated using this type of courseware as well (Van Campenhout, Johnson, & Olsen, 2021; 2022). Simply put, the formative practice students did would impact their learning outcomes on the summative assessments. Therefore, I created a more complex mixed effects linear regression model. Unlike the previous model, mindset became negative and significant ($p < .05$). Recalling that mean summative scores for the fixed mindset group was .723 while the growth mindset group had a mean score of .712, this small difference became significant in this model. While the reason for
this is not clear, I can point to several complicating factors such as the ratio of the growth to
fixed mindset groups and the interaction of the formative questions. Because, as shown
previously, what was positively significant was both the number of formative questions and
mean score of formative questions (both \( p < .001 \)). These were the only variables which had a
positively significant effect on summative scores.

**Table 18**

*Estimate and significance results from all learning metric regression models (Tables 12–15).*

<table>
<thead>
<tr>
<th></th>
<th>N Sessions</th>
<th>Amount of Practice</th>
<th>Mean Formative Score</th>
<th>Mean Summative Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mindset Score</td>
<td>0.505 *</td>
<td>2.979 .</td>
<td>0.001 .</td>
<td>-0.001</td>
</tr>
<tr>
<td>Self-Efficacy Score</td>
<td>0.153 -0.098</td>
<td>-0.002 ***</td>
<td>-0.001</td>
<td></td>
</tr>
</tbody>
</table>

Significance codes: *** \( p < .001 \), ** \( p < .01 \), * \( p < .05 \), . \( p < .1 \)

As summarized in Table 18, the self-efficacy variable proved not to be positively
significant for the outcome variables, and therefore, not meaningful for immediate practical
application. So few students selected low-self efficacy that the categories were not balanced and
could not be used for certain types of analysis. As a continuous variable, the self-efficacy scores
were not significant in the regression models in all but a single case. These findings indicate that
this self-efficacy variable (as gathered by this instrument) may not be beneficial for further
research and application.

The mindset variable, on the other hand, showed a few encouraging results. In the
regression models, mindset was significant at \( p < .05 \) for the number of sessions, and was
significant at \( p < .10 \) for both number and mean scores of formative practice. While not
significant for summative assessments, these results are still encouraging for future investigation.

Mindset is a meaningful variable for engagement metrics, and the formative practice outcome
variables most strongly predicted mean summative scores.
Implications

The implications of this study should be contextualized within the theoretical frameworks and research identified as foundational to this study. The vast majority of the research discussed in the following sections focus on traditional educational settings, even when technology is used as part of the learning process. If we conceptually compare the design of traditional research, the surveys are given to students, some unknown type of instruction and learning activities are completed, and an assessment or grade is used as the metric of evaluation.

In contrast, courseware captures vast amounts of clickstream data that is not be captured in traditional learning settings. Engagement and learning outcomes are easily gathered and analyzed at a micro level, providing new ways to evaluate educational and learning science principles. Because of this, the analysis in this study provides a unique addition to existing theory.

Self-Efficacy. Nearly all students identified as having high self-efficacy, regardless of their growth or fixed mindset selection. Only 1.4% of students identified having low self-efficacy, and while this surprised me at first, it is mirrored by results from a large study of community college students, indicating that these results may be typical for college students. It is difficult to understand whether the number of students who reported high and low self-efficacy is expected or not, as these numbers are not often reported in published studies (e.g. Tseng et al., 2020). While some research has found that growth mindset was positively related to self-efficacy (Dweck 1999; Komarraju and Nadler 2013), Tseng et al. (2020) found that student mindsets did not have an impact on student self-efficacy towards their online course. The results from this study seem to confirm Tseng et al.’s findings. However, they also found that self-
efficacy was predictive of engagement, which was not the case in this study. The self-efficacy predictor variable was not positively significant for any outcome variable.

These results suggest that general self-efficacy is not a meaningful variable to assess and track in this learning environment. It could be that other measures of task-specific self-efficacy might prove useful as a predictive measure. Yeo and Neal (2006) found the relationship between self-efficacy and task performance to be dependent upon the level of analysis and level of specificity. Using other instruments to measure self-efficacy might provide meaningful insights (Tseng et al., 2020; CCCSE, 2019), but the results of this analysis lead me to believe this instrument for self-efficacy is not useful to pursue further in this learning context.

**Mindset.** Mindset did not prove to be a predictive measure of learning outcomes in the form of summative assessments. This is contrary to the many research studies where mindset was predictive of learning outcomes—Blackwell et al. (2007), Paunesku et al. (2015), and Yeager et al. (2019) all found positive results on course grades while Claro et al. (2016) found positive results on a standardized test. However, other studies have found null results related to learning outcomes. McCabe et al. (2020), Burnette et al. (2016), Li and Bates (2019), and King and Trinidad (2021) failed to find significant relationships between mindset and learning outcomes. As Dweck (2017) noted, there is a danger to oversimplifying and overapplying mindset, and it’s important to understand its complexity. It’s possible the null results for summative assessments in this study could have other unknown related factors.

For example, when the aforementioned studies evaluated learning outcomes, they were most often evaluating course grades, not module-level quizzes. I imagine that using course grades has its own complications because different courses most likely have different graded course components that aren’t standardized across subjects. Mindset is being evaluated against a
high-level outcome in which many types of graded activities likely contribute. In this study, the courseware provides all the learning content, practice, and quizzes. The null mindset results on summative assessments in this study could be related to the frequency and type of summative assessment being used for evaluation.

The mindset variable was significant for the number of study sessions \( (p < .05) \), as well as for the amount and accuracy of formative practice \( (p < .10) \)—all important engagement metrics in the courseware. A recent study by King and Trinidad (2021) used the Educational Longitudinal Study data set and found that mindset was significant for engagement but not for achievement. There are some issues limiting a direct comparison to this study, as this data set uses a one-item mindset evaluation and two-question student and teacher perceived engagement instruments. However, it is notable that growth mindset positively predicted both teacher-rated and student-rated engagement. King and Trinidad posited these results around engagement “provided strong evidence for the role of mindsets in facilitating optimal attitudes towards learning” \( (p. \ 646) \). This study also did not find growth mindset to be significantly related to standardized math scores. In light of these findings, King and Trinidad note:

> Although the non-significant effect of mindset on achievement might be construed as a direct challenge to the validity of research on mindsets, Dweck (2018) argued that even massive educational reforms yield only small effect sizes \( \text{e.g.}, \text{ Hanushek 2011; Nye et al. 2002} \). It is also possible that growth mindsets exert their influence on achievement via other motivational processes \( (p. \ 646) \).

The Center for Community College Student Engagement (2019) included both mindset and self-efficacy surveys in their surveys on student beliefs. While the survey items and benchmarks are different than those used in this study, their general findings were that, “Having
a productive mindset correlates with higher levels of engagement, and this finding holds true across all CCCSE benchmarks. The closer students are to a productive mindset, the more likely they are to be highly engaged and, thus, more successful” (p. 4). This report is more evidence that a single self-belief may not be the sole predictor of success, but that understanding the complex interactions of multiple beliefs could be meaningful on student engagement, and therefore, success.

In this courseware type of learning environment where every click is tracked and all engagement and learning outcomes are recorded for analysis, there is more clarity on the complexity of interaction. Mindset and self-efficacy did not make an impact on summative assessment scores according to these results. However, mindset did have a small significant relationship to some of the engagement metrics. We know those engagement metrics are highly significant to learning. So, this study can add to the literature on these theories by not only illuminating the interaction of these theories with their interaction with learning activity, but suggest that much of their value may be in influencing certain engagement behaviors that in turn could impact learning outcomes.

The Doer Effect. The doer effect was included as a theoretical framework because it is one of the primary learning science principles that drives the design of the courseware. The frequent formative practice embedded within lessons engages students in learning by doing, which has been shown to be causal for learning (Koedinger et al., 2016; Van Campenhout, Johnson, & Olsen, 2021; 2022). This principle needed to be considered in this study because the amount and accuracy of formative practice students do impacts their outcomes. Knowing I wanted to investigate self-beliefs on learning outcomes as well, it was important to be aware of the relationship of formative practice.
In this research, I included a regression model that included all other variables as covariates when using mean summative scores as the outcome, specifically knowing the relationship between formative and summative scores. The amount and accuracy of formative practice were the only positively significant variables for mean summative scores. These findings are consistent with correlational doer effect findings that found formative practice to be significant to outcomes in similar linear regression models (Koedinger et al., 2016; Van Campenhout, Johnson, & Olsen, 2021b).

Learning Engineering. Learning engineering was a framework for this dissertation, as it has shaped my professional and scholarly practice. This research study is the learning engineering process in practice, serving as a first iteration of gathering data and evaluating results. The existing research shaped the purpose and hypothesis for this research, and the data was gathered from a learning environment designed to include the self-belief surveys. The data analysis and results found some relationships between mindset and engagement metrics, but it also uncovered new questions and directions for future research.

The next step in the learning engineering process is the feedback loop back into the instrumentation, design, and implementation. This iteration could be making updates to the self-efficacy survey instrument, changing the survey question type format to facilitate easier response selection, or requiring survey completion to get a more complete picture of student self-beliefs. More complex cycles could include future research that expand on this study, such as intervention research.

Learning engineering also helped me to think systematically and holistically about the practical and theoretical procedures of this study by being aware of the various theories at work in this research. For instance, the doer effect is a learning science principle focused on the
cognitive process of learning. Yet it is also tied to a behavioral issue on what students choose to do in the courseware—which is in turn affected by instructor implementation practice. All of these issues needed to be acknowledged and accounted for in the analysis, as they were aspects of learning within the courseware environment which would interact with this investigation.

Thinking about this research from a learning engineering perspective helped me to maintain a systematic process and consider several different intertwining theories.

**Limitations**

A limitation to this research project is that the data were not collected from a controlled (or quasi) experiment, but rather from a set of naturally occurring courses. While this type of data is valuable for understanding how students behave and learn in natural contexts—and is increasingly common for practitioner research of large-scale, technology derived data sets such as this one—it does introduce specific issues into this study. A convenience sample, as opposed to random sampling, limits external validity. In contrast to an experimental design where a randomly assigned sample of students would complete the survey and later complete an assessment and their scores would be compared to the results for a control group, in this study students were able to choose whether or not to complete the surveys as well as how much of the courseware they completed. Therefore, a large portion of students in the original data set were filtered out according to the selection criteria of completing the surveys and the majority of the assessments. For the students who were filtered out, we do not know their motivation for not completing the surveys, and whether those motivations could be relevant to this investigation.

The descriptive statistics in Table 6 that compared the mean values of the outcome variables between students who completed the survey, didn’t complete, and didn’t attempt, showed there were sizable differences between engagement metrics, and small differences between learning
metrics. Because we only investigated students who completed the surveys, we do not know how those other groups may have contributed to the findings.

Another limitation is the variation and lack of clarity into the implementation practices used for each section included in this data. A data set that includes many different course sections also includes differences in how each course section introduced, used, and graded the courseware. This variation is known to cause differences in engagement (Van Campenhout & Kimball, 2021). The mixed effects linear regression models used in this analysis used the course section as a random effect, thereby controlling for those differences by course section. However, there is still a potentially unknown interaction between self-beliefs and the context of the course in practice.

**Recommendations for Future Practice**

My first—and most practical—suggestion for future practice is to update the delivery method for the survey instruments in the courseware. Some of the selection criteria were needed because the survey items were created with the multiple choice multiple select question type—due to a limitation of the courseware platform question type options. While this question type allows for any selection to be correct, like a survey item, it also allows for multiple options to be selected. As the survey instruments were not written with the intent of having multiple options, students who selected multiple were removed from this analysis. In the future, those students could have been included in the study if the question type were to be updated to properly facilitate responses.

My second recommendation is to identify a way of having all students respond to the survey instruments. While a large portion of students did attempt them, ensuring all students do the surveys would help provide a more holistic interpretation of results. This could be tackled
from two different angles: a technological solution or a teaching solution. There could be a
technology solution that could increase student completion of the surveys such as creating a
completion score for the survey items that would become part of the gradebook, or putting the
surveys in a pop-up modal that requested to be completed before continuing to the content. There
could also be an implementation strategy where the instructor required the students to complete
the surveys, using the data dashboards to hold students accountable. Maximizing student
participation is also a beneficial practice for investigating self-beliefs in any context to avoid
having a percentage of the sample with unknown self-beliefs as a limitation in analysis.

My last suggestion for future practice is to consider changes to how student self-efficacy
is collected. First, I would re-implement the self-efficacy survey items with a different order of
answer options, from disagree to agree, to see if this change impacts the answers students select.
It may also be beneficial to try a task-specific self-efficacy survey instead of a general self-
efficacy instrument. The general self-efficacy may be too general to be meaningful in this
specific context, and it would be worthwhile to explore this change in practice and as well as
future research.

**Recommendations for Future Research**

A first step in future research would be to repeat this analysis working with an instructor
and a single large class. By coordinating with an instructor, we could ensure all students
complete the surveys and summative assessments. This also has the additional benefit of having
a controlled and consistent implementation approach. The goal of this research project would be
to identify if the significant mindset results on engagement metrics could be replicated in another
analysis where all students are accounted for. Replicating significant mindset results would
provide additional confirmation to move forward with potential intervention research.
As noted in the literature review, there are studies which reported positive outcomes shifting mindset for at-risk student populations (Paunesku et al., 2015). While the platform does not collect demographic information, partnering with instructors to combine data could investigate if the mindset variable is particularly significant for specific student populations. Paunesku et al. (2015) found intervention treatment effective for at-risk students, but not among other students, so a similar analysis using courseware could show new insights. This research study does not include any student characteristics in the analysis, but it would be a worthwhile project to investigate if self-beliefs are meaningful for at-risk student populations.

Should additional research of mindset and engagement in courseware show similar results as this study, the identification of these significant differences could then be acted upon through intervention research. The literature review described mindset studies which have provided intervention and shifted student mindsets from fixed to growth. Given that the results of this research showed a relationship between mindset and engagement metrics, it would be a valuable research project to provide a mindset intervention and identify if students increase engagement. The goal of this research would be to investigate if shifting mindsets could increase student engagement, which would in turn benefit summative assessment scores. The ultimate benefit to students would be to use what is known about mindset and engagement to help students shift their mindsets and improve their learning.

Conclusion

Educational technology is rapidly evolving as digital learning increases and new methods of supporting learners at scale are being developed. As a professional in this field, my experience as a learning engineer framed my approach to educational technology to consider how I could apply research from the learning sciences to educational technology to better support learners.
Specifically, I began to wonder if the literature on mindset and self-efficacy—and the positive associations they claimed—could be identified in student use of courseware. As a scholarly practitioner, the first step in this journey was to engage in a research inquiry.

Researchers have found that whether a student has a growth or fixed mindset can influence engagement and learning outcomes (Blackwell et al., 2007; Paunesku et al., 2015; Yeager et al., 2019; King & Trinidad, 2021). Researchers have found similar links between self-efficacy and student characteristics and outcomes (Chen et al., 2001; Yeo & Neal, 2006). Using a historical data set of students who used a Probability and Statistics courseware at three different institutions, I analyzed student responses to survey instruments on mindset and self-efficacy in relation to learning metric data to identify if these same findings could be identified in the student engagement and learning data gathered by a courseware platform.

The most surprising findings of this study were on self-efficacy: 98.6% of all students in this analysis identified as having high self-efficacy. In this study, self-efficacy proved to not have a significant positive relationship with any of the outcome variables—from the number of formative questions students did to the scores on their summative assessments. This is contrary to the aforementioned research findings, however, it is in line with results from CCCSE (2019). It could be concluded that current college students do not suffer from a lack of general self-efficacy, and instead self-efficacy should be evaluated using a task-specific instrument.

Mindset similarly had results contrary to research that found positive relationships to outcomes. Roughly 80% of students reported having a growth mindset. Mindset was not a significant variable for mean summative scores. While contrary to the previously noted positive outcomes, this is consistent with several recent studies in which mindset had null results on learning outcomes (Burnette et al., 2016; Li & Bates, 2019; McCabe et al., 2020).
The most promising result of this analysis was the relationship between mindset and engagement metrics. Growth mindset was positively significant for the number of sessions students had and the number and accuracy of formative questions students did. King and Trinidad (2021) recently found a positive relationship between student mindset and perceived engagement, as did Tseng et al. (2020).

When we consider what these mindset results suggest in the larger body of research, the findings may not be a clear signal to take further action—in the case of the mindset variable, intervention. Yet when I contextualize these results within the learning technology from which the data came, there is more complexity to acknowledge. Unlike in an in-person learning environment, courseware produces data for every click a student makes and all practice and assessment responses are available for analysis. This type of data and learning analytics revealed the value of the learn by doing method for learning, which the regression model for summative assessment score in this study also confirmed. Student engagement with formative practice is critical to achieve the doer effect benefits, and therefore, additional research is warranted to identify if mindset could become a useful predictor for engagement using data from other courseware instances. Mindset intervention research has similarly produced mixed results of success (so I mention this with an abundance of caution), but if online intervention could shift student mindsets, that could potentially increase engagement activity, which in turn could increase summative assessment scores. While an optimistic hypothetical chain of events, the advances in learning science and in learning technology warrant a dose of hope.
References


Center for Community College Student Engagement. (2019). A mind at work: Maximizing the relationship between mindset and student success. Austin, TX: The University of Texas
at Austin, College of Education, Department of Educational Leadership and Policy, Program in Higher Education Leadership. Retrieved from:

https://doi.org/10.1177/109442810141004

https://doi.org/10.1111/jcal.12389


Olsen, J., & Shackelford, S. (2021). Intersectionality and Incremental Value: What combination(s) of student attributes lead to the most effective adaptations of the learning


Appendices

Data Approval Letter

6/22/2020

Approval of Acrobatiq Data Usage

This letter is to provide approval for Rachel Van Campenhout, employed as the Learning Science Specialist at Acrobatiq by VitalSource, to use data provided by Acrobatiq for her dissertation at Duquesne University.

The data set being provided is from courseware developed and owned by Acrobatiq by VitalSource. It will include data gathered by the Acrobatiq platform and stored by Acrobatiq—including login events, page visits, formative attempts, adaptive attempts, summative attempts, etc. These data points are associated with a unique numeric identifier generated by the platform. No student names, emails, or demographic information of any kind is included in the data set. The data set will include an identifier for the institution, but the name of the institution should remain anonymous in the final dissertation.

| Benny G. Johnson, Ph.D. |
| Director, Research and Development |
| Acrobatiq by VitalSource |
| benny.johnson@vitalsource.com |

PERTS Growth Mindset Items

<table>
<thead>
<tr>
<th></th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Somewhat disagree</th>
<th>Somewhat agree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>You can learn new things but you can’t really change your basic intelligence/how smart you are.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Your intelligence is something about you</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
that you can’t change very much.

3 You have a certain amount of intelligence, and you really can’t do much to change it.

### New General Self-Efficacy Scale Items

<table>
<thead>
<tr>
<th>Item</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Somewhat disagree</th>
<th>Somewhat agree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 I will be able to achieve most of the goals I have set for myself.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 When facing difficult tasks, I am certain I can obtain outcomes that are important to me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 In general, I think I can obtain outcomes that are important to me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 I believe I can succeed at most any endeavor to which I set my mind.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 I will be able to successfully overcome many challenges.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 I am confident that I can perform effectively on many different tasks.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Compared to other people, I can do most tasks very well.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 Even when things are tough, I perform quite well.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>