March 2022

UTILIZATION AND EFFECT OF MULTIPLE CONTENT MODALITIES IN ONLINE HIGHER EDUCATION: SHIFTING TRAJECTORIES TOWARD SUCCESS THROUGH UNIVERSAL DESIGN FOR LEARNING

Catherine A. Manly
University of Massachusetts Amherst

UTILIZATION AND EFFECT OF MULTIPLE CONTENT MODALITIES IN ONLINE HIGHER EDUCATION: SHIFTING TRAJECTORIES TOWARD SUCCESS THROUGH UNIVERSAL DESIGN FOR LEARNING

A Dissertation Presented

by

CATHERINE A. MANLY

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2022

College of Education

UTILIZATION AND EFFECT OF MULTIPLE CONTENT MODALITIES IN ONLINE HIGHER EDUCATION: SHIFTING TRAJECTORIES TOWARD SUCCESS THROUGH UNIVERSAL DESIGN FOR LEARNING

A Dissertation Presented

by

CATHERINE A. MANLY

Approved as to style and content by:

____
Ryan S. Wells, Chair

____
David D. Jensen, Member

____
Ezekiel W. Kimball, Member

____
Ezekiel W. Kimball
Associate Dean of Academic Affairs
College of Education

DEDICATION

I dedicate this work to my father, Collis, who inspired me to believe the doors to knowledge and understanding are always open to those who seek them. By simultaneously bringing science to my kindergarten class and allowing me to help him set up college level anatomy and physiology labs, he helped me see myself from a young age as someone who could understand and participate in the intersection of science and education. Dad, you encouraged me on my doctoral journey and knew I would get here, and I celebrate this moment with your spirit.

I also dedicate this work to my former college roommate, Michelle, whose memory inspires me every day to work toward making positive improvements for college students who think and exist in the world very differently from the way I do. I hope my work contributes to supporting all students, including those who face significant challenges during college, mental or otherwise.

ACKNOWLEDGMENTS

Working on this dissertation has had a profound impact on my life and during this journey, I have benefitted from the support of numerous people. There are way too many who have aided me on my path to completion to name everyone here and I apologize to those whose names I will miss mentioning. If you played a supporting role for me over my time in this doctoral program, as a teacher, mentor, friend, companion student, or member of my scholarly community know that I extend my gratitude to you! I would not be at this point without all of you. I particularly want to thank those of you who have been writing buddies over the years, including (but not limited to!) Kelly, Kris, Aimee, Genie, Ben, Tiffany, Rudi, and all those who participated in the AERA Division J Graduate Writing Group Zooms once the pandemic hit. It's been wonderful to get to know you all.

I also thank Lynne for her help learning to think about statistics. This dissertation would not be what it is without your patient and thorough explanations early on in my doctoral program as I was curiously learning about conducting research. I likewise thank Becky, Jenn, Amy G, Amy A, Brita, and Helen, without whom this dissertation would never have lifted off the ground or kept flying. Additionally, I thank Ellen, Deb, John, and Becky for listening and asking insightful questions as I worked through the thorniest aspects of working with these data.

I want to especially thank Maura for suggesting that I consider using data from The American Women's College at Bay Path University and for shepherding me through the process of getting what I needed. I also thank Amanda for her enthusiasm and support at every step, Jeremy for making the modality data possible to collect on relatively short notice, and Rich for wrangling the data complexities over and over again (including troubleshooting problems that various companies thought weren't technically possible!). You have given me an amazing opportunity to gather interesting data and design interesting research. (At least I think so!)

I also want to acknowledge the recognition of my work by the Northeastern Educational Research Association who awarded me the GSIC Best Paper by a Graduate Student Award for a preliminary version of the study presented in chapter two.

I extend deep gratitude to my committee members, without whom this study would not have happened. David accepted the challenge of teaching me about quasi-experimental thinking and design with laughter and insight. I am humbled and grateful to have been the beneficiary of a truly masterful teacher and mentor. You have a knack for asking just the right probing questions–guiding but not leading too much and prompting thought about the next issue at just the opportune time, planting seeds that will bear fruit later at the right moment. I have learned both from your approach and your teaching. Zeke guided me toward becoming knowledgeable about issues involving students with disabilities which crystalized my knowledge about UDL. You also supported my taking this work to a practical level through our UMass grant to provide a workshop series and conduct related research. Getting a glimpse into the positive impact that spreading the ideas of UDL has made for the faculty and staff participants has fueled my passion for this dissertation, reinforcing the importance of this kind of work. Ryan, my dissertation chair, has been with me at every turn, both positive and negative. Your ongoing support has been critical in getting me to the point of finishing. Through our collaborative research and publications, you have taught me what it means to be a scholar and what the process of research is all about. I hope this dissertation reflects the best of that! My life has been shaped in positive ways because of knowing you. I remain grateful that we have both learned deeply and profoundly from each other over the years.

David, Zeke, and Ryan, my success with this dissertation rests on your expert guidance. Any muddled thinking or execution is mine.

Perhaps most importantly, I thank my family, who have supported me in innumerable ways over many years. I particularly acknowledge the support and sacrifices of my husband, John, without whom this whole endeavor would not have been possible. You have been there in ways small and large, day in and day out for me, for our kids, Miller and Nathan, and for my parents. In particular, you have been there for my mother, Carolyn, after my father passed. You have adjusted and readjusted and readjusted our lives again and again as we have worked out what was needed for me to finish this dissertation. I remain deeply grateful and thankful to be sharing our lives together.

I offer a final bit of gratitude to the heron on the campus pond who glides gracefully over the water and patiently fishes for lunch every day–thank you for centering me in the present during my final writing efforts as we all emerged from the challenges of the COVID-19 pandemic.

ABSTRACT

UTILIZATION AND EFFECT OF MULTIPLE CONTENT MODALITIES IN ONLINE HIGHER EDUCATION: SHIFTING TRAJECTORIES TOWARD SUCCESS THROUGH UNIVERSAL DESIGN FOR LEARNING

FEBRUARY 2022

CATHERINE A. MANLY, B.A., AMHERST COLLEGE

M.B.A., UNIVERSITY OF PHOENIX

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Ryan S. Wells

The idea that offering multiple means of representing course content will assist students of all abilities constitutes one pillar of Universal Design for Learning (UDL), a framework intended to address needs of students with disabilities while also holding relevance for all students. The efficacy of this UDL guideline lacks a verified empirical basis and therefore merits rigorous examination. My dissertation investigates the effect on learning outcomes of students using multiple modalities while learning course content (e.g., text, video, audio, interactive, or mixed content), targeting improving educational success for non-traditional online students.

I investigate this effect for older undergraduates from a women's institution who are predominantly low income and working mothers returning to school, many of whom are racial/ethnic minorities. Notably, challenges resulting from a lack of disability diagnosis and accommodation may be prevalent but hidden among these students.

Traditional higher education typically does not serve such students well. Use of multiple modalities in class activities holds potential for improving their outcomes.

Results show positive effects of using multiple modalities for learning content in courses across the curriculum presented in an adaptive learning system. Using a within-subjects study design, I found a medium-large positive effect size for knowledge gained across adaptive activities. Using an instrumental variables approach, I found a very large positive effect size for weekly assignment and quiz grades, and results suggest a large positive effect on course grade as well. I illustrate how combining knowledge of this effect with other information from the adaptive learning system and online tutoring in a Bayesian network analysis can predict where students may benefit from tutoring. This can inform potential support recommendations that would be particularly relevant when implementation of UDL-based design does not yet fully address students' learning needs.

These results provide the first evidence confirming an effect of UDL's multiple modalities guideline on collegiate learning outcomes and illustrate how this information could be used to provide recommendations to students using a learning analytics perspective. Results have implications for researchers, faculty, course developers, instructional designers, analytics professionals, and institutions aiming to improve learning outcomes through a design-based approach.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

> As educators, administrators, and service providers, we must continually unpack and reflect on barriers to educational access so that effective and socially just change can take place in our institutions. To act in the interests of social justice, we must be willing to collaborate with all the essential personnel in a higher education setting to enact positive and long-lasting change that will benefit all students. Until we make education accessible to all historically underrepresented groups, we will not be able to engage in a pedagogy that is truly inclusive of us all. (Pliner & Johnson, 2004, p. 111)

United States higher education offers many benefits for those who fit the expected student mold, but its meritocratic ethos hides systemic inequality for students who do not match the mainstream. Historically, that mainstream has included students of traditional age attending college full time who do not have significant physical or mental impairment. Unfortunately, too often the learning requirements of students with disabilities are not anticipated by or adequately supported by the expectations and processes that have been codified in course design.

By rigorously investigating the connection between course learning outcomes and one key aspect of the higher education course environment–the representation of content–this dissertation offers new insight into a challenging perennial educational question: how do we enable students to learn successfully? Informed by the Universal Design for Learning (UDL) framework that addresses learner variability by intentional design, I aim to improve understanding of this aspect of the complex and perplexing problem of enabling learning success for students both with and without disabilities, thereby informing student support efforts and future course design.

My dissertation focuses on students at a women's institution that welcomes many traditionally underserved students, particularly nontraditional-age undergraduates juggling work and families, most of whom are low income, many of whom have non-dominant racial and ethnic identities, and all of whom attend college part-time online. Since many are older students returning to complete their undergraduate study, they may not have current disability diagnosis information or may have never been diagnosed because disability awareness was less strong during their earlier schooling. Given this, my expectation is that some of these students have disabilities as well, though as will be discussed later, their number is difficult to discern.

UDL, the framing which orients my dissertation, views all individuals as capable learners given an environment that supports rather than disables their capabilities (Burgstahler & Cory, 2008). UDL's empowering approach to course design, promoted by many who study disability and education, has been widely recognized as having applicability to all students (Rose & Meyer, 2002). As a design-oriented framework, UDL encourages faculty and course developers to adopt attitudes, methods, and materials accessible to the full range of students (Fovet & Mole, 2013). It inherently supports those who might otherwise exist on the margins of educational practice, intentionally emphasizing design for learner variability among all students (Tobin & Behling, 2018). Sometimes, however, the learning environment's design intentionally or unintentionally "weeds out" students who fail to conform to expectations about how to achieve and demonstrate content mastery (Bettencourt et al., 2018; Etzkowitz et al., 2000). Epistemologically, UDL asserts each student has educational potential that faculty and institutions remain morally bound to foster. Understanding what practices support the goal of providing effective learning experiences thus becomes an imperative step toward equity.

While universal design pertains to on-campus as well as online learning, it holds notable importance in online educational settings where the learning environment substantially shapes students' experiences (Dell et al., 2015; Kelly & Zakrajsek, 2020; Rao et al., 2015). Importantly for this dissertation, the online setting facilitates data collection about student learning activity throughout a course. Despite the importance of understanding the need for and practice of accessible online design (Burgstahler, 2006; Quality Matters, 2020), many faculty have not been trained accordingly (Gladhart, 2010). Additionally, the COVID-19 pandemic has further highlighted the need for educators to employ practices relevant in technologically mediated learning environments that flexibly meet learners where they are in life as well as in their learning capabilities (Hodges et al., 2020; Levey et al., 2021). Particularly in crisis situations, but also in more normal times, the learning environment that students encounter may present bottlenecks to successfully achieving their educational goals (Fishkin, 2014). To progress, students must pass through these bottlenecks, but their progress along a given educational path may be delayed or derailed due to nonconformity to environmental norms not constructed to meet their needs.

Countering such systemic challenges, UDL offers an approach intended to support students who have impaired physical, psychological, perceptual, or processing abilities, for example, while also benefitting learners across the full range of those abilities, thus helping all students (Rose & Meyer, 2002). UDL's flexible approach holds potential to benefit everyone associated with the learning environment: students in the process of learning as well as instructors supporting students with diverse learning requirements.

This dissertation informs instructional design efforts to institute UDL and suggests intersections with student support. Specifically, I study the modality of course material presentation, such as text or video, a key perceptually based component of content representation. After establishing the effect on students from the practice of including multiple modalities for content, I offer an example of how knowledge of this impact might inform targeting of future tutoring recommendations by illuminating where students struggle. Given the practical realities that often limit UDL's full implementation in course design and development, such focused research is needed to indicate areas to emphasize in faculty training and development, as well as provide guidance for supporting students when efforts to date do not meet the needs of all learners. To this end, my research illustrates the type of predictive analysis targeting tutoring interventions that is possible when learning activity data are collected on an ongoing basis throughout a course, as can be facilitated by online adaptive learning platforms. The individualized design enabled by such adaptive platforms has similarities to the flexibility and options advocated for by UDL, though the computer science and education communities advocating each have typically not communicated closely with each other (Seale, 2013). In addition to identifying where particular students may benefit from tutoring in a course, such analysis has potential to inform course revision by identifying places in the sequence of course activities where patterns of struggle are identified or observed. Such work has the potential to benefit students who have non-traditional characteristics, along with all other students.

1.1 Statement of the Problem

While it seems reasonable to expect that designing higher education learning experiences with the intention of being universally accessible to students of all abilities would lead to improved student success, well-designed empirical research corroborating this intuition is surprisingly sparse. The Universal Design for Learning (UDL) framework challenges higher education institutions to design students' learning experiences to include multiple means of engagement and representation, as well as action and expression. Of these areas, engaging students in multiple ways in college has been shown to have great value in general, including both curricular and co-curricular aspects (Kuh et al., 2007). In contrast, comparatively little research has been done on providing multiple means of representation relating to perceptual cognitive processing, particularly for curricular content beyond multimedia (Mayer, 2001). Consequently, the connection between college student course outcomes and presenting course content through multiple modalities is not yet well understood.

Approaching course design from a UDL perspective is particularly needed in online settings which can otherwise see increases in accommodation requests for students with disabilities (Barnard-Brak et al., 2012).[1] Technological advances make it increasingly straightforward to gather data about student interactions with course material, particularly in online courses where most student activity and interactions leave recordable and analyzable traces. Such automatically collected electronic data make

[1] I respectfully acknowledge differing opinions about the relative merits of person-first or identity-first language when discussing people who have disabilities (Association on Higher Education and Disability, 2021). This dissertation uses person-first language as a group signifier since I do not discuss specific individuals who identify as having a disability.

investigating the connection between content representation and student success more tractable than before. However, complexities may exist such as subject or student-content-modality dependencies. Given that so little is yet known in this area, investigating the existence of effects without considering such interactions would still be worthwhile. Given the potential for improving success outcomes for traditionally underserved students from a range of abilities and prior backgrounds, it is imperative to investigate teaching strategies in increasingly rigorous and holistic ways. Doing so should enable focusing instructional support changes on places where actions are informed by predictions of expected improvement.

1.2 Universal Design for Learning

To offer context for this dissertation's UDL focus, the concept of universal design originated from the field of architecture, encouraging design of accessible environments and de-stigmatization of disability (Hamraie, 2017). In this architecturally based framing, students with differing abilities were seen as disabled by an inaccessibly designed environment and not because they are inherently disabled as people. This framing follows a social constructionist understanding of disability that places people in interaction with one another and their surroundings rather than following the older medical model of disability that relies on medical diagnosis oriented around negatively viewed problems rather than positively viewed capabilities (Kimball et al., 2016). Universal design's seven principles include: (1) equitable use, (2) flexibility in use, (3) simple and intuitive use, (4) perceptible information, (5) tolerance for error, (6) low physical effort, and (7) size and space for approach and use (Burgstahler, 2015).

These principles have been applied to educational settings in numerous ways that are slight variants on each other. Universal Design in Education (UDE; Bowe, 2000) and Universal Design for Instruction (UDI) address diverse learning needs by applying the seven principles of universal design to higher education. Two UDI variants exist that were developed independently using the same name. One developed by the Center on Postsecondary Education and Disability (CPED) at the University of Connecticut added two principles to the seven universal design principles of (8) a community of learners and (9) instructional climate (McGuire et al., 2003). The other was developed by the Disabilities, Opportunities, Internetworking, and Technology Program (DO-IT) at the University of Washington (Burgstahler & Cory, 2008), though they have later focused more on UDL. Yet another framing, Universal Instructional Design (UID), brought to higher education from K-12 by University of Massachusetts Amherst researchers (Silver et al., 1998), additionally incorporates Chickering and Gamson's (1987) *Seven Principles for Good Practice in Undergraduate Education* (Higbee & Goff, 2008).

The UDL framework developed by the Center for Applied Special Technologies (CAST, 2011; Meyer et al., 2014) takes a slightly different tack on universal design compared to these other approaches. It's three components posit that students benefit from having multiple means of engagement, multiple means of representation, and multiple means of action and expression in order to achieve their optimal learning potential. For some students with functional impairments, such multiple means can enable an accessible learning experience.

While my studies focused specifically on women due to data availability at a women-only institution, UDL principles are expected to be applicable to all genders given that the brain research underlying UDL is not gender specific (Rose & Meyer, 2002). While there has not been much prior research on UDL and gender (for examples, see Couillard & Higbee, 2018; Glass, 2013), gender could be interrogated further in the future, such as by extending the present work to confirm applicability with male students.

Education scholars face a gap in understanding the specific, quantifiable benefits of UDL's elements as applied in postsecondary education. Greater clarity about the effectiveness of the recommended practices could assist in prioritizing efforts for those implementing professional development and course (re)design processes. Since the desirability for universal design and web accessibility for students with disabilities has been recognized by the educational technology community for several decades (Bohman, 2000; Miller, 1999; Nielsen, 2000), it may be surprising that more in-depth analysis of the implementation of components of UDL in higher education has not occurred. UDL research has often focused on kindergarten through grade 12 (K-12; Mangiatordi & Serenelli, 2013), mirroring the generally more extensive K-12 scholarship of teaching and learning (SoTL) research base (Swanson, 1999). However, research about UDL in higher education has increased over time as well (Cumming & Rose, 2021). While guidance for faculty about course development exists (e.g., Chickering & Gamson, 1987; Dell et al., 2015; Ko & Rossen, 2017; McKeachie & Svinicki, 2010), more attention to the elaboration of universal design practices based in extensive, high-quality, empirically-based research remains needed.

To foster research connecting universal design to student success, appropriate data need to be collected in a useful manner. Technology holds out the promise of collecting increasing amounts of data in multiple institutional spheres (e.g., with learning management system (LMS), advising, tutoring, disability services, technology help desk, and card swipe/location data). However, in practice, if this information is collected electronically at all, it is often housed in disparate systems. Some institutions have been working on systems to improve student success through electronic data aggregation for predictive analytics, including risk analyses, early alert systems, and dashboards to present data visually to students and other stakeholders, for example (Arnold & Pistilli, 2012; de Freitas et al., 2015; Sin & Muthu, 2015), but such dashboards still frequently focus on macro-level issues such as course completion rather than within-course progress (Bodily & Verbert, 2017; Jivet et al., 2018). While educational technology has permeated certain aspects of the learning environment (e.g. the now ubiquitous LMS; Dahlstrom et al., 2014), depending upon the design of such systems and what traces are collected of student activity and learning progress, the form of the data may not be conducive to constructive analysis of the effects of student activity and choices. In general, though efforts targeting improved within-course learning, course success, and reduced time-to-degree are growing, analysis of data from such systems has not typically led to course development revisions (Ferguson et al., 2016). The research and practice landscape continues to evolve as institutions begin to aggregate student-related data in data warehouses, and as traces of learning-related data become more comprehensive (Papamitsiou & Economides, 2014; You, 2016). Within the growing field of learning analytics which makes use of such data (Zilvinskis et al., 2017), this increasingly enables learner analytics where predictions inform interventions about actions a student might take to improve their academic performance (Pistilli, 2017). Collecting such data about the student experience of learning content can enable research about the relevance of

UDL's expectations around content representation, extending past research on other aspects of UDL theory. Next, I discuss how the UDL components of providing multiple means of engagement, representation, action and expression relate to my dissertation.

1.2.1 Multiple Means of Engagement

Student engagement has been viewed as a key aspect of UDL theory. While this area initially was considered the third of the three UDL guidelines, CAST moved this item in 2014 to the first position of the three principles to emphasize its importance overall as well as its foundational role underlying the other two guidelines (Evans et al., 2017). Engagement has appeared to be similar across cognitive ability variation, as Hendrickson et al. (2015) found that engagement was similar for students with intellectual disabilities and typical students. Regarding representation, the focus of this research, students who were not engaged in their educational activities may not have had the opportunity to learn the content to be mastered, no matter how artfully or diversely represented, because they may not be present with it. Similarly, no matter what options were provided for setting learning strategies or communicating, disengaged students may not take advantage of them.

Engagement has been a powerful, important aim of UDL incorporating several components. Important to my studies, this foundational aspect of learning has allowed students to demonstrate their motivation to learn. It has included encouragement of self-beliefs and self-expectations that serve to drive motivation to peak levels for each individual. By providing options whereby students have optimal possibilities to realize autonomy and personal relevance in distraction-free learning experiences, individual interest may be sparked, motivating sustained effort. By including elements such as varied support for optimizing learning challenges, students have been encouraged to achieve a state of flow in their learning whereby learning experiences engage them in ways that motivate persistence.

1.2.2 Multiple Means of Representation

The part of UDL theory that I address encompasses providing options for perception. This falls under the UDL guideline of providing multiple means of representation. It specifically includes providing alternative formats for perceiving content. Given that individuals' perception and preferred processing mode will vary, some students may require or benefit from content provided in a visual format, while others may benefit from auditory content, for example. Such content could either be expressly created for a particular modality format, or it could be the closest analogue when translated from a different format. For example, a visual format could include a video, graphic, image, or plot. For a traditional lecture, content presented in a written format could be taken and translated to be offered in spoken form, perhaps with graphics or interactive visuals in a multimedia format intended to help points be learned and retained by students. An auditory format could include the audio track from a video as long as all visual elements that were important for understanding were rendered explicitly. Auditory content could also include a verbal description of a concept, which could be someone speaking or could be provided through text-to-speech software. Alternately, a textual format could include content written expressly to be read, a transcription of a video, or a textual description of the options in an interactive exercise. By offering students ways to customize what content they encounter, they can act with agency in determining what format or combination of formats works best for their learning. While material quality is essential, designers of learning systems can also build in observations of student learning that might be analyzed for signs that students are struggling to learn the material. Students who demonstrate difficulty understanding the material could receive a recommendation to try again to learn the material through a different modality. This is what the adaptive system utilized in my dissertation does. Alternately, such signs of struggle may be indicators of quality or design issues that could be addressed during the course and material revision process.

In addition to perception, other areas of representation in UDL include providing options for comprehending material, particularly through means that reinforce cognitive pathways other than those relied upon in traditional delivery modes like lecture. This includes providing alternate paths for students lacking expected background knowledge or helping students transfer what they learned in one context to a new learning context. It also includes highlighting important aspects of the material in potentially non-linear (i.e., non-lecture-style) ways, such as emphasizing patterns, global thinking, and relationships, particularly between features of the material that constitute a critical path for comprehension.

1.2.3 Multiple Means of Action and Expression

The action and expression tenet of UDL focuses on learning strategy and orienting actions toward learning goals. It encompasses executive functioning, which constitutes psychological elements that help individuals achieve what they set out to achieve and recognizes that what facilitates learning progress for different people can vary. It also includes ways of communicating and navigating physical space that could differ dramatically depending upon functional impairments of individuals. It posits that everyone's learning could be facilitated when interactions between individuals or between individuals and their learning environments are designed recognizing that options appropriate for a wide range of abilities may be necessary, enabling people with different disabilities to learn and demonstrate their learning successfully.

While other UDL-based research has looked at the effect of UDL from a holistic standpoint encompassing all three areas (Chandler et al., 2017), the present research separates the contribution of multiple means of representation from the other areas. Thus, these other UDL areas are assumed to be implemented similarly in these courses given the consistent course design process employed. Additionally, their presence or absence is assumed not to confound the estimation of the effect of interest, though such assumptions could be investigated in future UDL research more explicitly investigating possible interactions between UDL's elements.

The many facets of UDL mean that practical realities of academia frequently limit aspirations of implementing all components of UDL at once. Faculty need support managing time pressures and responsibilities that often feel overwhelming (Austin & Sorcinelli, 2013), and adding UDL to their professional development mix can present practical implementation challenges (Gladhart, 2010). Given that UDL has often been implemented in pieces and over time (Tobin & Behling, 2018), more ought to be learned about specific benefits of UDL's components to prioritize best practices for implementing UDL principles effectively as well as to support students when the promise of UDL has not yet been fully realized or is still incompletely successful.

1.3 Overview

This dissertation presents the results of three interconnected studies aimed at understanding how use of multiple modalities can be utilized to help students learn course content. This first chapter introduces the dissertation overall, describing the problem, conceptual framework, purpose, and logic behind the research approach; offering a few orienting definitions for some terms; and outlining the research questions and associated hypotheses. This chapter introduces the idea of Universal Design for Learning and establishes its importance for course design as a framework that holds potential to help students across the full spectrum of abilities, guiding development choices such as providing multiple options for perceiving content. It establishes that there is still much to be learned about the effectiveness of this principle for aiding student learning despite the face validity of the concept in general, providing an opening for the present research.

Chapters two, three, and four present the results of three studies delving into this opening, starting at a basic, formative level and iteratively expanding the view. Each of these three chapters is written as an independent article, including a review of relevant literature, description of the data and methods, results of the analyses, and discussion of the findings and their implications.

Each chapter highlights a slightly different facet of the UDL-related literature. Chapter two introduces the idea of universal design, focusing on research about modality representation in the higher education and disability literature, as well as the cognitive science literature, including brain imaging and multimedia research relevant to understanding possible benefits of using multiple modalities. Chapter three argues that the existing case for the effectiveness of multiple modalities is basically suggestive rather than concrete. Chapter four brings in perspectives from instructional design, learning analytics, and tutoring to complement ideas from the UDL literature. Taken together, this literature offers an understanding of how multiple modalities may work, what we currently know about this in the context of UDL, and how such information is being used in practice.

These three chapters investigate complementary outcomes utilizing slightly different data. Given the lack of research demonstrating an effect from using multiple modalities, I first aim to establish the existence of such an effect, initially for formative learning activities and then for summative grades at the end of each week and the course overall. Different analytical routes are used to identify causal connections as supported by available data. This means that while the samples in all three analyses come from the same institution, different courses from different semesters and different variables were included for each analysis. Separate samples were necessary to match the type of data available for collection to the design requirements for each method. Chapter two introduces the institutional context for the study overall, while chapters two through four describe the specific data used for each analysis.

Regarding methods, I start by investigating the effect of interest for formative learning activities in an adaptive learning system using a change score panel data analysis in chapter two (Morgan & Winship, 2015). I next identify the effect for weekly assignment and quiz grades using an instrumental variables analysis in chapter three (Morgan & Winship, 2015). I then extend this to look at summative effects on course grade using a similar instrumental variables analysis, also in chapter three. After establishing evidence of a medium-strong effect across learning activities, a very strong effect on weekly grades, and indication of a strong effect on course grade, I investigate how this knowledge might be used in practice to help struggling students from a prescriptive analytics perspective in chapter four. Here, I utilize a Bayesian network approach (Darwiche, 2009; Pearl, 2009b) in a proof-of-concept example analysis for an English course, bringing prior student data about modality use and tutoring to bear to identify points where analysis predicts individual students may benefit from tutoring given their past performance. I describe how such information combined from disparate technological systems could inform recommendations for tutoring assistance at key moments in a course.

In the final fifth chapter, I discuss the findings in an integrated way. I synthesize results from the modality effects analyses, drawing overall conclusions from the findings about effects across the studies and discussing how these results could inform analytics practices such as discussed in chapter four's analysis combining modality use and tutoring. Finally, I draw common implications from these analyses for theory validation, future research that could extend these results, analytical practice that could be informed by the types of data and methodological approach utilized, and institutional policy and practice that could be informed by the directions explored here.

1.4 Purpose of This Research

Identifying elements of the educational process occurring within the course context that lead underserved students to success with their assignments and the course has been historically challenging. This constitutes a central aspect of why we describe these students as underserved. It can be challenging to figure out where best to put resources for faculty and course development such that they will have the most substantial impact for struggling students.

Exploring one avenue for addressing this, the present research utilizes data traces collected as students progress through an online course to investigate the effectiveness of representing content to students through multiple modalities. This approach holds potential to identify impediments to learning which could be addressed either systemically through course design or individually through supplemental tutoring when systemic approaches fall short. Given the relative novelty of extensive learning data trace availability, few prior studies have empirically investigated the efficacy of providing multiple modalities for presenting content (e.g., Webb & Hoover, 2015) and none have used such course data traces to target improved content presentation accordingly. Providing content in multiple modalities is an aspect of offering options for perception which UDL theory posits will make a difference in student learning, especially given that students have a range of innate perceptual and processing abilities.

Depending upon how courses are designed, UDL theory guides us to understand that variation in perception and processing need not inherently result in students encountering disabling or ineffective learning environments. However, very few researchers (e.g., Hall et al., 2015) have studied the relationship between different content modalities and student course outcomes, even as part of larger research on UDL (Capp, 2017; Rao et al., 2014). Additionally, peer reviewed research studies focusing on investigation of UDL's multiple means of representation implemented in courses have limited statistical analysis beyond looking at descriptive statistics (e.g., Fidalgo &

Thormann, 2017). Such studies have not used detailed information to explore the role modality might play in improving course implementation.

Since existing research has not addressed the question of whether an effect of using multiple modalities exists or how important it might be, course design efforts applying this aspect of UDL remain in the realm of plausible but unverified. Research such as the present study is needed to investigate the effect on student outcomes from a causal perspective. This original research aims to inform student support and faculty development efforts with a specific focus around UDL implementation in order to target such efforts toward actions leading to student success by: 1) extending knowledge of the efficacy of offering multiple means of representation, and 2) demonstrating how student modality use activity data can be used in combination with other student data to support struggling students while simultaneously informing revision to course content delivery.

1.5 Research Questions

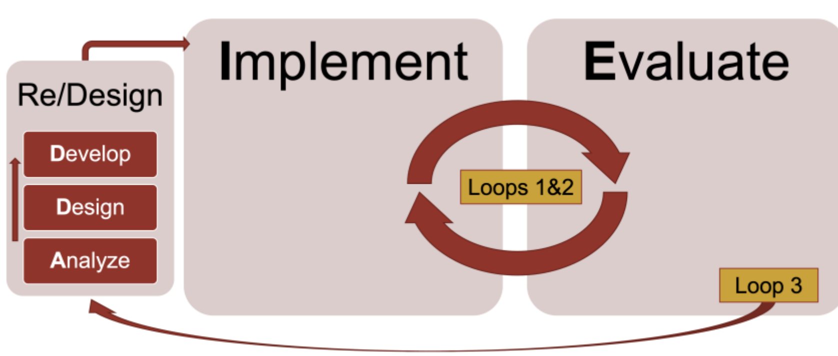
To investigate these issues, the following research questions guide my inquiry, which correspond to chapters two through four:

1. What are the effects of choosing more than one modality (either text, video, audio, interactive, or mixed) for learning course material on knowledge gain?
2. What are the effects of choosing more than one modality for learning course material throughout the week on subsequent weekly assignment grade outcomes (or throughout the course on final course grade)?
3. How can information about modality switches and tutoring be used to predict later learning module success in one week of an introductory English course?

1.6 Research Logic and Design

Overall, my research can be viewed as cycles of implementation-evaluation falling within the ADDIE instructional design model (Peterson, 2003). ADDIE separates the instructional design process into five categories: analyze, design, develop, implement, and evaluate, with both formative and summative evaluation occurring. To place the elements of my research within this larger design process, I organized the categories of ADDIE as a logic model that shows my studies in loops within a larger cycle (see Figure 1.1). This includes clarifying the resources needed for the research (i.e., Re/Design), the activities of the research (i.e., Implement), and the outputs expected (i.e., Evaluate; Cooksy et al., 2001). The outcomes which are goals of processing the feedback resulting from the research are described below. Organizing the internal components of my studies, this logical model positions my dissertation within the context of what has already been happening within the institution studied, illustrating how feedback from my results might be used within ongoing institutional improvement cycles. That is, my investigation occurred in the middle of a continuous institutional curricular development process that has already included designing the courses and system studied here. I use this logic model to situate my dissertation, showing how it contributes to knowledge and practice.

Figure 1.1 Re/Design-Implement-Evaluate Feedback Loops



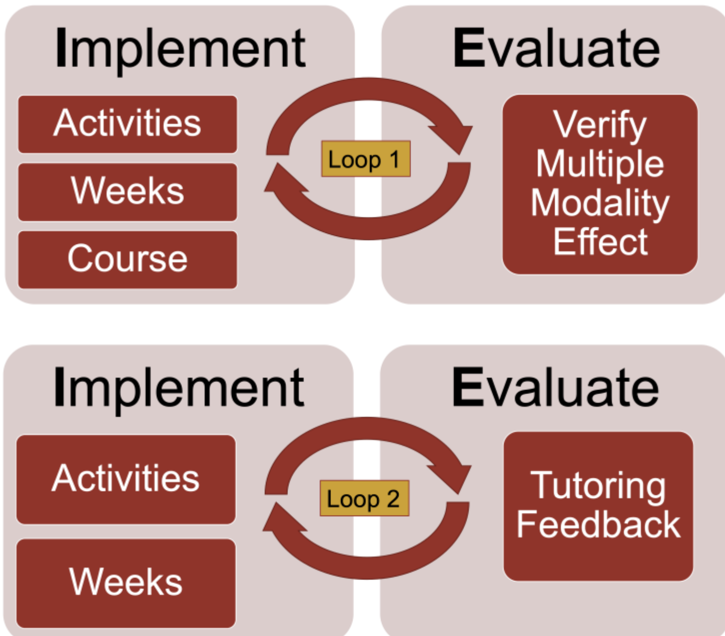
The inputs that are resources for my dissertation include the design process used by the institution to develop course content and the resources used for offering courses and collecting data. Using a structured team process funded by a U.S. Department of Education FIPSE First in the World Grant, a team of instructional designers and subject matter experts addressed the first three steps of ADDIE. This is the point at which the availability of multiple modalities was designed into the course content. In the context of my dissertation, these initial design steps have become inputs for my research. They have also become part of an ongoing redesign process that the institution plans to undertake periodically for all courses.

The activities and outputs of my research focus on two implementation-evaluation loops (in chapters two through four), with the intent of informing a longer-term third loop back to the redesign process. The implications of my research for this third loop will be discussed in chapter five. The goals of the three feedback loops are:

- Goal for loop 1: Knowledge about effect of using multiple modalities (addressed in chapters two and three)
- Goal for loop 2: Predict where to provide students guidance for tutoring (illustrated in chapter four)
- Goal for loop 3: Identify prime target areas for course redesign (discussed in chapter five)

The first feedback loop (see Figure 1.2) includes three cycles based on data from the courses being implemented, looking at the activity, week, and course levels to verify the existence of an effect of using multiple modalities at each level. If such an effect is not found to be present, then further research based on the existence of such an effect becomes unneeded. Thus, results from this first loop justify providing feedback to students and instructional designers based on information about use of multiple modalities in loops two and three.

Figure 1.2 Zoom in on Implement-Evaluate Feedback Loops 1 and 2



Assuming a beneficial effect is found, the second feedback loop (see Figure 1.2) again utilizes the data from the courses being taught, this time to develop a predictive model using information about multiple modalities combined with information about utilization of tutoring to provide the basis for predictions. This proof-of-concept analysis was intended to show how feedback could be offered to students throughout a course regarding when it might be helpful to seek tutoring assistance given their actions and performance in the course. Such prescriptive analytics results can provide a basis for the instructional design feedback in loop three. The third, longer-term feedback loop is not analyzed here, but could be pursued with institutional collaborators in the future. It is included to show the larger context within which the present dissertation's elements are situated, and to orient discussion of future implications.

The analyses in chapters two and three seek high internal validity and are designed to identify the extent to which a causal connection exists between use of multiple modalities for learning content and actual student learning, as represented by student performance on formative and summative assessments (Shadish et al., 2002). The analysis in chapter four expands on the causal model developed earlier and uses it for predictive purposes.

The quasi-experimental approaches used focus on identifying the effect at one institution. I do not directly seek to establish external validity. While some findings may be applicable to other settings, identifying such transferability other than by plausible argument is not part of the research design.

Considering construct validity (Shadish et al., 2002), I also do not directly address the question of whether the concept of multiple modality use investigated here, which should be highly internally consistent, is consistent with research in other settings. This is reasonable given how little research there has been in this area and the deliberate lack of standardization among researchers around operationalizing UDL constructs across settings in order to provide appropriate flexibility in implementation (Smith et al., 2019).

Attention is paid throughout, however, to establishing the validity of statistical conclusions. This is done by attempting to identify possible threats and describing potential issues. Associated topics addressed include attrition, baseline equivalence, appropriate modeling, power, and what would be required to alter conclusions drawn. While additional research would be warranted to explore different aspects of validity (particularly external validity), the multi-faceted approach to establishing validity taken here increases confidence in the findings overall and increases understanding of their appropriate interpretation and application.

A very strong quasi-experimental research approach involves a within-participants design. Such causally oriented investigation represents a particularly important direction for higher education research (Schneider et al., 2007). This approach uses data that are structured by panels of individuals across time where treatment occurred in some cases and not in others. In this type of research design, an individual (i.e., student) essentially acts as her own control, which means that student-level factors should not confound the results. While non-individual factors may still cause trouble with the estimate, in an educational situation, individual factors are likely to be a primary source of potential confounds, so being able to eliminate this source of bias is very desirable. Data for this type of analysis were available for the activity level outcome, so a change score panel data analysis was conducted as described in chapter two with data from the 2018-2019 academic year. By comparing the (first) difference in knowledge state before and after working to learn the material covered in the activity, the effect that using multiple modalities has on the gain in student understanding of the material could be estimated.

From a design standpoint, an optimal way to investigate a causal effect is to isolate the variation of interest so that it is not mixed with variation from other sources, either from other direct causes or by association due to common causes. Randomized experiments are intended to achieve this isolation by design since any source of variation other than the treatment should vary randomly across the sample assuming a good randomization process and will have a detectable effect given a sufficiently large sample. While data from a randomized experiment happen to be available (and are used in chapter three), the randomization procedure did not randomize the treatment of interest here. Instead, it randomized the *availability* of the treatment of interest. This situation can be described as an intent-to-treat instrumental variables design and constitutes a good candidate for an instrumental variables analysis since the instrument is known to be random (i.e., exogenous or external). The strength of this design for determining causal effects meant that I chose to conduct this type of analysis for both the week and course level outcomes analyzed in chapter three. The amount of data was sufficient but not large, so covariates were included to increase power and supplemental analyses were undertaken to verify the results' stability and strength.

In a proof-of-concept application using a learning analytics approach in chapter four, I combined modality use and tutoring information (students at this institution are offered a limited amount of free online tutoring) from a high enrollment English course in an example of the type of Bayesian network analysis that can indicate where students showed greatest signs of struggling with the content. This approach could be used to indicate places in the course where it is unfortunately not currently clear how to make the content work smoothly for all students from a standalone universal design standpoint (or else sufficient resources have not yet been applied to this end) and thus where individualized tutoring is predicted to benefit students. This holds potential to inform tutoring recommendations for particular students.

In investigating the use of multiple modalities, it is important to distinguish between possible alternative explanations for any observed effect. This serves to clarify the contribution of my work by situating it alongside other potential effects. The explanation that using different sense and processing capabilities may matter for learning concepts is the primary effect investigated by this dissertation. Use of multiple modalities allows some individuals to reinforce concepts via different mental pathways while enabling essential access for students with disabilities for whom certain functional capabilities are constrained or blocked. Within this explanation, there might be different mechanisms at work, but investigating such mechanisms is not the focus here.

While the issue of alternative explanations for an effect is explored further in chapter two, examples include the difficulty of the material, the quality of content, the quality of the UDL implementation, the quality of the online learning environment, the affective state of the student, the concentration level of the student, the executive function abilities of the student, identity characteristics of the student, feelings of self-efficacy around this type of material (which might be particularly salient for certain subjects like math), better performance under conditions of choice, worse performance under conditions of too much choice, variability in personal preferences for studying, students having a “best” modality (or a “worst” modality), content having a “best” modality (or a “worst” modality), spending more total time on task, repetition of the material, differences in memory storage and retrieval, or differences between the efficacy of short-term and long-term memory. Some of these possible explanations seem implausible given other evidence (such as best or worst alternatives), and others are not conducive to inquiry given the data available (such as memory functioning). Some (such as difficulty) were averaged over, while others (such as quality) were held to a minimum baseline by the institution’s design team which strove to achieve constancy in course design. Yet others (such as affective state) might be amenable to investigation through capturing additional data traces. A few would not be difficult to investigate and deserve further inquiry in the future (such as time on task or repetition) but are beyond the delimitations of this study. Further consideration of such alternative explanations, as in the Discussion and Implications section of chapter two, is an important part of developing certainty in claiming to find a causal effect (Shadish et al., 2002).

Designing my studies aiming to identify causal relationships responds to the UDL literature's multiple calls for better research designs. McGuire (2014) offers recommendations to build a systematic research agenda and "promote research that addresses the paucity of empirical evidence regarding the efficacy of these frameworks for promoting inclusion and learning" (p. 394). Capp (2017) calls for more experimental designs and curricular area studies to investigate learning outcomes. In a review of K-12 intervention research spanning 30 years, from when CAST was formed in 1984 through 2014, it is particularly striking that only five articles focused on causal effects through group comparison studies and single-subject designs. These authors advise:

> Last, researchers of future investigations should consider measuring isolated aspects of the UDL framework. In other words, for every guideline or checkpoint embedded into the design of interventions, researchers should also consider assessing the possible effects (e.g., disentangling embedded technological features used/not used). (Crevecoeur et al., 2014)

Additionally, McGuire, Scott and Shaw (2006) conclude that further rigorous research about UDL's aspects is necessary "to allow this potentially powerful model to be developed and proven *before* [emphasis in original] it is widely–and possibly ineffectively–implemented" (pp. 173-174). The research designs employed in the present dissertation aim to meet a high standard for rigor and begin filling this gap in the higher education UDL literature.

1.7 Definitions

To clarify description of my research, I offer definitions of the following terms, which are specific to understanding the technological and academic systems used here.

- **Activity**. A component of a course's weekly module expected to take a student approximately 20 minutes to complete.
- **Adaptive system**. The online platform developed by the institution in conjunction with an outside vendor (e.g., RealizeIT) that houses the delivery system for course content and checks of student knowledge. The system can include content presented in different modalities for each course activity.
- **Determined knowledge state**. The knowledge state of a student for a particular module as determined by a recommended assessment at the beginning of the week. This forms the basis for determining the knowledge gain across the first activity.
- **Knowledge gain**. The change in knowledge state as assessed across the beginning to end of a particular activity.
- **Knowledge map**. The mapping of connections between learning objectives for each course.
- **Knowledge state**. A student's current level of knowledge, updated after the end of each learning activity in the adaptive system. It includes knowledge demonstrated from the current course as well as any relevant activities from prior courses.

- **Learning management system (LMS)**. The online system (e.g., Canvas) in which the course is organized, including the syllabus, online discussions, assignments, and gradebook.
- **Modality**. The format by which content is presented to a student. In this dissertation, the possibilities include text, video, audio, and interactive presentations, as well as activities mixing these modes.
- **Module/Week**. One week of the class, including the content covered broken down into activities within the adaptive system. There are six modules in each course. The terms module and week are used interchangeably.
- **Randomized Control Trial (RCT)**. An experimental study design where individuals are randomly assigned to either treatment or control conditions.
- **Session/Subterm**. Three subterms occur during the fall semester and three in spring. The terms subterm and session are used interchangeably to refer to these six-week intervals when classes occur.

1.8 Research Overview and Hypotheses

Next, in Table 1.1, I re-present my research questions and connect them to the analytical approach and data used to address each. Based on the research literature discussed in subsequent chapters related to the research questions posed, I constructed hypotheses about what I would expect to be true for students. I address these three research questions and their associated hypotheses in turn in the next three chapters.

Table 1.1 Correspondence Between Research Questions, Methodological Approach, Data, and Hypotheses

Research Question	Method	Data	Hypothesis
1. What are the effects of choosing more than one modality (either text, video, audio, interactive, or mixed) for learning course material on knowledge gain?	Panel data analysis.	Fall 2018-Spring 2019.	Use of more than one modality has a substantively important positive effect on learning gains (>0.2 effect size).
2. What are the effects of choosing more than one modality for learning course material throughout the week on subsequent weekly assignment grade outcomes (or throughout the course on final course grade)?	Instrumental variable analysis (where instrument = assignment to RCT treatment group).	Spring 2018 RCT data.	Use of more than one modality has a positive effect on grade outcomes of at least half a letter grade (5 points out of 100).
3. How can information about modality switches and tutoring be used to predict later learning module success in one week of an introductory English course?	Demonstration of using a Bayesian network for decision making.	English course taught throughout Fall 2018-Spring 2019.	Combining modality switches and tutoring is predicted to benefit some students, showing potential to inform tutoring support recommendations.

In summary, addressing the research questions enumerated above will extend prior research literature by guiding me to: 1) examine the effects of use of multiple modalities on student learning outcomes in ways that have not yet been explored, 2) seek confirmation for the effectiveness of an aspect of UDL theory, and 3) provide a proof-of-concept analysis predicting results about the connection between an element of UDL and student success. I do this employing unusually comprehensive data using rigorous, advanced analytical methods that could be extended to other circumstances. Overall, this dissertation contributes to addressing systemic inequality in higher education, particularly for students of varied ability who have been traditionally underserved by that system. It does this by providing informed guidance for students with and without disabilities as they make choices about how they will engage with course content, while improving understanding about fruitful areas of focus for faculty development and course redesign efforts to provide appropriate options that support students.

CHAPTER 2

LEARNING ACTVITIY ANALYSIS

> If UDL is nothing more than providing students with alternatives, it fails significantly as a new paradigm for enhancing educational achievement, as it is simply another futile attempt to argue that schools [need] more resources. I choose to believe the critical focus of UDL is its emphasis on the variables that can be manipulated to produce high performance. I am inspired by Tomlinson's (1999 [2014]) conceptual work on the design of equalizers that could be utilized to manipulate key instructional variables to make curriculum accessible and engaging. (Edyburn, 2010, p. 39)

Considering the great variability among learners, finding instructional strategies that work well across ability levels holds importance for educators (Meyer et al., 2014). In the wake of COVID-19, many students have experienced learning challenges, including students in a variety of already underperforming groups at all levels of education (Dorn et al., 2020; Herold & Chen, 2021; Manly et al., 2021; Zhang et al., 2021). This has magnified the need to identify effective strategies for assisting learning acquisition in college (Office for Civil Rights, 2021). The present analysis probed one promising technique posited to support students across the full range of ability levels, including students both with and without identified disabilities (Tobin & Behling, 2018). Specifically, considering the Universal Design for Learning (UDL) framework's guideline to provide alternative modalities in representing content (CAST, 2018), I investigated the effect of use of multiple modalities on formative learning activities throughout a variety of online college courses.

Educational researchers have surprisingly little empirical evidence to show how specific educational design practices facilitated by new technological capabilities, such as practices recommended by UDL (Burgstahler, 2015), translate to learning success and other college outcomes (Kimball et al., 2016; Mangiatordi & Serenelli, 2013). I address this issue by investigating student learning success in online courses that use a learning-analytics-infused system for delivery, which facilitates implementing some aspects of UDL. This framework, following in the footsteps of efforts to improve universal design in the built environment (Hamraie, 2017), arose from work to improve educational circumstances for people with disabilities and is grounded in cognitive science (Rose, 2001). UDL theorizes students benefit from multiple means of engagement, multiple means of representation, and multiple means of action and expression in their studies (CAST, 2018).

Importantly for this study, UDL suggests that having content available in different modalities will result in beneficial outcomes. I investigate the causal effect of using multiple content modalities (i.e., text, video, audio, interactive, or mixed content) on student learning outcomes for nontraditional undergraduates at a women's institution. Combining data from the adaptive learning system and other campus systems, I aim to discover whether the multiple content presentation recommended by UDL is beneficial for these students. I undertake the present research because the efficacy of using multiple content modalities as proposed by UDL still needs to be rigorously and empirically investigated in practice (Rao et al., 2014; Roberts et al., 2011). A small research base has begun to be built, such as with the work of Hall et al. (2015) that involved formative assessment in 14 middle school classes including a control group. However, that work looked more broadly at online versus offline reading. Thus, while some UDL research has addressed content representation, overall, few researchers have studied the connection between content presentation modalities and student learning outcomes, even as part of more comprehensive research on UDL (Capp, 2017; Cumming & Rose, 2021; Rao et al., 2014). My research addresses this gap.

I conducted a theoretically informed panel data analysis of an authentic learning situation (Mayer, 2008), taking a short longitudinal approach. I look at change within student across two consecutive time points logged approximately 20 minutes or less apart across a single activity. I averaged across all activities to investigate an overall effect posited to be observable across learning activities in a variety of courses. My goal was to identify the extent to which any beneficial effect existed when a student utilized more than one modality when learning course content. My research question was: What are the effects of choosing more than one modality (either text, video, audio, interactive, or mixed) for learning course material on knowledge gain? I hypothesized that use of more than one modality (such as text, video, audio, interactive, or mixed content presentation) when learning content would have a substantively important positive effect on learning gains.

2.1 Theory and Literature Review

I framed my study by considering how online courses could be designed to support a universal range of abilities while simultaneously viewing each student individually and holistically. I drew upon the theoretical framework offered by UDL (Burgstahler & Cory, 2008), which facilitated systematic investigation and critique of intentional integration of different modalities in course design.

The concept of universal design came from the idea that designing a built environment accessible to people of all ability levels would produce a setting that enables rather than disables participation by all individuals. In the early 1970s, architect Ron

Mace's experiences as a college student at North Carolina State University, where "Mace had to be carried up and down stairs to attend classes and was unable to use the men's restroom because his wheelchair was too wide to fit through the entrance," led him to pioneer the idea of barrier-free design, which he and others later expanded to the idea of universal design applying to everyone (Evans et al., 2017, p. 277). Ramp structures such as curb cuts offer a standard example of universal design, as they make it possible for people requiring wheelchairs for mobility to access spaces that would be inaccessible via curbs or stairs. Although curb cuts were initially designed with disability access in mind, they are usable by and useful for many other types of non-disabled individuals, such as people wheeling a stroller or grocery cart, or those rolling luggage.

Similarly, in higher education, universal design has been seen by advocates as poised to become "a mainstream concern and a discourse serving the needs of students at large," partly because "the wider objective of increasing diversity on campus is exceptionally well served by the model," a conclusion drawn from an institutional case study of faculty, administrators, and other employees (Fovet & Mole, 2013, p. 124). Educational frameworks based on the idea of designing for a universal audience include Universal Design for Instruction (UDI), Universal Instructional Design (UID), and UDL (CAST, 2014; McGuire, 2014; Scott et al., 2003; Silver et al., 1998). In these frames, all people are viewed as having potential to benefit from design providing essential access to an otherwise disadvantaged subpopulation. This situates disability in the environment rather than in individuals, whatever their current level of physical or mental capability, and whatever their educational background (Evans et al., 2017). All these frames also value providing content to students in multiple modalities. While parallels exist between these formulations of universal design in educational settings, and each have different strengths, I focus on UDL because of its explicit articulation of multiple means of representation as a guiding principle, which can be explored empirically through offering options for perception as per the UDL guidelines.

UDL draws specifically upon brain imaging research, guiding learning design to facilitate academic achievement by diverse students whose capacities may vary significantly across the brain's affective networks, recognition networks, and strategic networks (Rose, 2001). Recognizing that each student possesses a unique combination of strengths and weaknesses in each cognitive area enriches understanding of the dimensions across which human capability varies, informing design of educational experiences (Rose et al., 2006). However, UDL literature has more frequently focused on arguments for UDL's importance than on empirical study of its effects and effectiveness (Mangiatordi & Serenelli, 2013; Roberts et al., 2011). What empirical research exists about universal design has often focused more on K-12 than college (Crevecoeur et al., 2014; Rao et al., 2014), and on perceptions or implementation activity rather than learning outcomes (e.g., Abell et al., 2011; Kortering et al., 2008; Lombardi et al., 2011). The modality use studied here falls under UDL's principle of providing multiple means of representation (CAST, 2014).[2] Within this, UDL recommends providing options for perception, connected to the brain's recognition capacity.

[2] I use the terminology representation (which is the official wording used by CAST) and presentation interchangeably.

2.1.1 Modality Representation and UDL

UDL assumes that students enter college with a wide range of ability and prior experience and that students benefit from flexible paths to facilitate their learning. Courses designed with UDL in mind offer students multiple means of representation, including alternatives for auditory and visual information (CAST, 2014). The point of such multiplicity is not to offer additional complexity to students or add detrimental cognitive load (Beacham & Alty, 2006; Greer et al., 2013). Instead, course design offering multiple modalities should avoid unnecessarily increasing cognitive load, so UDL designers should be aware of potential pitfalls when presenting options simultaneously that may actually increase barriers for some students (Kohler & Balduzzi, 2021). A design goal would be to allow students to pursue alternate paths through a course's content if they struggle to learn along the initial path or are functionally unable to follow a particular path.

This flexibility of approach aligns well with availability of multiple modalities for alternative content presentation in the type of adaptive system used in the present study (e.g., Cavanagh et al., 2020). Past small-scale experimental research on an adaptive system where content was available in different modalities found benefit to student learning (Mustafa & Sharif, 2011). However, that research focused on adjusting initial mode of content presentation to individual learning style rather than investigating the role that the availability of additional modalities may have played. Although there has been great interest in investigating how content in different modalities might be presented to students in e-learning systems (Khamparia & Pandey, 2020), this research has not yet evaluated student learning outcomes connected to such presentation alternatives. Thus, even though adaptive systems may be designed to facilitate use of multiple modalities, the effect of doing so in them remains unknown.

Modality has been included in prior research about UDL, although not as a sole research focus. When considering content presentation overall, Orr and Hammig (2009) searched specific peer-reviewed journals from learning disability and higher education fields, between 1990 and 2008, explicitly excluding K-12. Focusing only on quantitative or qualitative empirical articles, they found 38 articles with research pertaining to UDL and learning disabilities. Their content coding found 10 of those articles contained a theme of multiple means of presentation. Thus, although presentation is not typically the sole focus of studies, it is fairly common for it to be explicitly included.

Several studies have shown support for positive student outcomes associated with universal design overall. For example, University of Minnesota faculty ran several studies of UID that found positive results for students, including higher grades and lower need for accommodations (Evans et al., 2017). Four of 80 abstracts reviewed by Mangiatordi and Serenelli (2013) included assessing student academic improvement, providing general support for the expectation of positive learning outcomes for UDL practices, though apparently none of these explicitly studied providing options for perception. Also, since these authors only reviewed abstracts and did not provide a list of articles reviewed, the quality of these studies is unknown. In a meta-analysis by Capp (2017) that analyzed 18 pre- and post-test UDL intervention studies published between 2013 and 2016, spanning both K-12 and postsecondary education, UDL proved effective overall at improving the learning process (i.e., a positive effect size was reported). Two of the quantitative studies found positive effects for providing multiple means of representation via student self-reports of perceptions. One of these compared pre/post-questionnaires of almost 400 introductory psychology students' perceptions before and after the faculty teaching their course received UDL training, comparing these results to a control group of over 200 students whose faculty did not receive the training (Davies et al., 2013). The other study used a convenience sample of 60 students responding to a pre/post-questionnaire for one department's redesigned course study guide (Tzivinikou, 2014). Thus, studies connecting the efficacy of UDL practices to actual course outcomes remain scarce.

2.1.2 Modality Representation and Cognitive Science

The idea that humans process sensory input through multiple channels, including visual and verbal channels, has substantial evidence (Mealor et al., 2016). Likewise, the idea that human working memory has dual channels for these two pathways has years of experimental support (Mayer & Moreno, 1998). The existence of these channels, theorized as encompassing visual and verbal (i.e., dual) sensory input, as well as visual and verbal processing channels housed within working memory, supports the idea that use of these two input channels may connect to learning outcomes (Clark & Paivio, 1991; Mayer, 2001). Multimedia research tends to investigate simultaneous use of these modalities, reserving investigation of the sequential modality use I focus on here to the control situation (Mayer, 2001). I posit more remains to be learned about the benefits of sequential use than is currently known, however. While general multimedia research has found benefits of simultaneous presentation for learning certain types of content, other research has shown that neurodiverse individuals may process multimedia differently, calling into question a one-size-fits-all approach to multimedia design that assumes combining media benefits all learners in similar ways (Beacham & Alty, 2006; Wang et al., 2018). Multimedia research also tends to investigate fairly short chunks of content (e.g., a single sentence or short explanation), typically shorter than the 20-minute learning activities I study (Mayer, 2001). I draw upon the cognitive science and multimedia research literature to support studying use of multiple modalities in sequence as well as combination to assist struggling students with their learning.

Mayer and Massa (2003) investigated the idea that people fall into visual or verbal learner categories, and their factor analysis supported a visual/verbal distinction based on spatial ability differences, cognitive style differences, and multimedia learning preferences. Different neural information processing pathways have also been shown to operate for people with visual and verbal cognitive styles in MRI brain scans (Kraemer et al., 2009). Additionally, sensory input handling has been found to correspond to cognitive style preference by changing nonverbal information to verbal coding in the brain, for example (Kraemer et al., 2014). This supports the theory that dual pathways bring information from our senses to the point of long-term memory integration (Mayer, 2008). Additionally, memory has been found to be as good for verbal information of paragraph length presented in an auditory modality as for presention in a visual (textual) modality (Morris et al., 2015). It thus seems plausible that using these dual channels in sequential learning as studied here may aid long-term memory and associated learning performance requiring retention.

Adding to this, “choices made within the context of an authentic learning scenario” have been found to be distinct from preferences expressed on questionnaires (Mayer & Massa, 2003, p. 839). This suggests a difference between innate visualizer/verbalizer cognitive style and expressed preferences when learning. Likewise, experimental brain imaging research with 20 people suggested an individual's cognitive strategy in a given situation may be inconsistent with their questionnaire-determined cognitive style preference (Kraemer et al., 2017). These results suggest that while people are typically able to self-assess their learning style along the visual/verbal dimension, this preference may not correspond to the modality that works best for them when learning. This conclusion supports the idea that offering content in different modalities in an adaptive system may provide benefit for students.

Taken together, these research results suggest people use a variety of cognitive processes while learning, whether they are consistent or not with their preferred cognitive style. Combining different approaches may therefore be beneficial when individuals have a difficult time grasping information in the first way shown to them.

2.2 Context

The setting for the study was a well-established private, women's institution in the Northeast. Older than the average four-year college-aged student coming out of high school, these students typically juggled family and work responsibilities in addition to school. While such non-traditional students have frequently been underserved by higher education overall (Kazis et al., 2007), they have been supported and encouraged at this institution, where staff have continuously explored ways to structure course experiences and utilize data to better support student success.

The three-credit undergraduate courses studied were taught in an accelerated six-week format as part of a variety of degree programs. At this institution, a semester contained three sequential subterms of six weeks each, a format that allowed students to take multiple courses during a single subterm or multiple courses during a single semester by taking only one course at a time across several six-week subterms. During the year, courses were offered across a total of six subterms, which corresponded to three times during each semester. This accelerated format facilitated working students focusing on one (or two) courses at a time while still completing multiple courses in a semester.

All courses were taught using the same technological interface for students and faculty, combining a learning management system for discussion and overall course interactions (e.g., weekly assignments and grades) with an adaptive learning system for content presentation and learning mastery level formative assessments based on multiple choice questions. The courses studied had each been redesigned over the prior three years using a team-based course design process utilizing Open Educational Resources (OER) placed within the adaptive learning system. This process was informed by the Quality Matters (2020) online course design rubric and associated annotations. Content appropriate for each course was formatted utilizing a similar structure that allowed the redesign team to code the activity modality types as a field within the adaptive system's backend data, enabling the present analysis.

The courses analyzed for my study incorporated aspects of UDL by design as integrated in the Quality Matters (2020) rubric. This rubric, which aims to ensure high quality in online courses overall, guided the strategies undertaken to improve student success. The rubric specifically encouraged practices addressing multiple means of representation, and generally encouraged following other aspects of UDL in course design (Robinson & Wizer, 2016). This aligned with arguments for broad use of UDL as a design strategy beneficial for diverse learners (Bradshaw, 2019; Tobin & Behling, 2018). UDL adoption should go further than preserving a status quo of what constitutes good teaching that has not served some students well (Edyburn, 2010). Here this meant widespread redesign of courses incorporating UDL principles as well as concern for web accessibility in line with the deep mission-driven desire to improve educational success for students typically marginalized in higher education. The inclusion of multiple modalities for learning course content was a deliberate design choice made by the institution across the courses studied. Alternate paths for learning content through different modalities were part of standard course design. In an approach consistent with universal design principles around providing alternatives for perception, the adaptive learning system encouraged students showing signs of struggling to pursue paths using alternate modalities until they achieved successful content mastery. To illustrate the nature of the content modalities studied, I explain an example from an introductory English course that is the second course of a sequenced pair of courses. During an early week in this course, students were expected to gain competency in skills that would support their approach to writing. A structured sequence of activities took students through the concepts needed to develop competency in the targeted skills, as in the following example of a sequence entailing three connected activities.

In the first activity of this example, students learned about choosing a topic to write about. By default, the content for this activity was presented to them as text. When they sufficiently mastered the concepts covered in the activity, as demonstrated by achieving at least 70% on a series of three to five multiple choice questions, they were allowed to progress to the next activity about how to write a thesis statement. If they showed signs of struggling by not achieving at least 70%, they would have been redirected to review the material and a recommendation would have been made to view the material in a different modality. If the original content had been presented as text, they would typically have had the option to view a video if they were struggling to learn the concepts covered in the activity. For some activities, content in additional modalities including audio, interactive exercises, or an intentionally designed mixture of content types would also have been available in addition to text and video. These additional modalities would have been accessed in a similar manner if needed by the student. Mastery of the second activity on writing a thesis statement, again demonstrated through responses to a few multiple-choice questions, then would bring the student to the final activity of this sequence on writing a proposal.

In each activity, required concepts could be presented in multiple ways (i.e., different content modalities), as crafted by the course development team. As students progressed through the course, their knowledge score based on the questions answered, time spent actively working on the activity, and the modality used were recorded for each activity. If they repeated an activity in a different modality, that also was recorded. This type of learning path with multiple modalities was created by developing alternative activity content for each learning objective utilizing OER to the extent possible to reduce costs for students. Depending on the subject and course, additional assignments, quizzes, and projects were also assigned and graded, as well as required weekly online discussion participation. The result was a robust dataset with measures of student action and knowledge captured in an ongoing way throughout each week of a course. This allowed the analysis of student utilization of more than one modality when learning.

2.3 Data

The analysis sample included 1,278 women undergraduates enrolled in 283 sections of 51 online courses taught during the 2018-2019 academic year. These courses spanned 14 subjects, including sciences, social sciences, humanities, and professionally oriented courses. Student performance data allowed study of the impact of using multiple modalities for representing course content on course-related success.

Several technical features facilitated the collection of these performance data, including a data warehouse and student anonymization. Student-level information was gathered from multiple campus systems and combined into a data warehouse, including data from the learning management system, the adaptive platform for course content and formative assessment, and the administrative student information system. Student information was anonymized prior to the researcher having access, addressing privacy concerns and facilitating approval for this secondary data analysis by the institutional review board (IRB) at the institution providing the data.

Data were collected across multiple instances of all courses using the adaptive system during the 2018/2019 academic year. Each six-week course was broken down into learning activities each anticipated to take about 20 minutes, with approximately 5-15 activities per week in the adaptive system. This resulted in 199,396 cases for analysis.

A student’s prior knowledge of the upcoming content was assessed at the beginning of each week and a starting knowledge state score was assigned. Their knowledge of the content covered in an activity was also assessed at the completion of that activity. Information was captured about when and for how long students worked on the activity, as well as any repetitions of the activity, and the modalities utilized each time they went through an activity. These features made these data well-suited for an aggregated analysis of modalities and learning across multiple courses.

Data were analyzed for each student at the activity level across all courses. Each activity instance completed by each student was given its own row in the data, with variables identifying whether a second modality was used at any point during that activity's completion along with the student's knowledge gain for that activity. Each activity typically had three to six content sections including a short introduction, a long section where most of the content was presented, and a summary. Questions assessing formative understanding were also asked. Sometimes at the end of a main content section, the student was asked if she would like to view alternative content. If she chose to do so, she was often presented with content in an alternative modality, such as video if the main content was presented as text. Given that my research question aimed to identify an overall connection between use of more than one modality and learning gains, aggregating the data in this way across courses was sufficient.

2.4 Variables

This section explains the outcome variable, primary independent variable being studied, set of independent conditioning variables, and how missing data were handled. Appendix A describes the operationalization of these variables.

The outcome was change in knowledge score. An initial weekly knowledge score was assigned after determining the student's prior knowledge of that module's concepts, and exit assessments occurred at the end of each 20-minute learning activity in the adaptive system. The knowledge gain for an activity was calculated as the difference in a student's knowledge score before and after going through that activity.

The primary concept of interest was the use of multiple content representations. Up to five alternate paths for learning content through different modalities were designed into each learning activity in each course. The use of multiple content representations was operationalized as student use of any second modality of content representation for a given 20-minute activity in the adaptive learning system. While not all activities had the same number of modalities available, most had at least two modalities, making this treatment operationalization relevant for the greatest number of activities possible.

Two additional course-related variables were considered for model inclusion: 1) the amount of time spent on the activity since time on task may impact learning, and 2) the combination of year and term for the course since content may have been updated between terms but not during terms per institutional policy. The final analysis model excluded these conditioning variables, however, as explained below.

Missing data were not a pervasive problem in this study. Data were only missing on the dependent variable, and such cases could be missing because they were missing either the starting or ending knowledge state score. When students worked on the initial assessment at the beginning of the week that determined their starting knowledge level for that week's material, this activity legitimately had no beginning knowledge state score, and since this initial assessment activity was not associated with modality use while learning content, these cases were dropped from all analyses. Some students elected not to or were unable to complete the ending formative assessment after working on an activity, and this resulted in missingness for the ending knowledge score in 22.4% of cases. This type of missingness was expected due to the work and life demands of these non-traditional students. Given that the student had no end score in this situation, these cases were also dropped from analysis. This left only cases of activities where the student had both a beginning and ending knowledge state score to be analyzed.

While I acknowledge observational research always has the potential for bias due to unobserved and unknown selection effects that could be associated with missing data, sources of such bias are not anticipated for this study, as it seems probable that random life events led to the missing ending score. However, if struggling students were more likely to have given up and not completed the ending assessment, then that might positively bias the results. It is also possible though, that such missing data came from students who completed the activity with a sufficiently high score to continue along the activity sequence, but who chose to review material without completing another assessment, potentially negatively biasing the results. As a sensitivity analysis utilizing all available information about these students, missing data for the ending knowledge score were also handled via multiple imputation (Manly & Wells, 2015), with similar analytical results (see Appendix A).

2.5 Methods

2.5.1 Descriptive Analysis

Given the range of courses studied, I began by breaking down the number of students, activities, and uses of multiple modalities seen across different fields of study to gauge the spread of the data across fields. Calculating means and standard errors for the analysis variables offered a descriptive sense of the data. (See Appendix A for a correlation matrix.)

Additionally, I compared demographic differences between groups that do and do not use a second modality. This allowed me to investigate the potential for threats to validity caused by confounding effects of latent variables that might have caused systematic differences in outcomes of interest between groups. While there was not much that could practically be done if such problematic latent variables were unobserved, investigating systematic differences in who chose to use multiple modalities allowed me to probe for potentially problematic areas that might warrant future investigation.

Aspects of the data were investigated that relate to the nature of these panel data as well. These included panel balance, the amount of variation within subjects, and whether potentially problematic time-related trends were discernable through plots.

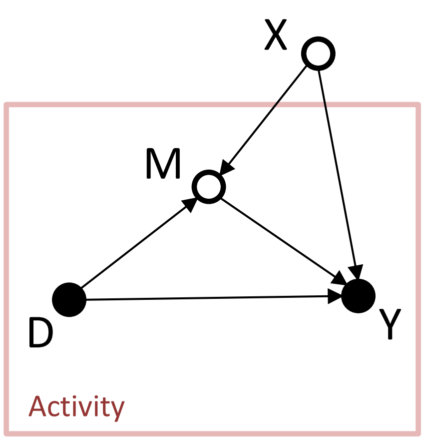
2.5.2 Panel Data Analysis

I utilized both associational and causally oriented approaches to statistical inference while addressing the clustering present in the data. Investigating causal effects offers a particularly important and often overlooked direction for higher education research that has become increasingly possible given the more nuanced individual learning data now available through online learning systems such as those used in this study (Schneider et al., 2007). After beginning my inquiry with an ANOVA analysis to gauge the basic relationship between use of multiple modalities and knowledge gain, I explored several causally-oriented modeling approaches.

To more fully understand the relationships in the data, my approach utilized causal graphical modeling (CGM) to represent alternative causal hypotheses that might be investigated and determine which to pursue further (Pearl, 2009b). Using CGMs to represent alternative structures facilitated investigation of causal effects by aiding my modeling choices. CGMs represent random variables as nodes and causal relationships between random variables as uni-directional causal arrows between those nodes. When necessary, bi-directional arrows can also be used to indicate latent confounding. CGMs are explicit about the direction of causation whereas those relationships are either implicit or unclear in many other types of models (e.g., structural equation models). The pattern of connections among random variables that is asserted in the model directly implies marginal and conditional independencies that can be tested with data. Knowledge of the data generating process can be used to constrain the potential space of possible CGMs. For example, theory, prior research, and knowledge of time ordering can be used to infer the existence or direction of causation. In addition to the conceptual benefits CGM can provide for developing and understanding models based on subject-matter knowledge, with a large enough dataset, relationships can be learned algorithmically from the data (Pearl, 2009b; Spirtes et al., 2000). This was done here, given that the dataset had almost 200,000 observations. Known logical relationships provided constraints on this learning process to speed the processing and ensure the resulting model conformed to reality, with logical characteristic-based and time-based relationships being reflected properly. This process of model-building and testing was conducted iteratively and flexibly to determine the most appropriate model for subsequent analysis.

Using this CGM-based approach, I began by representing the variables I expected to be related, including the use of multiple modalities (treatment, D), time spent on the activity (a potential mechanism, M, through which use of multiple modalities may have operated), the year and term to reflect possible changes in the curricular material (exogenous control, X), and the knowledge state gain score (outcome, Y). The structure of connections between these variables was also learned from the data through several Bayesian network structure learning algorithms using the *bnlearn* R package and the results were compared to each other. These included structure learning algorithms that were constraint-based (grow-shrink, PC, and incremental association), score-based (Tabu and hill climber greedy search), and hybrid (two-phase restricted maximization, max-min hill climbing, and hybrid HPC). The models learned were constrained by prior knowledge about temporal-based relationships as well as the assumption that static characteristics (e.g., year and term) would not be predicted by other variables. After learning the edges representing relationships between these variables (i.e., nodes) from the data, the conditional probabilities of the nodes were learned by the algorithm. Testing found that the network structures from the different algorithms belonged to the same equivalence class, which means these learned models from each algorithm imply the same set of conditional independencies. The resulting model of the underlying data generating process, shown in Figure 2.1, indicated D and X were independent in the data.

Figure 2.1 Graphical Model Learned from Data



Where:

D = Use of multiple modalities (treatment)

M = Time spent on activity (mechanism)

X = Year/term (exogenous)

Y = Knowledge state gain score (outcome)

● = in panel data model analyzed

Inspecting this model led to the conclusion that neither X nor M should be included in the analysis model. In this study, I am interested in the overall effect of use of multiple modalities on knowledge gained by the students, not specific mechanisms that might partially explain that effect, although exploring potential mechanisms, such as time on task or task repetition, could be investigated in future research. While year/term could have been included to increase precision of the effect estimate by reducing the variance in the outcome, this was deemed unnecessary given the very large sample size. Additionally, year/term is not needed to estimate the treatment effect under a model without M. An additional exploration to learn the model when including many more of the variables that were in the full dataset did not reveal any available variables that cause both use of multiple modalities and knowledge gain (i.e., parent variables that are common causes in the language of CGMs) that ought to be included in the model. This means that whatever causes a student to use multiple modalities is either not relevant to include given my research design (such as student-level variables) or not observed and therefore not amenable to empirical investigation at this time (such as course-related variables like quality of the material or recommendations made to the student by the adaptive learning system to review the material in another modality). My conclusion from this model exploration was that the most appropriate model given my research question and the directional relationships that were learned from the data was a very simple panel model with only treatment and outcome, taking the clustering by student into account. Thus, I assumed that the data generating process could reasonably be modeled utilizing a clustered regression analysis based on this simple graphical model (see the black dots in Figure 1). I compared results from ordinary least squares (OLS) and panel data analyses conducted as follows.

The OLS regression was adjusted for clustering by student using Stata's *regress, vce(cluster id_student)* command (Cohen et al., 2003). To confirm the appropriateness of regression for the continuous dependent variable of the gain in knowledge state score across a single activity, I verified that the assumptions of regression were met sufficiently. I also found no potentially problematic outliers.

I probed the causal connection between treatment and outcome using a panel data analysis with Stata's *xtreg, fe* that accounts more appropriately than OLS for the clustering of the data within individuals (Cameron & Trivedi, 2009). I took a short longitudinal approach, looking at change within student from before to after each learning activity expected to take approximately 20 minutes. The longitudinal nature of these data facilitated calculation of a change score across these two consecutive time points, and so a panel data analysis was appropriate to estimate the causal effect of use of multiple modalities (Hsiao, 2014). This approach is known to econometricians as "a panel data variant of a difference-in-difference model" (Morgan & Winship, 2015, p. 364). Given that I had such longitudinal data from many students over courses each lasting six weeks, I estimated the effect across all activities to investigate an overall effect posited to be observable across heterogeneity in course settings and activity types.

2.6 Limitations

Several features of the data and method should be noted when interpreting the results. Although the sample likely contained many students who have disabilities, their number was unclear. This lack of clarity limits conclusions from these results for students with disabilities specifically. However, the sample's atypically low rate of official course accommodations (0.6% compared to 19% nationally; Snyder et al., 2019) may be due to the intentional design of these courses incorporating UDL principles and being guided by the Quality Matters rubric for online course design (CAST, 2018; Quality Matters, 2020). That is, students who may have felt the need to receive accommodations in other circumstances may not have needed them for these courses. Alternatively, while it is possible that few students with disabilities chose to attend this institution in the first place, past research indicates that many students with disabilities choose not to disclose their disability in college for a number of reasons even if they had accommodations earlier in their education, and many do not know that such supports exist (Gierdowski, 2021; Newman & Madaus, 2015). Additionally, pursuing updated diagnosis and arranging for accommodations can be time consuming and expensive, and such costs may have been perceived as prohibitive, particularly for students with jobs and families who may not have much time flexibility to pursue the required process (Fox et al., 2021). Unfortunately, the reason is not possible to distinguish from available data.

Given the very large amount of data employed (almost 200,000 cases), significance tests are nearly meaningless, as even very small effects can be significant with enough data. Because of this, to aid interpretation, confidence intervals are reported to indicate estimate variation and effect sizes are emphasized.

Additionally, a potential issue with the panel data approach was that treatment assignment may have been "fuzzy" since students who received a recommendation to use a second modality might not have followed that advice. Unfortunately, it was not possible to obtain additional data about the recommendation offered to students, as the adaptive learning vendor considers this proprietary information. This limitation of the present research could be addressed in future research where such data became available.

2.7 Results

Of the 1,278 students in the sample, many took courses during both Fall and Spring sessions. As shown in Table 2.1, 2,566 learning activities were engaged in by these students. Almost 200,000 instances of activities with modality data were logged across the humanities, professional studies, math and sciences, and social sciences, and more than one modality was used over 100,000 times, again spread out by field.

Table 2.1 Fields of Study, Students, Activities, and Modality Use

Field of Study	Students	Activities	Activities Engaged by Students	Times >1 Modality Used
Humanities	745	594	56,821	31,301
Professional studies	557	1,168	59,466	41,074
Math and sciences	417	343	40,682	12,061
Social sciences	585	467	42,427	31,187
Total	1,278	2,566	199,396	115,623

Table 2.2 shows that across all activities, the mean knowledge improvement was 0.131 (on a 0-1 scale). 58% of the students used more than one modality while working on an activity. On average, students spent about 7 minutes (0.124 hours) on an activity, and data were spread reasonably evenly between the two semesters.

Table 2.2 Estimated Means and Standard Errors of the Estimates

Variable	Mean	SE
Knowledge state gain across activity	0.131	<0.000
Use of >1 modality?	0.580	0.001
Hours spent on activity	0.124	0.001
Fall 2018	0.530	0.001
Spring 2019	0.470	0.001
Observations	199,396	

When investigating the data's panel nature, since the fixed effects estimation relied on having good variation within subjects, a variance decomposition was conducted which confirmed that sufficient variation existed. The panel was unbalanced, with varying case numbers for students, because different courses had different numbers of activities and students could choose whether to complete them. Within the courses, data were clustered for each student. Checking time-series plots showed no potentially problematic discernable trends over time. Testing for heteroskedasticity suggested that using clustering by student was indeed appropriate for these data ($\chi^2 = 868$, $p < 0.001$). A simple cross-validation check splitting the data by semester confirmed that results in each semester were quite similar to those presented below. Finally, a robust Hausman test confirmed the appropriate use of fixed effects for these data ($\chi^2 = 39$, $p < 0.001$). Thus, my analysis focused on a fixed effects panel data model.

As shown in Table 2.3, the average marginal effect of the use of more than one modality to learn the content in an activity was 0.049 when calculated with a fixed effects panel approach (model 2) accounting for student-level factors that might influence results. Clustered regression results (model 1) are presented for comparison. (See Appendix A for discussion of additional methodological considerations.) The panel coefficient (0.049) is equivalent to a standardized effect size of Hedges' $g = 0.224$ (see

Table 2.4). This can be interpreted as a reasonable effect size for education since Cohen's labeling of 0.20 as small and 0.50 as medium "can be misleading in educational policy contexts, in which effect sizes of 0.20 or smaller are often of policy interest" (Hedges & Hedberg, 2007). Recent guidance for educational interventions considers effects over 0.20 to be large (Kraft, 2020), though some in higher education would argue for slightly larger values (Mayhew et al., 2016). The effect found corresponds to an improvement index of +8.9 (above 50th percentile), which is equivalent to a comparison student improving from the 50th to the 59th percentile (What Works Clearinghouse, 2020a).

Table 2.3 Average Marginal Effects for Knowledge Gain Across an Activity – Clustered Ordinary Least Squares Regression and Fixed Effects Panel Data Analyses

Variables	(1) Regression with student clustering (SE)	[CI]	(2) Fixed effects panel data analysis (SE)	[CI]
Use of >1 modality?	0.069***	[0.061,	0.049***	[0.043,
	(0.003)	0.076]	(0.003)	0.056]
Constant	0.091***	[0.084,	0.103***	[0.099,
	(0.003)	0.098]	(0.001)	0.106]
Observations	199,396		199,396	
R^2	0.023		0.009	

Note: *** p<0.001, ** p<0.01, * p<0.05
ANOVA results were identical to (1). Standard errors in parentheses. Confidence interval in brackets.

Table 2.4 Hedges' *g* Effect Sizes Corresponding to Analysis Models in Table 2.3

	Comparison		Treatment		Impact			
Model	Mean	SD	Mean	SD	Estimate	SE	*p*	Effect Size
(1)	0.091	0.206	0.160	0.228	0.069	0.003	<0.000	0.314
(2)	0.091	0.206	0.140	0.228	0.049	0.003	<0.000	0.224
N	83,773		115,623					

The sensitivity of results to choices made in the analysis process were probed through several additional analyses (see Appendix A). Substantively similar conclusions to those presented were drawn when using two alternate operationalizations of the “mixed” treatment category, running OLS regression including additional covariates, multiply imputing missing values on the dependent variable, and adjusting for clustering by activity. Additionally, learning the structure of more complex models through a Bayesian network approach did not suggest potential confounders that should be included for an analysis of the overall effect, although it did suggest possible mediating factors that could be investigated in future research.

To check for balance in demographic and prior educational factors across treatment groups, I tested for differences between cases that did and did not use multiple modalities in race/ethnicity, Pell grant status, age, withdrawals and failures in the prior semester, prior GPA, and the number of credits transferred in when the student entered the institution. Finding only a significant difference by age, I probed further and found a difference between students under and over the median age of 31, although there was no difference within each of these groups. This suggests future research might explore differences in modality use between younger and older students.

2.8 Discussion and Implications

My analysis found a medium-large, educationally important effect of using multiple modalities on the knowledge gain students exhibit across a learning activity. This work extends calls to scientifically validate aspects of UDL, supporting guidance to provide flexible options for perceiving content as a way to deeply engage students with the material they are learning (Edyburn, 2010; Rao et al., 2014). On average, use of more than one modality predicted a 0.05 increase in knowledge score on a 0-1 scale across a learning activity over students using only a single modality. This corresponded to a student improving almost 10 percentiles above the median on the activity, a meaningful boost. This result is in line with expectations that providing content in multiple modalities will assist student learning (Rose, 2001). These results make one of the benefits of UDL for formative student learning outcomes concrete, offering a meaningful contribution to the universal design literature, which has been notably lacking in efficacy studies (Cumming & Rose, 2021; Roberts et al., 2011). Overall, the results of this study support UDL's claim that providing multiple means of representing content will be beneficial, quantifying that benefit for women in the adaptive learning context studied.

Given the large sample size, in determining how confident to be that the result indicates a real effect, I also investigated the amount of bias it would have taken to switch from a significant to non-significant finding (Frank et al., 2013). I note that the effect would have needed to be biased by 89.95% to invalidate the inference. Alternately, it would have taken a confounding variable correlated at 0.199 with both treatment and outcome to invalidate the result. Such a correlation with the outcome would have been stronger than the outcome's correlation with either the treatment (0.153) or the amount of time spent on the activity (0.039). This gives confidence that the result is quite strong, even considering the large sample size. The volume of data is a strength of this study, while also being large enough to warrant emphasizing effect size interpretation over statistical significance.

While caution is always warranted when making causal claims, the panel nature of the data employed means that person-centered variables that are difficult to measure and often confound observational studies should not bias these results. That is, in typical regression modeling accounting for clustered data, collecting data about personal background factors may be challenging or practically impossible. Such factors could include motivation, personality-based predispositions, or prior experiences that serve to increase engagement with the material. While observable characteristics can be measured and models adjusted appropriately, potential exists for unobserved characteristics to introduce bias. A panel approach essentially allowed a given student to act as her own comparison, automatically adjusting for person-related factors so they would not confound conclusions drawn. Non-student factors may still have biased the results, such as the quality of either the material, the course design, the implementation of UDL principles, or the instructor's teaching. However, use of the Quality Matters (2020) rubric by the institution in the development of these courses supported the assumption that such quality measures were held constant in this analysis, supporting a causal interpretation of the results. Thus, the panel approach held notable strength for a person-oriented outcome as studied here, particularly when coupled with approaches to ensure baseline course quality, even while future research about possible alternative explanations beyond student-level factors remains warranted.

Additionally, both a data science-oriented approach of learning the model from data and multiple sensitivity analyses probing the influence of a variety of choices made during analysis suggested confidence in the conclusions drawn. Although online course-taking during remote learning at the height of the COVID-19 pandemic may not reflect the voluntary nature of the choice to study online by the students in the earlier time period studied (Hodges et al., 2020), it is reasonable to assume future students will once again return to choosing online courses on a voluntary rather than forced basis. This has already started to happen, so the conclusions drawn here are expected to have relevance going forward. Thus, I claim with reasonable conviction that non-traditional, undergraduate women students of differing ability levels taking online courses benefit from the opportunity to learn content by utilizing multiple modalities across a range of humanities, professional, social science, and scientific disciplines.

2.8.1 Alternative Explanations

Multiple causal paths may underlie the improvement seen in learning gain given use of multiple modalities and it is worthwhile to consider such alternative explanations when interpreting the results of this treatment effect study. For example, learning might improve when students repeat the activity, giving them more exposure to the material. Although the importance of time for learning might have face validity and has generally been considered good practice to encourage (Chickering & Gamson, 1987), prior research has sometimes found positive (Wellman & Marcinkiewicz, 2004) and sometimes negative (Greenwald & Gillmore, 1997) relationships between time-related factors and student achievement, so the potential influence on the present study is unclear. Preliminary investigation of possible alternative explanations for the results reported here such as time on task and activity repetition did not appear to explain away the effect of use of multiple modalities when included in preliminary sensitivity analyses, reinforcing confidence in the claim of a positive effect that may be durable even when considering potential mediators. However, future mediation-focused research could investigate the extent to which these and other factors may be operating in concert to aid students' learning. From a causal perspective, such potential alternative explanations should be researched to determine the extent to which they are also important in understanding UDL and modality use.

Alternatively, an argument might be made that a particular modality is simply "better" at conveying certain content. For example, a faculty member who learned a concept in a particular way may believe that way to be "the best." However, UDL principles "[reflect] the fact that there is no one way of presenting information or transferring knowledge that is optimal for all students" (Rose et al., 2006, p. 137). Based on the reality that perceptual capabilities differ between individuals (Mealor et al., 2016), I would not expect that certain material would be found to be most effectively conveyed through a particular modality for all students. If future discipline-specific or course-specific research found evidence to the contrary, this would point to an alternative explanation that might confound the results of the present study and challenge this foundational UDL principle, though this seems unlikely.

As another candidate cause to consider, it is possible that the second modality used by students was better suited to their learning needs. Students may not have been guided to an optimal initial choice for conveying content by the default learning path in the adaptive system. Recognizing this possibility, over time, the adaptive learning system notes which mode of content presentation works better for a given student based on their performance and will begin presenting material in that modality first when alternative content is available (Cavanagh et al., 2020). Based on results refuting the matching hypothesis in the learning styles literature (Cuevas, 2015; Pashler et al., 2009), I would not expect that matching students who prefer a given modality type with material presented solely in that modality would improve learning. It is less clear whether using particular combinations or sequences of modalities might be beneficial given that such combinations may tap into the different brain regions people use when processing visual and verbal information (Kraemer et al., 2009). It is also unclear whether any such combinatory effect might differ for students who report particular learning preferences, such as visual or verbal (Mayer & Massa, 2003). The present results suggest that future research investigating specific sequences of modality use would be warranted.

As another possible cause, opportunities to make choices have been considered a component of student agency leading to improved academic performance (Jääskelä et al., 2021). Here, the agency that comes with freedom of choice to pursue different modalities may have been operating to aid students' learning. While this could be investigated in future research, to put this and other possible alternative causes in context as already noted, such a potentially confounding variable would need to have had a 0.2 correlation with both treatment and outcome to nullify the treatment effect. In that case, such a correlation with freedom of choice would have been stronger than either's correlation with time on task.

2.8.2 Future Research

The results suggest numerous additional intriguing directions for future research. The possibility exists that factors such as motivation to earn a high grade may moderate the results. That is, the institution's learning design team is aware that some students who are very motivated to earn a high grade will repeat activities over and over until they earn high grades on every activity. This anecdotal information is in line with prior research on agency, self-efficacy, and high performing students that has found motivated students do better academically and competitive students will work hard to achieve a high grade

(Alkış & Temizel, 2018; Ayllón et al., 2019; Baumann & Harvey, 2021). In the present research, a latent factor for such grade-based motivation was used in a regression sensitivity analysis (with no substantive difference in result; see Appendix A), but was not employed in the panel analysis since that factor was constant for a given student. However, future research could consider stratifying the sample by a measure of grade motivation to investigate the possibility that use of multiple modalities may operate differently for students with higher or lower motivation to achieve a high grade. In a similar vein, other potential moderating factors such as prior academic achievement could be investigated to gain a fuller picture of the circumstances under which use of multiple modalities makes the biggest positive difference for students.

Research taking a more nuanced look at the specific modalities used by students could also investigate whether use of more than two modalities offers benefit (e.g., in a dosage analysis). Existing theory about dual-channel visual/auditory processing suggests that the largest cognitive difference may come from use of modalities offering complementary visual and auditory sensory input (Mayer, 2001). From this standpoint, a third, fourth, or fifth mode that uses different combinations of sensing and cognition to process (e.g., video involves both visual and auditory elements) may duplicate the sensory input of either a single visual- or auditory-based presentation. It remains unclear whether the impact of use of a second modality is related primarily to a dual input distinction (i.e., eye and ear), to a dual processing distinction (i.e., visual and auditory), or to dual-channel pathways within working memory overall (Kraemer et al., 2009, 2014; Mayer, 2008). Given these dualities, the benefit of using multiple modalities may primarily be a benefit of using at least a second modality. Consistent with this supposition, a preliminary look at dosage suggested the biggest benefit may appear after use of any second modality. However, not all courses studied had content in more than two modalities available, so this characteristic of the data may have had a confounding influence on these preliminary dosage explorations. Future research should distinguish the limits and causes of dosage effects further.

Several other directions left unexplored by the present study could also be targeted in future research. For example, the effect may be stronger for some courses or subjects than others. The timing of presentation of content in alternate modalities by the adaptive learning system might matter, involving analysis of recommendations made to struggling students to use another modality. The number and type of modalities offered for a given activity may matter as well, presenting a potential confounding influence which could be researched, informing future course design.

Past research investigating learning through brain mechanisms involving multiple sensory pathways to memory supports the idea that the present results may have wider applicability (Mayer, 2008; Mayer & Moreno, 1998). When investigating the effect of the simultaneous presentation of media utilizing dual-channel sensory pathways in complementary fashion, multimedia research has found benefit in utilizing both visual and auditory sensory modalities when learners are to remember and integrate content information (Mayer, 2001). Although such work has focused on simultaneous presentation of multiple media, the content presentation in the present study also makes use of more than one sensory channel for learning, but primarily for consecutive presentation. Although multimedia may be more effective than single-mode presentation for some types of students learning some types of content, students with certain types of disabilities, such as dyslexia, may encounter difficulty comprehending material presented simultaneously in different modes due to the required cognitive load (Beacham & Alty, 2006). So, while the sequential type of presentation studied here is perhaps not as efficient a method of comprehending material as a full multimedia presentation for some individuals, it deliberately offers choices and alternative learning paths to students, giving them agency to utilize what works for them. Additionally, the adaptive learning system does not overwhelm students with too many options initially, keeping the cognitive load down, which can otherwise challenge some neurodiverse students depending on how options for multiple modalities are implemented (Kohler & Balduzzi, 2021). The results of the present study reveal potential benefits of combining ideas about dual-channel processing for memory with dual coding for cognitive load, supporting the idea that memory function is not necessarily dependent on the type of sensory input (Morris et al., 2015). That is, memory benefits that exist when utilizing both visual and auditory channels to reinforce learning appear to operate under the conditions studied here in ways that support retention of material without a potentially overtaxing cognitive load, thus effectively addressing a wide range of cognitive abilities.

It seems reasonable to suspect these results may hold more broadly even while acknowledging the limits of external validity for a single research study of one institution and the need to extend this work to a variety of student populations. Given the paucity of research literature addressing the effectiveness of practices based on UDL guidelines for improving student learning (Mangiatordi & Serenelli, 2013), it would be helpful to undertake studies exploring the extent to which these results can be replicated in other settings. Additional research could be undertaken to investigate the effect on outcomes at different time scales and in different institutional contexts as well as confirming the result for men as well as women. While more work remains to confirm these results with students who have known disabilities, current results support a claim of broad applicability for providing content in multiple modalities. The demonstrated benefit realized by many students in this institutional context suggests that course design steps such as those taken by this institution may reduce the need for specific accommodations, and future research should explore this further.

The current research also leaves open the question of why students choose other modalities. Having identified that use of multiple modalities provides benefit for non-traditional undergraduate women, future qualitative research could interview students to investigate why they chose to use multiple modalities, what they hoped to gain by doing so, and how they perceived the benefit obtained. Better understanding student motivation to engage in working through content in different modalities may help future educators design courses that encourage the positive aspects of this practice more explicitly.

2.8.3 Implications for Practice

The strong evidence presented here for an educationally meaningful positive effect of use of multiple modalities has important implications for practice. These results provide a compelling argument that faculty development and curricular design efforts should include the UDL principle of providing multiple means of representation for course content. That is, there are demonstrable benefits for formative learning gains when students are given the opportunity to encounter course content in more than one modality. Faculty development increasingly includes exposing faculty to universal design principles, and widely used guidelines for good online development incorporate UDL ideas (Higbee & Goff, 2008; Robinson & Wizer, 2016). However, even though faculty are often aware of the need to learn about and implement UDL ideas, this does not always translate to actual implementation (Cook et al., 2009; Izzo et al., 2008). Encouragingly though, faculty who have received UDL training are more likely to include multiple means of presentation in their teaching (Lombardi et al., 2011). In line with what has been termed the “plus one” strategy for approaching UDL implementation (Tobin & Behling, 2018), identifying key material where students typically struggle and adding an alternative for learning content in a different modality for that material may be a good place for faculty to start as they add to their UDL-informed practice and work toward fully incorporating UDL concepts. This study provides clear and compelling support for making options for content available, and action to achieve this can be encouraged in faculty training.

This study also provides concrete evidence that curriculum development efforts should include making content available in multiple modalities, particularly in adaptive learning systems, because students can see an improvement index of almost +10 above the median in their learning. At the institution studied, a systematic and comprehensive approach to including multiple modalities was strategically undertaken, with a design team adding such material to over 50 courses. Such modality options can include alternate text, video, audio, interactive, or mixed modality representations of content. The benefits seen suggest other institutions would be well advised to consider devoting resources to systematically developing options for students to go through material in science, social science, humanities, and professionally oriented fields. Offering students options for how content is presented is a commonsense UDL tenet with demonstrable benefit that is straightforward for faculty and institutions to implement if they have allocated sufficient resources for implementation. Such clear opportunities to improve practice are all too rare in postsecondary education and should be a call to action.

2.9 Conclusion

This study investigated the relationship between use of multiple content representations and formative student outcomes for 20-minute learning activities in an adaptive learning system. The goal was to better understand and help confirm UDL's proposition that providing multiple means of content representation benefits student learning. This work extends knowledge about UDL in practice by identifying the effect of the use of multiple modalities on formative learning done by women undergraduates as they engaged with content for online courses across multiple fields. By combining data from several campus systems, a comprehensive within-course dataset enabled estimates of effects to be revealed through a within-subjects analysis approach. Results support UDL's claimed benefit of providing options for perception by demonstrating quantifiable learning gains for students. This suggests that time spent by faculty and course developers modifying course material to incorporate different modalities offers clear benefit to students. These results should bolster administrative efforts to direct resources, such as faculty development funding and support, toward efforts to provide content to students in multiple modalities.

CHAPTER 3

WEEK AND COURSE GRADE ANALYSES

> Certainly, the current literature is starting to give definition and shape to what a UD educational model-based project or intervention looks like, but eventually researchers will need to address whether instruction incorporating UDL actually causes better results than conventional lessons and courses by conducting high-quality experimental studies, including true experimental, quasi-experimental, and single-subject designs. (Rao et al., 2014, p. 164)

The idea of addressing learner variability through course design facilitated by technological innovation holds potential to disrupt postsecondary education in ways that positively benefit learners (Christensen & Eyring, 2011; Rose & Meyer, 2002). Widening educational gaps due to the COVID-19 pandemic have made such concerns more urgent (Basham et al., 2020; Bruns et al., 2021; Hodges et al., 2020; Manly et al., 2021). A universal design approach has been posited to benefit students along the full spectrum of ability, addressing accessibility issues otherwise requiring special accommodation (CAST, 2014; McGuire, 2014; Scott et al., 2003; Silver et al., 1998). Educational design framed around options enabling widely accessible learning experiences can assist faculty practice, particularly combined with appropriately flexible technology (DeSilva et al., 2017; Tobin & Behling, 2018). While such practice holds potential to improve student outcomes like grades, well-designed empirical research corroborating this intuition remains surprisingly sparse (Kimball et al., 2016; Rao et al., 2014). The present study takes aim at this gap, investigating the effect of one aspect of Universal Design for Learning (UDL) on course outcomes. This quasi-experimental investigation aimed to better understand the efficacy of offering multiple means of representation, specifically investigating how use of multiple modalities in over 40 undergraduate online courses at a women-only institution impacted weekly grades and course grades for adult students.

For all courses studied, an instructional design team included content presented in multiple modalities (e.g., text, video, audio, interactive, or mixed), a choice aligning with UDL. Alternate paths for learning content were designed into an adaptive learning system and developed with Open Education Resources (OER) in different modalities (Navarro et al., 2016). As students worked through the material, if they showed signs of struggling as detected through a formative assessment at the end of each 20-minute learning activity, they were able to repeat that activity or engage alternate modalities. Such an approach, consistent with one of UDL's tenets, allowed investigation into the efficacy of UDL's guideline of offering multiple means of representing content (CAST, 2014). In this study, grades of students using multiple modalities when learning content in the adaptive learning system were compared to grades of students in a "business as usual" condition using a traditional learning management system (LMS).

The research question was: What are the effects of choosing more than one modality for learning course material throughout the week on subsequent weekly assignment grade outcomes (or throughout the course on final course grade)?

3.1 Theory and Literature Review

Informed by the universal design movement dealing with the physical built environment (Hamraie, 2017; Mace, 1991), UDL's foundation in cognitive neuroscience focuses on the brain's affective networks, recognition networks, and strategic networks (Rose et al., 2006). UDL posits students benefit from educational environments offering multiple means of engagement, multiple means of representation, and multiple means of action and expression, corresponding with these brain networks (Meyer et al., 2014). Designing from this neuroscience-informed perspective potentially aids students with and without disabilities taking courses online when faculty plan for learner variability to manifest in a technologically mediated environment (Rao et al., 2015). Instead of designing for the average and assuming students will accommodate themselves to that norm (perhaps with specialized assistance), a UDL approach deliberately designs in alternatives to achieve the learning objectives, facilitating a range of knowledge construction approaches.

Within UDL's principle of providing multiple means of representation, modality is emphasized in several practical checkpoints (CAST, 2014). Alternatives for auditory and visual information are recommended so that students can perceive content, along with ways for students to customize what they encounter so that it works for their perceptual and processing abilities. Providing options for perception has historically not been an emphasis of content presentation in many college classes (Davies et al., 2013). In addition to addressing functional impairments, this approach may support students who process and transfer information to long-term memory differently than their instructors.

The theoretical foundation for the present study combines the UDL framing around offering alternatives for perception with the idea that human brains perceive and process information through dual channels, one for visual information and another for auditory information (Mayer, 2001). Neural patterns have been found to differ for people with visual and verbal cognitive styles (Kraemer et al., 2009). Multimedia research experiments have shown memory benefits when combining information through these channels (Mayer, 2008), although such results have been found in other experimental research to depend on cognitive function, being different for people with dyslexia, for example (Beacham & Alty, 2006). Considering these ideas together, when students utilize options to learn content using their dual sensory brain channels in different but complementary ways, learning may be enhanced.

The universal design literature overall remains notable for its lack of effectiveness-oriented peer-reviewed research (Crevecoeur et al., 2014; Kimball et al., 2016). Multiple reviews spanning several decades have uncovered surprisingly few empirical articles pertaining to universal design in higher education given the face validity of the ideas (Capp, 2017; Mangiatordi & Serenelli, 2013; Orr & Hammig, 2009; Rao et al., 2014; Roberts et al., 2011). The most recent international review of higher education literature examined thirty empirical articles about UDL, none of which conducted experimental or quasi-experimental research investigating effectiveness for student learning (Cumming & Rose, 2021). As the majority of UDL research has focused on student and faculty perceptions, this review repeated an oft-heard call for studies of UDL's efficacy. As an example of difficulty encountered in past UDL research, one study found no difference in student scores before and after implementing UDL-based strategies in a supplemental biology course website, perhaps because providing lecture notes enabled students to skip lectures and engage in cramming rather than sustained learning activities (Bongey et al., 2010). Thus, implementation choices have been important when evaluating UDL effectiveness. The present study addressed such issues through careful attention to research design around a focused construct while recognizing possible alternative explanations will still need further research.

McGuire, Scott and Shaw (2006) argued for rigorously developing universal design theory by "refining and validating the [universal design] principles" iteratively by "testing of suppositions (i.e., [universal design] principles)" (p. 172). From a practical standpoint, faculty are frequently advised to begin implementation of UDL in tractable pieces (Tobin & Behling, 2018), which holds importance given that even faculty who are supportive of inclusive teaching practices, including providing multiple means of presentation, have not always reflected those values in their teaching (Lombardi et al., 2011). Knowing which aspects of UDL offer substantial benefits for student learning by themselves holds importance (Crevecoeur et al., 2014), as those would be appropriate places to encourage faculty to begin course redesign. This study investigates one such aspect: offering content through different modalities.

3.2 Data

The data come from a private, women-only institution in the Northeast, including 41 instances of 17 three-credit undergraduate online courses in a variety of disciplines, including business, economics, English, history, health, religion, psychology, and sociology. These courses were selected because they had sections taught in both treatment and control groups in a randomized control trial (RCT) that could be used as an instrument for the present study. 185 students took these courses during two Spring 2018 sessions, some taking courses in both sessions.[3] Student-level information from multiple campus systems was gathered from the institutional data warehouse, including data from the learning management system, an adaptive platform for course content and assessment, and the administrative student information system. Student information was anonymized prior to the researcher obtaining access, addressing privacy concerns.

[3] Often students will take one course per session since these are accelerated six-week courses.

The students participated in an RCT funded by a grant from the U.S. Department of Education's Fund for the Improvement of Postsecondary Education that investigated the effectiveness of the adaptive learning system in improving institutional outcomes including credit accumulation, year one to year two retention, and degree completion. Students in the treatment group had access to multiple modalities designed into the adaptive platform's course content, while students in the comparison group were placed into "business as usual" courses that had not been redesigned using the adaptive platform and did not deliberately offer content in multiple modalities. Although this RCT did not randomize my treatment of interest directly, the RCT's design provided an opportunity to instrument the variability in my treatment of interest, isolating the variability caused in the outcome for those who complied with treatment, as discussed further below. While data for this RCT experiment were collected across three years, the modality use information key to the present analysis was only collected during the last two sessions of the experiment in Spring 2018, and so data for this study were restricted to those two sessions.[4] Attrition after randomization was investigated (see Appendix B) and found to be low per national standards for quasi-experimental design studies of this type (What Works Clearinghouse, 2017, 2020b).

3.3 Variables

The outcomes included two summative measures of the student's understanding of the course material: a) mean grade on the week's assignments and quizzes, and b)

---

[4] During this time, the institution was still developing the adaptive courses through the grant, so fewer courses were available for this analysis than in the panel data analysis presented in the previous chapter which used data from the subsequent academic year.

overall course grade. See Appendix B for additional information about these and other variables, including sensitivity analyses performed to investigate how operationalization choices might have affected conclusions. For both week and course outcomes, the primary analysis dropped zero scores assuming many students earning zero credit do so not because they do not understand the material but because other life factors intervene. This potentially de-emphasizes student learning within the outcome, thus increasing noise when attempting to detect the treatment's impact on student learning. An alternate, policy-relevant practical impact was estimated by including zeros, anticipating a potentially diluted effect.

Restricting week-level analysis to weeks two through six allowed incorporating a factor addressing the potential confounding effect of motivation to achieve a high grade. Indicators of latent grade motivation from week one were used in a principal components factor analysis to generate a standardized grade motivation score variable (reliability $\alpha = 0.76$; see Appendix B). Equivalent indicators of grade motivation from prior courses were not available for the course level analysis.

When analyzing weekly grade, treatment was operationalized as any use of more than one modality on seven or more activities during that week. A threshold of seven was chosen since many courses were designed with a small number of activities that included a second modality by default, and thus would have been encountered by all students. Seven was the weekly median during weeks two through six. To check the sensitivity of results to this threshold choice, alternate values were analyzed from the distribution of the minimum number of times multiple modalities were used, including one (25^{th} percentile), five (75^{th} percentile), and nine uses (95^{th} percentile).

When analyzing overall course grade, treatment involved any use of more than one modality during the course. Following a similar rationale, the median of 38 was chosen as the threshold. Sensitivity analyses across the distribution of the minimum number of multiple modality uses included one, 23 (25^{th} percentile), 34 (75^{th} percentile), and 68 uses (95^{th} percentile) during the course.

In addition to indicators of the course session and RCT cohort, since the RCT utilized a blocked design by cohort, several measures of prior education and demographic characteristics were included to increase the precision of the treatment effect estimates (see Appendix B). Most variables included in the primary analysis model for the week and course level outcomes had no missing data. Seven covariates had missing data, with a maximum of 4.5% for the survey-based technical competency score. While this small amount of missing data was not considered a major threat to validity, results via multiple imputation (M=100 imputations via Stata 16 *mi impute chained*) were generated and compared to listwise deleted results. Multiple imputation facilitated an alternate analysis including an indicator of first-generation college student status which had 80.5% missing data. Despite the presence of so much missing data, this variable was deemed of interest since it is a component of socioeconomic status.

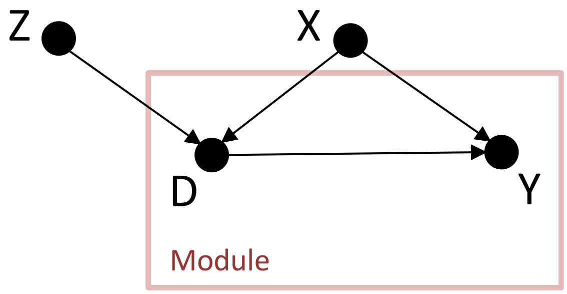
Two baseline equivalence measures were checked per What Works Clearinghouse (WWC; 2017) guidance for quasi-experimental analyses. Academic achievement, proxied by the number of credits transferred in upon entry to the institution, required adjustment per WWC guidelines (week dataset baseline effect size, $|ES_{Base}| = 0.03$; course $|ES_{Base}| = 0.15$). Socioeconomic status, proxied by Pell grant status, satisfied WWC standards for baseline equivalence for the analysis of weekly grade with statistical adjustment (week

$|ES_{Base}| = 0.21$), but the baseline Pell difference for the analysis of course grade was higher than recommended (course $|ES_{Base}| = 0.65$; see Appendix B), warranting caution in interpretation. Both baseline variables were included in all models.

3.4 Methods

Initially, means and standard errors were investigated for each variable at both week and course level. Tabulations and counts checked distributions across treatment and comparison. After obtaining basic understanding of the data, an instrumental variable analysis allowed isolation of variation due to treatment through two-stage least squares regression (Angrist & Pischke, 2009; Morgan & Winship, 2015; Murnane & Willett, 2010). The first stage regressed the treatment (use of multiple modalities) on the instrument (assignment to treatment/control groups in the RCT). The second stage regressed the outcome on predicted treatment values from the first stage. Analysis focused on outcome variation induced by the treatment as seen between groups defined by the instrument, thus partitioning variance in a way revealing the treatment's effect on the outcome for individuals similar to those in the sample (i.e., local average treatment effect (LATE) or complier average treatment effect (CATE)). Given that knowledge of the full distribution of student learning was limited based on grading of failure and As, a tobit approach accurately represented the censored nature of the dependent variable (Long, 1997). An upper limit was used when zeros on the dependent variable were dropped from the sample (week UL=1 and course UL=4). Both upper and lower limits were used otherwise (LL=0 for both). Figure 3.1 presents the model.

Figure 3.1 Model for Instrumental Variable Analysis



Where:

Z = Participation in RCT treatment

D = Use of multiple modalities

X = Exogenous control variables

(e.g., (X_1, X_2,… X_k)

Y = Grade

Note: D and Y represent the appropriate time frame for the week and course models.

RCT treatment gave students the opportunity to go through a particular activity's material a second time with a different modality because only treatment students had access to multiple content modalities within the adaptive system. The control group's "business as usual" scenario did not involve the adaptive system. This meant the instrument (Z) would induce variability in the treatment (D), meeting the correlation requirement for an instrumental variable analysis to have relevance.

Furthermore, participation in RCT treatment is reasonably assumed not related to how well the student learned material as demonstrated by their grade, meeting the non-correlation requirement (i.e., exclusion restriction). Specifically, the only causal path between instrument (Z) and outcome (Y) would be through the treatment (D). Here, all exogenous control variables (X_k) were pre-treatment measures, so cannot be either a mediator or a common effect of treatment and outcome. If the adaptive learning platform might affect grades other than via a path through treatment, that would affect this study's validity. However, the RCT found no impact of the adaptive system on grade-related outcomes such as credits earned, supporting the claim that exclusion was not violated.

Even so, it is worth considering whether alternate mechanisms might exist within the adaptive learning system providing another path between instrument and outcome. Such paths might be due to a) the system's adaptation to the students, or b) the quality of OER material or course design, for example. For this analysis, I made the plausible assumption that such factors did not interfere. For a), system administrators expected minimal adaptive changes when this study was conducted since the system was new and such adjustments take time to emerge from data. For b), the same institutional instructional design team developed both business-as-usual and treatment courses, presumably minimizing such issues to the extent possible. Additionally, the design team attempted to ensure a similar standard of quality for the courses utilizing OER material by designing them to meet the Quality Matters standards (Robinson & Wizer, 2016). This offered assurance of similar quality across courses, even though the institution did not pursue the full external review process for QM certification. For these reasons, I make the same assumption as the RCT that quality issues were not confounders, although quality concerns could be investigated in future research. Additionally, another analysis (reported in chapter two) using within-student panel data from the same adaptive learning system found an effect of using multiple modalities on formative within-week learning outcomes. Taken together, these considerations bolster confidence that the adaptive learning system itself did not act as a confounder for the present study and thus that the results have internal validity. Accepting this assumption, the instrument (Z) met the two main relevance requirements of correlation with treatment (D) but not outcome (Y).

Considering additional required assumptions (Porter, 2012), RCT assignment was random by design, so no similarity-to-randomness argument is needed. Additionally, by the Stable Unit Treatment Value Assumption (SUTVA), an individual's treatment status and outcome must not have been affected by anyone else's treatment. While this often poses problems in education research, RCT participation was randomly assigned, so SUTVA applies.

Lastly, assuming monotonicity meant change in RCT group assignment either did not affect use of a second modality or had the same effect across students. RCT assignment determined treatment access, which was only available through treatment courses using the adaptive system. Students could thus not switch groups mid-course. While no control students ended up in a treatment course, treatment students could defy their assignment. Only four students assigned to treatment took a control course instead, and one took both a treatment and a control course. Thus, there was high fidelity to initial RCT group assignment. However, 16% of students assigned to RCT treatment never took the opportunity to use multiple modalities. Given this incomplete correspondence between RCT treatment assignment and the current study's treatment assignment, an instrumental variables analysis is appropriate, and results should be interpreted as complier average treatment effects. Overall, the necessary assumptions were deemed met with this instrument.

Augmenting the primary week and course analyses, additional sensitivity analyses were conducted. These checked sensitivity of results to the choice of dependent variable, number of covariates, assumption of homoskedasticity, type of inference model, treatment threshold as discussed above, whether zero grades were included, and whether courses without a treatment-control match during the two subterms studied were included.

3.5 Limitations

Considering issues associated with internal validity, while the sample size is low overall, the study design allows relatively low-variance estimation of the effect of multiple modalities on the weekly assignment grade. Although the lack of sufficient baseline equivalence on Pell status for the small course level sample calls for interpretive caution, the results' suggestive positive nature warrants further research in this area.

Notably, only one student with a known disability course accommodation was in the analysis sample. This number was surprising and far lower than the rate that might be typically expected, particularly for non-traditional students, potentially affecting the study's external validity. Nationally, 19% of undergraduates have recognized disabilities, a rate that rises for older students (NCES, 2021). Anecdotal evidence from faculty and staff indicated students self-identified as having disabilities, so it is unclear whether the sample's low number reflected a reticence of students with disabilities to participate in the RCT. Given that some students may be unprepared generally to request accommodations in college, they may have wondered if an experimental setting might present more barriers rather than fewer (Marshak et al., 2010). Alternately, while some may have received accommodations earlier in school, given their older average age and often underprivileged status, past diagnoses may have needed potentially costly and time consuming updating in order for current accommodation eligibility, interfering with some students seeking otherwise relevant accommodations (Fox et al., 2021). Additionally, students may have chosen not to request accommodations for fear of stigma or other reasons, or they may not have known how to access services for which they would have been eligible (Gierdowski, 2021; Newman & Madaus, 2015). On the other hand, it may also be the case that some students with disabilities did not feel the need for accommodations given the intentional attention to course design approaches potentially mitigating accessibility issues. Although unverified, this seems plausible, and if true would indicate additional support for the UDL approach overall. Whatever the reason, the lack of students known to have requested accommodations despite the expected presence of students with disabilities should be kept in mind when considering the results.

3.6 Results

Students' mean weekly score on assignments and quizzes (not including zeros) was 88%. Mean course grade was 3.3 (B+). Close to 20% received treatment in both analyses. Appendix B shares additional descriptive results.

Given the RCT data, understanding the influence of students leaving the study gave insight into instrument strength. The 3-year RCT had low attrition, and about 30% of RCT participants were analyzed here. Low group differential attrition (0.019 percentage points) and low overall attrition (2.3%) across this study's two subterms mean it likewise has low attrition overall (What Works Clearinghouse, 2020b), supporting the instrument's strength.

Instrumental variables tobit analyses of weekly grades (Table 3.1 top panel) and course grades (bottom panel) are presented. Most impact estimates were statistically significant at $p<0.05$. The rest carried significance at 0.1, which may be reasonable given the small sample. Appendix B presents sensitivity analyses and full regression tables.

Table 3.1 Hedges' *g* Standardized Effect Sizes (ES) for use of Multiple Modalities

Analysis	N_C	$\overline{Y}_C$	N_T	$\overline{Y}_T$	Impact Est.	SE	p	Std. ES	I-Index	% Bias
<u>Week</u>										
Assign./quiz -with 0 & first gen.	878	0.794	219	0.943	0.150	0.065	0.020	0.539	+20.5	15.4
Assign./quiz -with 0	878	0.794	219	0.896	0.102	0.054	0.061	0.366	+14.3	[6.47]
Assign. Only-no 0	674	0.900	191	0.991	0.091	0.024	<0.000	0.903	+31.7	48.37
Quiz only-no 0	270	0.862	41	1.207	0.345	0.115	0.003	2.355	+49.1	34.33
Assign./quiz-no 0 & first gen.	795	0.876	207	1.047	0.170	0.038	<0.000	1.283	+40.0	56.1
Assign./quiz-no 0	795	0.876	207	1.015	0.139	0.032	<0.000	1.045	+36.2	54.1
<u>Course</u>										
Grade-with 0 & first gen.	239	2.995	55	3.992	0.998	0.541	0.065	0.821	+29.4	8.9
Grade-with 0	239	2.995	55	3.888	0.893	0.493	0.070	0.735	+26.9	8.9
Grade-no 0 & first gen.	215	3.329	52	4.089	0.760	0.382	0.047	0.979	+33.6	17.0
Grade-no 0	215	3.329	52	4.048	0.719	0.344	0.037	0.926	+32.3	21.4

Notes: Impact Est.=impact estimate of the treatment effect; Std. ES=WWC percentile standardized effect size; I-Index=WWC improvement index; % bias=percent bias that would be necessary to invalidate the inference at α=0.05 for week and α=0.10 for course, and results in [brackets] indicate % bias needed to sustain the inference at that α level. Assign/quiz=outcome combining mean of weekly assignment and quiz grades. The full set of covariates for week and course level analyses were used for all models except for first-generation status as noted above (see Appendix B). Missing data handled by multiple imputation.

For the weekly grade, assumed treatment endogeneity was supported through an endogeneity test (Durbin χ^2 = 13.140, p = 0.003). As expected for an RCT-based instrument, testing after the first stage (*ivregress*; *estat firststage*) indicated a strong instrument: F(1, 920) = 161.857 under listwise deletion; median F under multiple imputation was 169.132 with minimum 166.280. Thus, F exceeded the recommended 104.7 for single instrument studies (Lee et al., 2020). Testing for a weak instrument (*ivtobit*; *weakiv*) by comparing the Anderson-Rubin and Wald test statistics also demonstrated a strong instrument; the confidence intervals of the two tests were close across all imputations, meaning the instrument was partitioning most available variance successfully (Finlay & Magnusson, 2009). As will be discussed further later, the instrument strength was less for the course grade analysis.

Effects of use of multiple modalities ranged from Hedges g = 0.37 standard deviations to 2.36 SD depending on the model and sample. Specifically, using more than one modality seven or more times across a given week was shown to have a strong positive standardized effect size of 1.05 SD on that week's mean assignment and quiz grade. This represents a large effect for social science, and particularly large for intervention research in education (Kraft, 2020; Lipsey, 1990). When incorporating first-generation status as a covariate, the magnitude of the effect detected was 1.28 SD. This higher estimate was likely a more valid treatment effect estimate. Despite the large amount of missing data for the first-generation variable, multiple imputation utilized all available data, incorporating this component of socioeconomic status, while accounting appropriately for uncertainty caused by missingness. Alternately, incorporating zero grades reduced the effect's magnitude to medium strength (0.54 including first-generation status or 0.37 without it). Additional sensitivity analysis indicated that a moderate effect existed starting with just a single use of multiple modalities during the week (see Appendix B).

Course level effects ranged from Hedges g = 0.74 to 0.98 SD. Specifically, using more than one modality at least 38 times throughout the course, corresponding to over 6 times a week, resulted in a strong effect of 0.93 SD on course grade. As with week, and as might be expected, including zero grades in analysis reduced the detectable effect. While dropping zeros entails a form of selection bias, it also means potentially confounding factors that cause people to fail because they stopped educational activities would not influence estimation of how the treatment helps students learn course content.

3.7 Discussion

Interested in addressing the lack of effectiveness-oriented UDL research (Roberts et al., 2011), I probed causal connections by taking advantage of the randomization process of an RCT related to the availability of the treatment of interest in an instrumental variables analysis. Isolating one aspect of the UDL framework for investigation (Crevecoeur et al., 2014), I found that using different modalities when learning content had a statistically significant and meaningfully large positive effect on learning as demonstrated through content-related weekly graded activities including assignments and quizzes. These findings are consistent with prior research on adaptive learning where material was available in different formats (Mustafa & Sharif, 2011). Though that study did not focus on modality use, present results suggest this feature possibly provided a mechanism for these prior positive results of adaptation including media type.

The main week-level finding demonstrated a large effect size over one standard deviation for students who did not earn zero on the material for the week and a moderate effect when including zeros. An improvement index of +36 for this finding means this is like a student moving from the 50^{th} to 86^{th} percentile (What Works Clearinghouse, 2020a). To gauge confidence in this effect, it would have taken a confounding variable causing 54% bias in the estimate to invalidate this inference at the $\alpha = 0.05$ level (Frank et al., 2013). Such an omitted variable would have to have had a stronger correlation with both outcome and treatment (0.281) than most correlations between covariates used in this study. It would be similar to the negative correlation between being White and Pell eligible, but less strong than either GPA or grade motivation and the outcome (correlations over 0.4). These examples offer intuition about the kind of missing confounder necessary to explain away the effect detected. While it would be relevant to investigate the possibility of such an unobserved factor, policy-relevant factors with such potential seem few. The takeaway: this result should inform practice.

To unpack this weekly grade result given its strength, a basic decomposition investigated assignments and quizzes separately (see Appendix B). While both individual effects were strong, the effect of using multiple modalities when learning content on quiz performance was well over two SD. This kind of two-sigma effect is atypical in educational research, where researchers often deal with interventions demonstrating small-to-moderate gains due to the complex nature of learning (Hattie, 2015). While interpreting effect sizes across studies carries complications, and issues such as cost per student and scalability warrant attention as well, recent guidance suggests 0.2 might be considered a large effect for standardized achievement outcomes (Kraft, 2020), whereas for treatment effectiveness research, 0.9 has been considered large (Lipsey, 1990). The effect seen here is much larger, equivalent to a student moving from the 50^{th} to the 99^{th} percentile on their quiz score, a very dramatic and educationally important improvement if it proves to be replicable in other contexts. This result suggests benefits for quiz performance and content mastery deserve further attention, though caution against overgeneralization is warranted based on a single study of women with few known disabilities which did not focus on this quiz-based effect (Shadish et al., 2002). Even given these caveats, results point to an intriguing area for future intervention research.

The longer-term course grade effect was lower than the week effect by a surprisingly small amount given the lengthier time and the other aspects of learning beyond content acquisition that course grades also reflect. In contrast to weekly grade's effect just over one SD, the effect on course grade was just under one SD. Over 21% of this estimate would have to have been due to bias to invalidate this inference (considering a threshold of $\alpha = 0.10$ since the sample is small), so the result is robust to anything but a moderately strong confounder. The effect corresponds to a student moving from the 50^{th} to the 84^{th} percentile, which is also a very policy-relevant magnitude.

Interpretive caution holds relevance here, however. Course level data did not exhibit desired Pell grant status baseline equivalence, meaning socioeconomic status may influence results. Course results also did not meet current standards for the desired strength of the first stage ($F = 33$), although the instrument was stronger than the threshold (10) previously considered acceptable in many published instrumental variable studies. Coupled with results on the verge of acceptable statistical significance (i.e., significant at the 0.10 level), readers are advised to view the results of the course level analysis optimistically but with informed caution. Given the strong effect on weekly assignment and quiz grades as well as the results of other research (in chapter two) showing a positive effect when students are engaged in learning activities, the course-level results suggest an effect on the edge of detection given the sample and design. Research confirming this course grade effect remains warranted.

3.8 Implications

By confirming broad benefits for a key aspect of UDL's focus on the brain's recognition network, this study extends empirical higher education UDL literature (Rao et al., 2014) and concretely affirms the idea that UDL elements benefit students across the ability spectrum (Tobin & Behling, 2018). Initial results looking at different treatment levels found increased impact as usage of multiple modalities increased at both week and course levels (see Appendix B). This suggests dosage conditions may influence the magnitude of the effect and should be more formally investigated, particularly given that past research has shown that not all modalities work effectively for all students (Beacham & Alty, 2006). While the strength of results in other settings remains speculative, future experimental or quasi-experimental research could investigate conditions which foster this benefit and the types of students who benefit most (Roberts et al., 2011). The extent that results from an online adaptive setting translate to other more traditional on-campus conditions remains unknown without further research. However, UDL's foundations suggest relevance in a variety of settings (Meyer et al., 2014) and UDL principles have been successfully implemented in non-online settings (Davies et al., 2013). Likewise, implementation quality's importance remains unknown and may have relevance for external validity. Importantly, this study was conducted at a women's institution, and so demonstrates an effect for women, but research is needed including men too.

Additional research with students known to have disabilities would be an important extension, particularly to confirm the effect's magnitude. Such investigation would ideally be aware of the fluidity of disability while looking at the magnitude for different disability types (Fox et al., 2021). I was unaware of the small number of students with disclosed disabilities when this study was conceived. Presenting an unexpected and notable limitation, the study was conducted anyway given that more students with disabilities were assumed present, if unidentified (Marshak et al., 2010), and UDL's tenets were assumed relevant throughout the population. However, future research specifically investigating the effect for students with known disabilities and/or elucidating the extent to which this institution's students self-identify as having a disability would be warranted.

Possible mechanisms through which the effect operates should be investigated as well, particularly factors with potential to explain away the effect. As mentioned, motivation to achieve a high grade is one such mechanism, and a crude measure of this was included in the week level analysis. Attempting to identify an improved measure of this would be relevant for future research. The present study highlights the need to investigate how to construct a strong measure from learning management and adaptive system log data to better understand how such grade motivation may be reasonably accounted for with observed data in future research.

With a solid effect detected at the week level and a reasonable effect likely at the course level, the implications for course design and faculty development are clear. Students benefit when institutions and individuals devote time and energy to aspects of UDL encouraging options for perception. Current results should encourage faculty to consider how to add options for learning content in a variety of modalities. Prior experimental research has shown faculty make improvements in offering content through multiple means of representation that are noticeable to their students with only five hours of UDL training (Davies et al., 2013). Training has also been shown to assist faculty in utilizing OER materials for this purpose rather than creating such material themselves (Navarro et al., 2016). Thus, the bar for making improvements in this area is not unreasonably high. Present results should encourage institutions to fund efforts in this direction, supporting cultural change toward a universal design mindset (Silver et al., 1998). They should also bolster arguments that funders should invest in high quality research supporting empirical investigation of UDL guidelines to clarify the extent of the benefit of UDL recommended practices, guiding future resource allocation (McGuire et al., 2006). To date, much of the literature about UDL's components, including offering multiple means of presentation, has entailed plausible arguments along logical rationales and perception-based support from students and faculty (Lombardi et al., 2011; Rose & Meyer, 2002). The present study complements this literature (McGuire, 2014). The empirical basis upon which faculty development programs can base recommendations and upon which faculty can decide to devote their time is growing but should still be expanded.

The current study reinforces the idea that good practice supporting students with disabilities is good practice to support all students (Tobin & Behling, 2018). Particularly given that college students with disabilities often choose to go unidentified (Gierdowski, 2021), evidence of the widespread effectiveness of providing multiple means of representing content across a variety of fields of undergraduate study should be a wake-up call to practitioners and administrators throughout postsecondary education: become more widely informed about UDL-based approaches and participate in action-oriented research to better understand the effectiveness of such practices. It would be beneficial for the scholarly community to widely embrace this orientation within higher education and the disciplines.

3.9 Conclusion

Overall, this study extends prior research by deepening understanding of UDL's supposition that providing multiple means of representation through options for perception benefits the full range of student abilities. Importantly, models at the week level detected a noteworthy, statistically significant positive effect of using multiple modalities on grades with a moderate to very strong effect size, and these results likely extend to the course level. The ability to learn content by utilizing different modalities may be particularly important and helpful in student populations where students with disabilities may not be formally identified in high numbers. Results support straightforward practical action by faculty and course designers in line with UDL principles resulting in improved grades. Such action holds promise for utilizing course design in addressing systemic inequality in higher education outcomes for adult women students in particular.

CHAPTER 4

ANALYSIS COMBINING MODALITIES AND TUTORING

> When the goal is to design for "everyone," I ask, who counts as everyone and how do designers know? (Hamraie, 2017, p. xiv)

> Faculty development [in the future] will be linked to the capacity of the field to engage in more research about best practices that influence student learning, and to work programmatically from a research base on learning and teaching. (Austin & Sorcinelli, 2013, p. 94)

Adaptive learning systems have become well positioned to assist in effectively teaching learners with wide variation in perceptual and processing ability, but their effectiveness remains tied to the material designed into them. Given that the college student mix increasingly includes students with disabilities (Kimball et al., 2016), an approach such as Universal Design for Learning (UDL) can help make course material accessible in ways that are beneficial for all students (Rose & Meyer, 2002; Tobin & Behling, 2018). However, UDL and adaptive features can each be time consuming to implement, often occurring through iterative (re)design cycles over multiple semesters. Thus, educators face the task of helping struggling students in courses not (yet) fully universally designed. The present study illustrates a data-informed approach to identifying predicted points when recommending tutoring may be beneficial for students.

Higher education must go beyond traditional responses of providing accommodations for students with disabilities given that only about 35% choose to tell their institution about their disability (Newman & Madaus, 2015). In this chapter, I build on the results from chapters two and three that demonstrated a positive impact on student learning of one aspect of UDL: providing content via multiple modalities (i.e., text, video, audio, interactive, or mixed content). Practices such as UDL comprehensively encourage faculty and course designers to expand the idea of good teaching to include making courses accessible to students who have a wide range of capabilities, including disabilities. These practices provide essential access to some students with functional impairments (whether or not those students have been formally diagnosed as having a disability that requires accommodation) while benefitting all students. UDL is not a panacea, however, and while it provides guidance, full implementation of universal design often remains aspirational. In practice, this means faculty can expect students will run into challenging areas where the course design may not fully meet their needs for learning that particular content. This leads to the question: What can we do to help students when our efforts at universal design fall short or are in process?

This proof-of-concept study illustrates a learning analytics-informed approach that combines formative data traces from tutoring and adaptive activity to build a prescriptive analytics model that would identify points during a course where recommending tutoring may be warranted to support students. For this study, I assume practices such as tutoring, use of multiple modalities within an adaptive system, learning activity repetition, and time on task all benefit learning. Within that framing, I illustrate how analyzing within-course data may help make recommendations to students. Institutions increasingly use predictive analytics to inform feedback to students, often through vendor-driven systems that may involve proprietary algorithms with unknown characteristics. Guidance for educators regarding evaluating proprietary predictive analytics systems has begun to explain use of Bayesian networks, which facilitate modeling of causal relationships and evaluation of hypothetical scenarios, but such efforts are not yet well known (How & Hung, 2019). Prescriptive analytics extends the predictive analytics approach to include modeling and simulating alternative possibilities to investigate optimal decisions while accounting for uncertainty (Frazzetto et al., 2019). Analytics approaches have been receiving increasing attention by academic and institutional researchers, informing the individual-centered approach investigated here (Dawson et al., 2019; Gagliardi et al., 2018). While making use of predictive data analytics has become increasingly important within higher education, “analytics for the purposes of improving student learning outcomes… remain sparsely used in higher education due to a lack of vision, strategy, planning, and capacity” (Gagliardi, 2018). We have an opportunity to increase understanding about ways predictive and prescriptive analytics could be implemented and extended, particularly at the course level, which has not been the focus of most previous analytics efforts (Schwartz et al., 2018). Institutional mechanisms to make use of predictive capability often remain nascent, of limited scope, or in early stages of development, reflecting the state of the analytics field generally (Dawson et al., 2014). The novel analytics approach illustrated here aims to let students know when tutoring might be beneficial, augmenting whatever assistance they already receive from universally designed course elements.

This projected prescriptive method presents a different approach to analysis than a typical treatment effect study like those presented in the previous two chapters. Standard deductive quantitative research involves identifying an open research question along with data that can address it, and then statistical analysis methods estimate an average treatment effect result. In contrast, this study takes a prescriptive analytics approach using different datasets for model construction and predictive analysis, and aims to identify simulated predicted treatment effects for individual students taking a particular class rather than an average treatment effect across multiple students. Such individual predictions may be used to inform interventions with those students.

This prescriptive method combines a machine learning approach with the idea of simulating worlds for prediction to determine the best course of action to recommend to students. In machine learning, one dataset trains a model that can then be used to predict what would happen with other data. Applying this idea to simulated worlds allows exploration of what would happen in a hypothetical experiment on a simulated educational system where the same student simultaneously received different interventions. Treatment and comparison conditions can each be specified and evaluated in a different simulated world and then compared for a given student to determine the anticipated optimal path.

This prescriptive approach starts by developing nuanced modeling capability for a particular course, using existing data from prior semesters of the course. A graph of the Bayesian network (see Appendix C for more explanation about Bayesian networks) specifies the assumed causal relationships between each variable (i.e., node) in the educational system. This provides the basis for modeling each node as the outcome of its own structural model. Parameters are determined and stored for each node's predictors (i.e., its parent variables) in the Bayesian network using existing data (i.e., the training data in this example). During subsequent course offerings (i.e., the testing data in this example), predictions can be projected into the nodes for future time points using these models, providing predicted data for subsequent calculations.

The Bayesian network structure and associated assumptions about the causal mechanisms representing the data generating process for each node facilitate such modeling and predictions. The network illustrates the posited causal structure between nodes in the educational system, which in turn determines the appropriate models for different nodes. Once an appropriate node structure for a course has been determined, the associated parameters necessary for predicting the values of those nodes are learned and stored. While the example then predicts values for students who have already finished the course, this approach could be implemented dynamically as a course is offered, making real time recommendations to students about seeking support.

At any point in the course, two potential worlds can be simulated for a given student. In one simulated world, the student would receive tutoring and in the other simulated world, they would not. In each simulated world, the outcome of interest would be predicted. In this example, the outcome is that week's assignment grade, but we could make predictions about any outcome where appropriate training data could be collected. We could investigate an outcome occurring at any future time point in the course as long as it has been incorporated into the Bayesian network. Given data collected up through the time point chosen for analysis, values of future nodes can be predicted from the learned models' parameters and these predicted values propagated through the network and used to calculate values for subsequent nodes. Running the simulation multiple times (e.g., 500) allows a set of projected predictions to be accumulated in each potential world, representing the distribution of the predicted outcome in that world for that student and reflecting the uncertainty associated with the simulation. These outcome predictions in the two worlds can then be compared with a statistical test to determine the expected benefit should the student choose to receive tutoring at the specified point in the course.

The example presented features an introductory undergraduate English course where system logs captured students' learning actions in both an adaptive learning system and a traditional learning management system, along with information about the online tutoring they received. While UDL includes numerous guidelines for practice, the focal element here included presenting options for perception by offering content through multiple modalities including text, video, audio, interactive, or mixed content presentation. As the research in chapters two and three has shown, use of multiple modalities can help students to learn. Knowing this, I posit that identifying patterns in students' use of multiple modalities combined with their utilization of tutoring should offer insight into where students struggle to learn the course material, and thus where future students may benefit from seeking additional support.

As a proof-of-concept, the investigation conducted illustrates the type of analysis that could be done to offer predictive suggestions based on past data about modality switches and tutoring. As such, the illustration presents preliminary results with a more argumentative than analytical focus. Through this example I argue that institutions should be thinking creatively and expansively about how to use the wealth of student learning data now collected through online systems that have become increasingly available in aggregated institutional data warehouses in the service of further assisting struggling students. The example presents an English course taught during one academic year at a women-only institution that collects such data. It provides a first look at the kind of analysis that could be expanded to other circumstances with different characteristics but similar technological capacity to merge data across campus and vendor systems. My approach combines the idea of "closing the loop" to students from the learning analytics field (Clow, 2012; Mattingly et al., 2012) with the idea that it ought to be possible to utilize a network of prior and current data to make dynamic predictions about key intervention points in a course with a goal of improving student course success.

The research question for this study asked: How can information about modality switches and tutoring be used to predict later learning module success in one week of an introductory English course? I hypothesize that combining modality switches and tutoring will be predicted to benefit some students, showing potential to inform tutoring support recommendations.

4.1 Theory and Literature Review

UDL practice builds on the educational implications of natural learning variations. It has been explored for several decades by a community of scholars and practitioners interested in universally designing educational experiences (Burgstahler, 2015; Higbee & Goff, 2008; Meyer et al., 2014; Silver et al., 1998). From a social justice standpoint, supporting student learning across the full spectrum of ability takes a step toward inclusive practice equitable for all (Levey et al., 2021). Recently, the worldwide disruption to educational systems resulting from the COVID-19 pandemic exacerbated practical challenges faced by educators committed to sharing and enacting such principles (Basham et al., 2020; Bruns et al., 2021; Hodges et al., 2020). Often enacted from a backwards design standpoint in which faculty and other course developers start from the learning objectives they want students to grasp, the UDL framework facilitates intentional course design that encourages faculty to consider alternate means to achieve equivalent learning ends.

Support for multiple approaches to learning advocated by UDL ideally would become integrated into all aspects of regular instruction and would sufficiently address the learning needs of all students in a course. However, faculty and support professionals must recognize and plan for what happens when existing design efforts and resources prove insufficient to meet that goal (Burgstahler & Cory, 2008). While demonstrably more could be done to improve educational practice toward achieving fully individualized support without singling out particular students, fully universally designed instruction remains a high aspiration that can be difficult to achieve in practice (Evans et al., 2017). Realities of existing courses and historical course development practices mean that courses may need multiple partial revisions or redevelopments. Given that many institutions remain far from making all content accessible in a universal fashion, even intentional implementation of UDL principles may not fully meet the needs of all students. To address this, supplemental individualized support may also be provided, such as through tutoring or accommodations, in instances where the options available do not (yet) encompass a wide enough array to meet the needs of particular students.

Tutoring can be an important augmentation to UDL-based instruction when students demonstrate individualized learning needs not sufficiently addressed through existing course design. This is consistent with Edyburn's concern that, "we need to renew our commitment to equitably serving all students in the event that our UDL efforts fall short" (Edyburn, 2010, p. 40). Tutors have individualized instruction of content, customizing content presentation to an even greater degree than otherwise currently possible, even with adaptive learning technology, as used in the present study. Given extensive research evidence showing positive effects of tutoring prior to college (Gordon et al., 2007), and demonstrated benefits for students both in college course outcomes (Abrams & Jernigan, 1984) as well as longer-term college persistence (Laskey & Hetzel, 2011), benefits of tutoring for student learning were expected. I posited that combining content through multiple modalities with tutoring may have provided the additional assistance that struggling students needed to be successful if the material was not presented in ways that addressed sufficient learning variability. That is, I viewed tutoring as augmenting the design of the course in ways that held potential to address gaps in the universality of content presentation, since a human tutor would be able to explain material in a highly interactive and personalized way that went beyond other ways of presenting the content.

In general, improving college student learning, as well as subsequent retention and success, including for low-income students and those with disabilities, constitutes a well-acknowledged challenge in higher education (DaDeppo, 2009; Kuh et al., 2007; Tinto, 2006; Wessel et al., 2009). Refining instructional design offers one avenue with potential to improve achievement, particularly for disadvantaged groups (Edyburn, 2010; Tobin & Behling, 2018). Inspired by the desire to better support people with disabilities and grounded in cognitive science, UDL challenges higher education faculty and staff to design students' learning experiences intentionally including multiple means of engagement, representation, and action and expression (Burgstahler, 2015). UDL's empowering frame arises from considering disability as a social construction (rather than a medical diagnosis; Jones, 1996) and views all individuals as capable learners given a supportive environment that does not disable their capacity.

In recent years, institutions have welcomed students with an increasing range of student disabilities, diversity in neurological functioning (i.e., neurodiversity), forms of engagement, and cognitive approach, with the rate of known disabilities among college students rising from 11% in 2003-04 to over 19% in 2015-16 (Snyder et al., 2019; Snyder & Dillow, 2013). Inclusive educational design for content presentation improves the success potential for students with disabilities. Such design can benefit students whether or not institutions have identified them as having a disability, as individuals do not necessarily fall into neatly diagnosed cognitive and affective bins that either constrain or empower them in all learning situations. Given that many students with disabilities choose not to identify themselves as such to their postsecondary institution, faculty will frequently not know who among their students has a disability (Newman & Madaus, 2015). Given the widening range of abilities, needs, and ways of knowing that aspiring students bring to higher education given the shift from mass toward universal higher education across the past century (Trow & Burrage, 2010), it becomes imperative for educators to design courses while viewing a broad range of abilities and experiences as normally expected. Doing so holds the potential to foster greater success for typically underserved groups such as students with disabilities and nontraditional-age students. Given that online students frequently come from traditionally underrepresented populations (Barnard-Brak et al., 2012; Wladis et al., 2015), online education offers a salient environment for investigating alternative content presentation, as in this study.

Issues of learning accessibility and variation have found expression in several related theoretical frameworks, including UDL (CAST, 2014), Universal Instructional Design (UID; Silver et al., 1998) and Universal Design for Instruction (UDI; Scott et al., 2003), among others (McGuire, 2014). UDL has been intimately connected to educational technology support for learning because technology facilitates complying with UDL principles, as in the present study (Mangiatordi & Serenelli, 2013). Despite enough interest to generate a variety of alternatives, such universal design frameworks still struggle to gain acceptance in academic culture (Archambault, 2016) and remain understudied in postsecondary education (Rao et al., 2014). Given widespread interest in universal design, including implementation guidelines and many arguments calling for its adoption (Burgstahler, 2015; CAST, 2014), the lack of research is surprising. Despite plentiful general course design guidance for faculty, including for online education (e.g., Chickering & Gamson, 1987; Dell et al., 2015; Ko & Rossen, 2017; McKeachie & Svinicki, 2010), universal design practices are still working their way into this literature. Development and incorporation of universal design practices based in high-quality, empirically based research remains beneficial to pursue (Cumming & Rose, 2021). The present study extended one aspect of the research inspired by universal design work.

When conducting research about UDL, it remains unusual to study student learning outcomes directly. With regard to the UDL outcomes typically studied, even in higher education, subjective perceptions constitute a heavy research focus, typically of faculty (e.g., Ben-Moshe et al., 2005; Higbee & Goff, 2008; Izzo et al., 2008; Lombardi & Murray, 2011) and occasionally of students (Higbee et al., 2008) or employees (Parker et al., 2003). In one relevant study, graduate students were found to both recognize the benefits of having content provided in multiple formats, and to a lesser extent, reported using them (Fidalgo & Thormann, 2017). Additionally, Webb and Hoover (2015) studied application of UDL principles to library instruction, specifically intending to address multiple learning styles through multiple means of representation, and did usability testing on the resulting library tutorial in order to identify improvements. One of very few studies expressly investigating multiple content representations, it aimed at improving library tutorial instruction rather than classroom instruction, as focused on here.

More interesting for this study, faculty training in UDL appears to matter in terms of student perception outcomes regarding content presentation. Content presentation would be one of multiple topics covered in typical UDL training and subsequent evaluation, and some studies break down subtopics, allowing understanding of content presentation within the larger study's context. One such study of over 1,000 students surveyed before and after their professors received 5 hours of UDL training indicated that faculty improved in areas such as providing material in multiple formats, among other positive results (Schelly et al., 2011). A follow-up study using treatment and control groups and more detailed questions about aspects of UDL practice also found that 5 hours of UDL training led to improvements perceived by students, again including offering materials in multiple formats (Davies et al., 2013). However, as an acknowledged limitation, this study did not start by evaluating a baseline condition, so pre-existing differences in faculty knowledge and practice may confound the results.

UDL theory encourages faculty and others involved in course design to engage students through a range of teaching and learning practices that are based in brain science. While inspired and informed by the general universal design movement, UDL's foundations lie in cognitive neuroscience, focusing on the brain's affective networks, recognition networks, and strategic networks (Meyer et al., 2014; Rose et al., 2006). Recognizing that varied student learning needs too frequently are not adequately addressed through course design, UDL advocates for incorporating multiple means of engagement, multiple means of representation, and multiple means of action and expression into course experiences. The UDL guidelines and checkpoints in each of these areas are intended to help educators recognize and support the full spectrum of variability within these brain networks related to learning (CAST, 2018; Rose et al., 2006).

As researchers learn more about the scale and scope of these dimensions and their practical application through observation and experiment, conceptions of what constitutes good educational practice can be refined. While targeted at improving the experience of students with disabilities, UDL advocates posit it extends beyond addressing students with disabilities to hold relevance for all students (Edyburn, 2010; Rose & Meyer, 2002). The present study investigates the benefit for all students. This is particularly salient in postsecondary education where many students with disabilities elect not to disclose their disability, facing potential sigma and negative faculty attitudes toward providing accommodations for disabilities (Bettencourt et al., 2018; Evans et al., 2017). Older adult students may also be deterred from seeking and receiving accommodations due to the cost of acquiring current disability documentation (Bittinger, 2016). An approach offering the ability for students to perceive information in a variety of ways (e.g., textually, auditorily, or visually) supports students with functional impairments as well as students who “may simply grasp information quicker or more efficiently” in alternate formats or who improve the depth of their learning because they more effectively transfer what they see or hear to making connections that embed those concepts in long-term memory (CAST, 2014, para. 1). Thus, designing courses assuming a range of ability levels as UDL does should better support students with undisclosed and unaccommodated disabilities who may otherwise struggle to succeed as well as students with a variety of learning abilities generally.

While acknowledging the individuality of each person and their unique experiences, it seems like common sense to identify recognizable patterns of variation in the way humans learn that may assist educators in identifying a broad array of alternative learning experiences and then apply that knowledge to teaching and learning (Rose et al., 2006). The need for flexible approaches to instruction stemming from recognition of this inherent learning variation constitutes a key foundation of universal design for learning (UDL; Rose, 2001). UDL intentionally fosters a learning environment inherently accessible to all students that everyone can navigate seamlessly. Different students might take alternate paths through that environment, just as some people may use a physical ramp while others who are able to may choose to use the nearby stairs. Educationally, it should be possible to achieve the learning objectives in multiple ways. To educational theorists, this means creating learning environments where all students are enabled to reach common learning goals articulated for a course, avoiding generating disabling circumstances in a class context. The point is not that everyone ought to go through the same learning path (i.e., a single path that everyone universally follows); it is that everyone ought to be able to follow a learning path that allows them to successfully navigate the territory (i.e., a design allowing *all* students to be guided to or to find an individually appropriate path). Such an environment presumes individualized paths facilitate success and structurally enables this. The point is to empower each student to follow a learning path they can navigate to learn successfully.

4.2 Data

The undergraduate women in this study were nontraditional age students, typically juggling family and work in addition to school, a group frequently underserved by higher education. They attended a women-only university in the Northeast, and all sections of one basic English course (which I call ENG1) from the Fall 2018/Spring 2019 academic year were analyzed. This course combined a learning management system for discussion and overall course interactions (e.g., assignments and grades) with an adaptive system that included content and mastery level learning assessments. The six-week course contained learning activities designed to take about 20-minutes, with two to nine activities per week. Log files captured the modality used by a student in the adaptive system each time she went through an activity's material.

The course had an articulated knowledge map of the learning expectations for students that made connections between content activities explicit. This meant that in the analysis model, it was possible to map the structural connections between the activities within the course. In addition to modality information, tutoring session information from the tutor.com platform was available from within the institutional data warehouse. The course website in the learning management system presented a link to tutor.com, and each student had several hours of free tutoring available to them. The data warehouse collected information about the tutoring subject, the start date and time, and the duration of the session. These data were combined with data from the learning management system (for the weekly grade outcome), adaptive learning system (for learning activity information), and student information system (for covariates).

For each activity, the student had a known knowledge state score upon entering that activity in the course site on the adaptive platform. Upon completion of the activity, the student's knowledge state was updated after they completed a brief assessment.

In an approach consistent with universal design principles around providing alternatives for perception, students who showed signs of struggling to understand the content were encouraged by the adaptive learning system to pursue paths along alternate modalities until the student found a path guiding them to successful content mastery. The system log recorded each of these paths traversed through the material with an identifier, timestamp, and duration. Additionally, any student could use additional modalities that were offered by choosing to repeat the activity even if she was not struggling. The log similarly captured such repetitions, providing a detailed account of student activity within the adaptive system. Even though adaptive systems have the potential to make content available in multiple modes when such content is designed into them, this possibility not been a focal point of prior research that investigated presentation mode (Mustafa & Sharif, 2011).

4.3 Variables

The aim of this proof-of-concept analysis was to argue for the potential benefits of identifying predicted tutoring intervention points within a course where multiple content modalities were also available and illustrate a potential process for doing so. The outcome analyzed was the predicted probability of achieving either an A, B, or C, versus a D or F, on the mean grade for week one course assignments (see Appendix C). This analysis investigated the full sequence of learning activities in week one of the content presented by the adaptive learning system. Week one had two activities in the adaptive learning system.

Every student had a knowledge state score upon entering each activity in the adaptive system that was either determined at the beginning of the week through an assessment or else was based on her knowledge state at the completion of the prior activity. Upon completion of the current activity, the student's knowledge state was assessed again. These states were determined by the adaptive system using a proprietary algorithm to which I did not have access. The adaptive platform then determined whether to recommend that the student go through the material again in another modality, and if she followed that advice, her knowledge state score was updated when she completed the exit assessment.

A variable representing whether the student received tutoring provided the simulated intervention focus. A binary indicator represented whether tutoring was obtained after the student began to work on the current activity and before she started working on the subsequent activity. This variable was handled differently during different steps of the analysis process. Actual data initially trained the models when identifying parameters for later prediction. Subsequently, different parts of the simulation set whether the student received tutoring to either yes (1) or no (0) to compare the potential outcomes predicted under these scenarios.

The other main variables in the Bayesian network included use of multiple content modalities during an activity and repetition of that activity. Use of multiple content representations was operationalized as student use of at least any second content modality (e.g., a second of either text, video, audio, interactive, or mixed presentation) when learning the material within a given activity. Repetition of an activity was a binary indicator of whether the student went through part or all of an activity two or more times.

Demographic and prior educational independent conditioning variables were also used. These included measures of race/ethnicity, age, Pell grant status as a proxy denoting low income, and the number of credits transferred into the institution upon entry as a proxy for prior educational experience.

Given the complexity of this type of analysis and its proof-of-concept nature, missing data were handled via listwise deletion. Listwise deletion was applied individually for the regression equation for each variable in the network when determining the appropriate parameters for subsequent prediction.

4.4 Methods

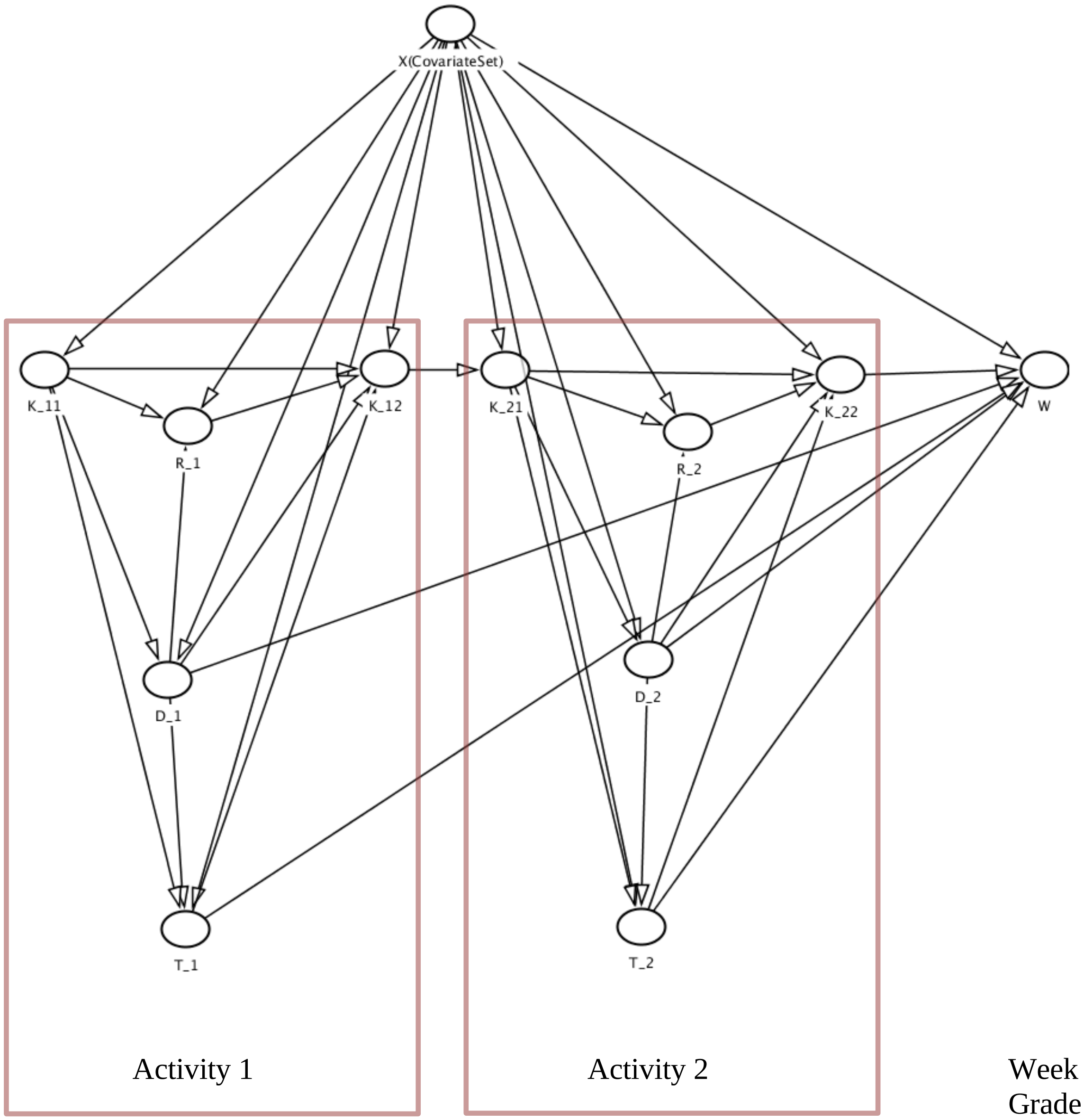
The adaptive system log data captured sequences of usage, and patterns of student use of content representations were investigated visually (Theus & Urbanek, 2009). This included a descriptive plot of tutoring, modality information, and general activity information across all the activities in the course, the first two weeks of which are presented here. Consistent with an analytics approach focusing on what happens for individual students, heatmap plots allowed visualization of cases for all students individually. Rows of the heatmap plots correspond to each student. Variables displayed included the same four as on the sequence plots. The heatmaps were clustered by similarity on both rows and columns to show patterns of the combinations of modality use and tutoring. Additional methodological information is provided in Appendix C.

The analytical approach I describe as a proof-of-concept illustration was based on a Bayesian network using a directed acyclic graph (DAG) to illustrate the assumed causal relationships in a model of modality use and tutoring intervention (Pearl, 1995, 2009a). The analysis offers a proof-of-concept description of updating a network of probabilities for individual students on the fly during a given instance of a course when a presumed tutoring intervention is simulated from data combined across multiple systems in the campus data warehouse. This illustrates the kind of calculations that could be employed to identify opportune moments to make recommendations regarding tutoring when use of multiple modalities may be predicted to be insufficient to help a student master the content.

A Bayesian network approach allowed incorporation of a holistic view of students progressing through a given module of the course. The first module (i.e., week) of the course was analyzed in the example. Figure 4.1 shows a DAG sketch of the Bayesian network for this module, which had two activities, drawn in Daggity.

This model was used twice, once to estimate the parameters of the model from data, and a second time to infer the values of unobserved variables after intervention. (See Figure C.4 in Appendix C for a modified DAG representing the network after the intervention was applied.) Each prediction included adding an error term drawn from a normal distribution with mean zero and standard deviation equal to the root mean square error of that variable's model equation.

Figure 4.1 DAG of a Basic Bayesian Network Showing Combinations of Modality Switching and Tutoring for Week One of ENG1



In the DAG, the binary week grade outcome is represented by *W*. The four *K's* represent the knowledge state scores in the adaptive learning system at different time points, two for activity one and two for activity 2. Since the adaptive system sometimes made additional adjustments to this score in between what it recorded in the log at the end of one activity and the beginning of the next, given other information it had about the student in the system overall, these two knowledge state scores (which exist separately in the log file) were modeled separately.

The *R's* are binary indicators of whether the student repeated that activity. The *D's* are binary indicators of use of multiple modalities at any point while engaging in the activity. Each indication of multiple modalities includes use of up to five different types of modalities. The *T's* are binary indicators of whether the student received tutoring after beginning that activity and before beginning the next activity in the adaptive system. The second tutoring variable had its value set in the simulated educational system by "doing" *T* (or do(T_2) in the notation of causal graphical modeling). This denotes setting the value of this variable to 0 and 1 in different branches of the simulation and comparing the results of those branches (Lübke et al., 2020). For all variables analyzed, the set of covariates, X, was also assumed to be relevant, consisting of student demographic variables and prior academic achievement.

The model depicts the two activities assigned during week one of this English class. Activity sequences assigned during weeks two through five of the six-week course were not analyzed for this proof-of-concept example to keep the illustration simpler.

This approach builds on the knowledge mapping done by the institution's course design team prior to putting material into the adaptive learning system. This example focuses on one course (ENG1), with activities connected to each other in sequences according to this course's curricular knowledge map. While the education-oriented Bayesian network analysis conducted by Xenos (2004) had similar complex contours to this example overall, for the purpose of this proof-of-concept description, the potential complexity was substantially reduced by using a relatively simple causal structure with a limited number of variables in the model for illustration purposes.

Figure 4.2 presents an overview of the steps involved in using a Bayesian network approach to simulate an intervention. This overview corresponds to the model presented in Figure 4.1 under the assumption that the arrows in that figure represent causal relationships without unobserved confounding.

Figure 4.2 Bayesian Network Intervention Simulation Overview

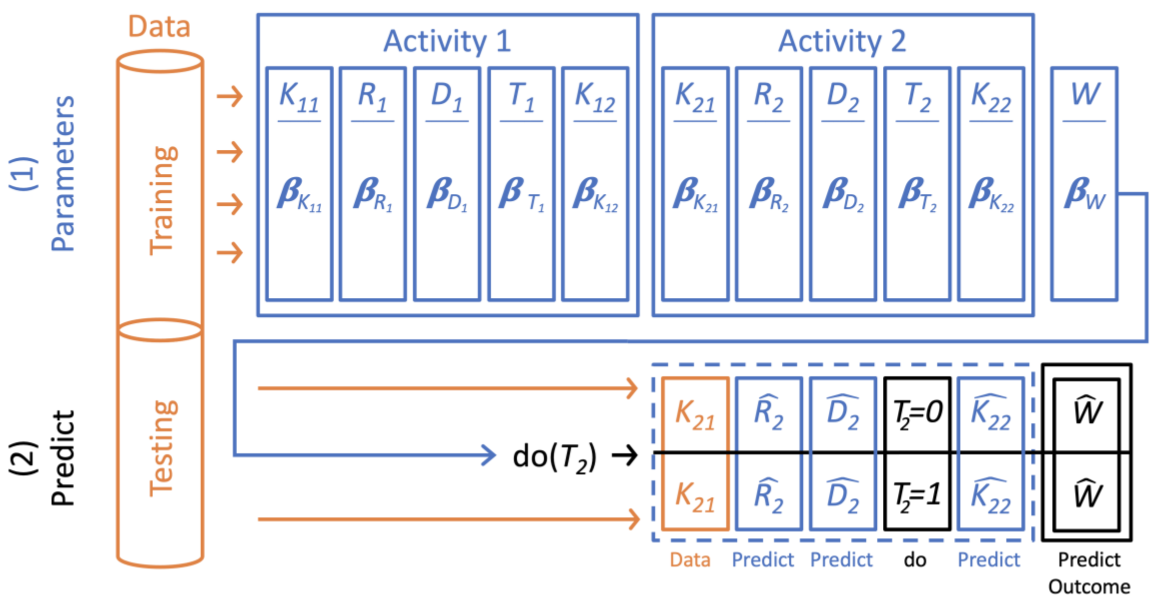


Figure 4.2 presents a way of thinking about the data, the models involved with identifying the conditional probability distributions of each node of the Bayesian network, and the simulated intervention that predicts two complete datasets under the two scenarios entailed in the experiment on a simulated educational system. The top panel (1) depicts the first step in the analysis process where the parameters of each node in the DAG are determined. For a more complex DAG including more weeks of the course, appropriate activities and variables could be added to determine additional parameters.

Training data were used to determine the parameters ($\boldsymbol{\beta}$) for the variables in the models for each node in step 1. Testing data was held out from the full dataset and not used for training these models. In this example, half the data were used for training and half for testing given the very small amount of tutoring information available. To verify each individual model's predictive ability, I conducted 10-fold cross validation on the entire dataset to determine goodness of fit for the models with several metrics. Results of these model validity checks may be found in Appendix C.

The testing data formed the basis of the predictions in step two in the lower prediction panel (2) of Figure 4.2. In this example, I chose to set the values of tutoring during the second activity, T_2. That is, I chose to set tutoring to 0 and 1 for a given student, do(T_2), to compare potential outcomes in the two worlds thus created for that student in the simulated system. This meant the simulation began after the initial knowledge state score (K_{21}) for activity 2 was determined. Since the simulated intervention happened on tutoring, values for this variable were set to $T_2 = 0$ in one half of the simulation prediction step and were set to $T_2 = 1$ in the other half. At this point in the simulation, there were two submodels representing the two simulated scenarios. Other variables were predicted (e.g., $\widehat{D}_2$) within each corresponding part of the submodel. These predicted values were used as "data" in the predictions of subsequent nodes (i.e., also called child nodes) in the submodels, such as the week grade outcome (*W*). See Appendix C for further information about the Bayesian network analysis in this example.

Once the outcome predictions were calculated, the effect size of the outcome of the intervention was evaluated for individual students. I report results for four students as examples. For each student, results in the $T_2 = 0$ condition were compared to the $T_2 = 1$

condition using kernel density plots of the distributions of the predicted values for a given student as well as statistical t-tests of the difference in the means.

4.5 Limitations

As with other analyses in this dissertation, the number of students with disabilities was very small. Only two out of 385 students in the sample were diagnosed with disabilities, but these two were dropped in the listwise deleted sample used in the example. As already discussed in prior chapters, this presents an obvious limitation for understanding the implications of this type of analysis for the students with disabilities who the UDL framework was designed to assist, and future research should investigate this further.

The Bayesian network assumes that the causal structure of the model is accurate and is not missing any confounders. However, this example analysis was intentionally made relatively simple to focus on the concepts involved in the analysis and illustrate its components. As an example of a variable that might be missing from the model, results for a given activity may depend on the difficulty level of the material covered. Future research could attempt to include a measure of difficulty by perhaps including the average length of time spent on that activity by all students in the training data, the historic mean change in knowledge state score for the activity, and/or an estimation of difficulty by the course development team. Future research should further consider such potential alternative explanations for the predicted effect of a tutoring intervention, seek to evaluate their effect on student learning outcomes, and include measures or develop a study design that would reduce any impact on results.

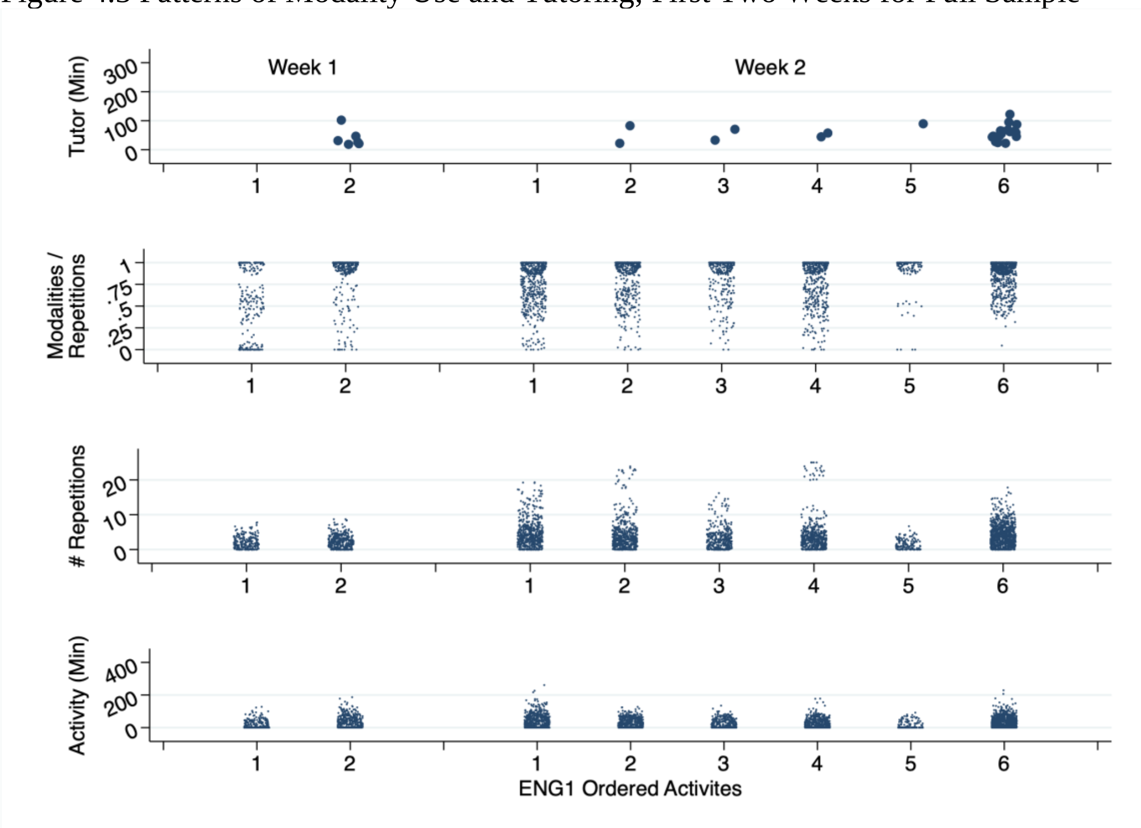
As an illustrative study, these results are not meant to be generalized to other courses or settings despite being quantitative in nature. Only one course is analyzed at one institution. Within that course, only one week is emphasized. However, given the modular nature of the type of analysis conducted here, it should be possible to scale this approach straightforwardly to additional weeks in the course or to different courses, as well as apply the concepts to different institutions. While this study offers an example that could be followed, additional work would be needed to integrate results with a dashboard presentation or email notification for students, faculty, support staff, or course designers so they could easily make use of the information.

4.6 Results

Figure 4.3 illustrates patterns of modality use and tutoring aggregated across all instances when the course was taught. The x-axis shows the series of activities for the course ordered from the knowledge map designed for the course and instantiated in the adaptive learning system. The two activities assigned during the first week appear under the Week 1 heading, and the six activities assigned during the second week appear under the Week 2 heading. (See Appendix C for additional descriptive results, including a plot for all five weeks of the course.) The four plot rows show the amount of time spent getting tutoring, ratio of the number of times multiple modalities were used to the number of repetitions of an activity the student chose to do overall, number of activity repetitions, and amount of time the student spent working on the activity in the adaptive system.

Many students used multiple modalities upon repeating the activity. Students frequently repeated activities. The amount of time spent varied per activity, but the density of higher values did not always correspond to either tutoring, modality use, or repetition. Notably, few students in this sample received tutoring, presenting a limitation for analysis with this sample (Morgan & Winship, 2015). However, enough students received tutoring during the second activity to enable illustration of the technique with appropriate extrapolation. In actual application, evaluation of sufficient overlap between treatment and comparison cases should occur with a larger sample than available here. As data continues to be gathered in the data warehouse over time by students taking the course, model estimation could continue to be refined, improving predictive ability.

Figure 4.3 Patterns of Modality Use and Tutoring, First Two Weeks for Full Sample



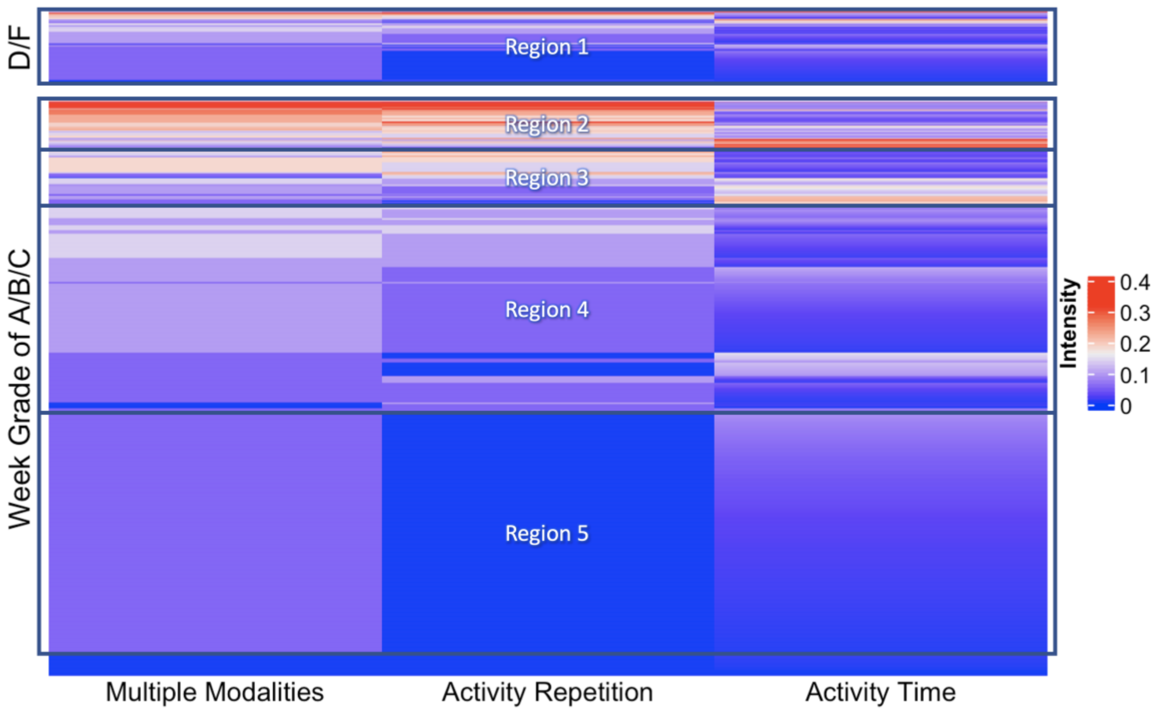
Note: Points jittered, and tutoring points enlarged for visibility. Activities for weeks one and two shown in sequence; there were two activities in week one and six in week two. The four rows of plots from top to bottom show: 1) the amount of time each student spent receiving tutoring after beginning to work on the activity (zero tutoring times not displayed), 2) the ratio of the number of times multiple modalities were used when working on the activity to the number of repetitions of that activity overall by each student, 3) the number of times each student repeated the activity, and 4) the amount of time each student spent working on the activity overall.

Given that the Bayesian network simulation analysis aims to present predictions for individual students, a descriptive approach looking at individual cases across the entire dataset holds more salience than the typical descriptive approach of aggregating the data using statistical summary measures. Heatmaps offer a compact yet complete description of the full dataset through visualization consistent with a prescriptive analytics perspective. To get a visual sense of the entire dataset, Figure 4.4 presents a heatmap of all individual cases, clustered by variable on the horizontal axis and case on the vertical axis. The variables analyzed for clustering included whether a student used multiple modalities for a particular activity, whether repetitions of that activity occurred, and total time spent on that activity and on tutoring. To facilitate comparing these variables' distributions, their ranges were converted to [0,1] using min-max scaling.

Tutoring was not displayed in Figure 4.4 because so few cases involve tutoring overall that they would be difficult to see since individual cases display as very thin rows and few case clusters were visible given the amount of data not involving tutoring. Instead, cases involving tutoring were featured in Figure 4.5 so that patterns in the data involving tutees would be more visible, including total time spent on tutoring. Both Figures 4.4 and 4.5 were created using R's *ComplexHeatmap* package with Euclidean distance as the clustering distance metric (i.e., the shortest difference between two points in the multidimensional space including the analysis variables) and the default "complex" agglomeration method of clustering.

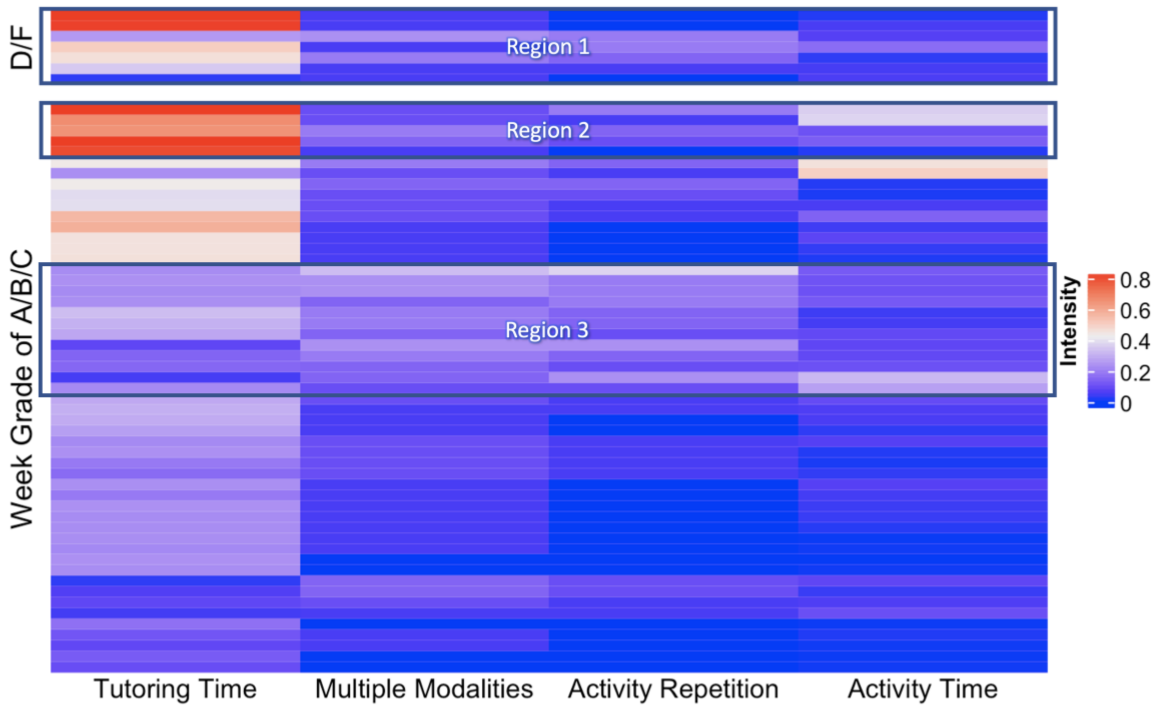
Additionally, both plots are split by grade group. The lower panel shows cases where the student received a grade of A, B, or C on their mean weekly assignment grade, and the top panel shows cases where the student received a grade of D or F. Two students from each of these groups in the test data were selected for the intervention simulation.

Figure 4.4 Clustered Heatmap of Adaptive Activity for Full Sample, Split by Week Assignment Grade



Note: Rows display individual student cases, with each student having five rows displayed for the five course weeks that included adaptive activities. Multiple modalities, activity repetition, and activity time variables were each scaled to range between zero and one. Color was assigned based on these normalized values, with each row/column line showing the intensity value of that variable for that case for a particular student in each week of the course. Rows and columns were clustered by similarity using the Euclidean distance between pairs as calculated by R's *ComplexHeatmap* package; this determined the order of the rows and columns, displaying similar students together along the vertical axis. Discussion of the regions of interest highlighted in numbered boxes may be found in the text. Cases where students earned a weekly grade of either A, B, or C were grouped together in the bottom panel (in region two and below), and cases where students earned a weekly grade of D or F were grouped together in the top panel (in region one).

Figure 4.5 Clustered Heatmap of Adaptive Activity for Students Receiving Tutoring, Split by Week Assignment Grade



Note: The description in the note for the previous Figure 4.4 applies here as well, although this Figure 4.5 displays only students who received tutoring and adds the amount of time spent receiving tutoring (scaled between zero and one). See the text for interpretation of the regions highlighted in numbered boxes.

While numerous patterns exist, I highlight several groups of students visible from inspection of the clustering in Figures 4.4 and 4.5 as labelled in the numbered box regions. In both figures, region one includes those students who struggled the most, receiving a D or F for their weekly assignment grade. These students typically spent less time in the adaptive system than other students and were not the heaviest users of multiple modalities.

Looking at Figure 4.4 region two, the group with high levels of activity who received a weekly assignment grade of A, B, or C, these students frequently repeated activities in the adaptive learning system and used different modalities while doing so (i.e., high/red in both multiple modalities and activity repetition). In this region, a pattern appears where students were higher on either modality use (and associated repetition) or time spent on adaptive learning activities, but not both. That is, students who chose to use multiple modalities most frequently (or repeated the activity most frequently) did not take the longest time working in the adaptive system. This pattern held true across the highest activity group in region two, across the moderate activity group in region three, and as a less pronounced pattern in the lower activity group in region four as well (i.e., redder/lighter groups for modality use and activity time variables in each region do not correspond with each other, instead corresponding with bluer/darker groups in the other variable). This suggests that students typically use different strategies when working through the material, either choosing to go through the material in different ways or choosing to go through the material more slowly. Additionally, in region five, there seem to be a substantial number of students who used multiple modalities but did not repeat the activity, suggesting that numerous activities may have been designed with use of multiple modalities built into the expectation of learning along the main path of content.[5]

Looking specifically at students who received tutoring in Figure 4.5, most were not high on the distributions of the other three variables. That is, tutees were not particularly high relative to others on use of multiple modalities, repetition of activities, or time spent on adaptive activities (i.e., darker colors for adaptive activity). For example, the tutee group in region two spent noticeably more time than the rest getting tutoring but only spent moderate time in the adaptive system (i.e., high/red in tutoring time, but bluer

[5] This finding from region five about use of multiple modalities without activity repetition supports the merits of the choice in chapter three to analyze more than just a single use of multiple modalities across each week.

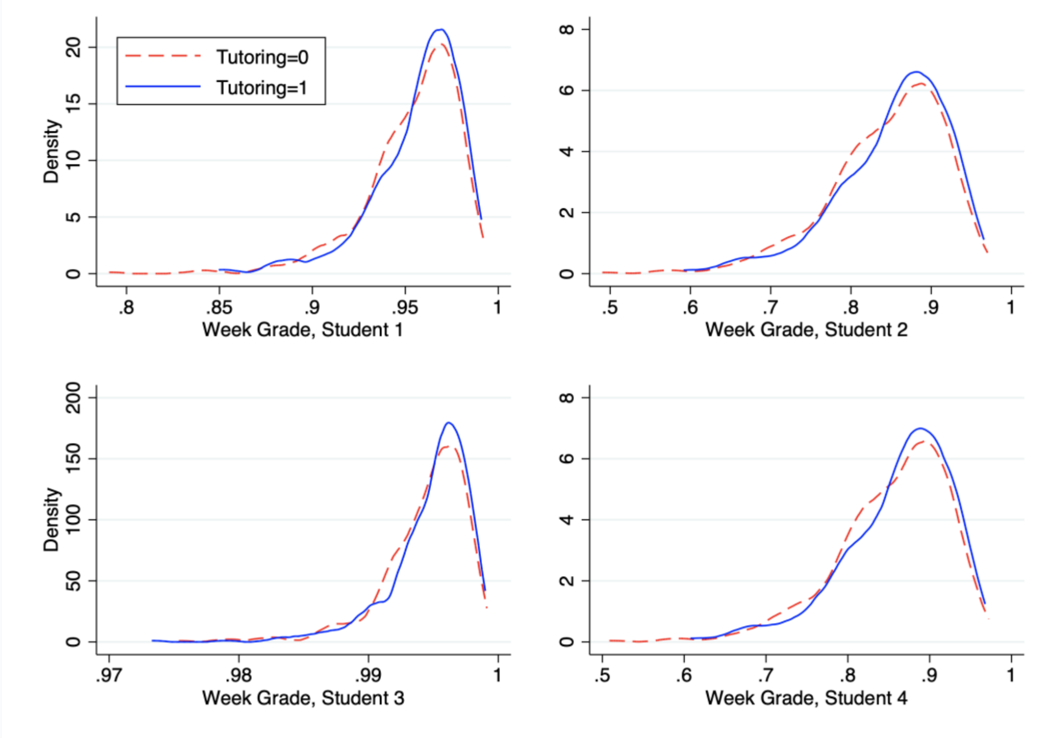
for the other variables). Tutees in region three spent moderate amounts of time receiving tutoring and were slightly higher (i.e., lighter blue) on the distributions of adaptive activity such as modality use than other tutees. However, tutees were not among the highest in adaptive activity across the entire class. In general, patterns seem similar for students who received a D or F for their week grade (region one) and those who got higher grades. However, it is striking that most tutees who received lower grades spent relatively more time getting tutoring.

While other students also appear to need assistance, the groups highlighted stand out, either because of their noticeable difference from others given the highly skewed distributions of these variables, or because of similarities between students. From this descriptive look at patterns across tutoring, modality use, and overall adaptive activity, it does not appear that information amenable to data collection from one type of learning support mirrors that from other types of support. Each provides unique information that could be useful in combination for understanding student actions that may support success. This means modeling intended to predict future outcomes within the course should include all these indicators.

For the Bayesian network simulation showing predictions projected into the future, there were 142 students in 22 sections in the training data and 103 students in 15 sections in the testing data. The analysis conducted here aimed to make predictions for individual students, and 500 predictions were calculated per student under each simulated scenario. This revealed the effect distribution, facilitating comparison of the predicted outcome in the two simulated worlds considering associated uncertainty. Since a small number of students would suffice for a proof-of-concept analysis that predictions can integrate tutoring, adaptive, and administrative data, four focal test students were selected. They received a range of grades in the course (B, C, D and F) as well as a range of grades on their week one assignments (0.7, 0.44, 0.96, and 0).

Figure 4.6 shows kernel density plots for each of these four students comparing the predicted week grade for the two simulated interventions of not receiving and receiving tutoring. These distributions show the outcomes in the two simulated worlds for each student, representing the uncertainty in the treatment effect estimate in these simulations. These plots visualize the heart of the prescriptive approach being illustrated, which involves identifying an optimal choice. While such plots would probably not be shared with students, they facilitate understanding the approach. In real-world applications, additional processing would turn such results into an analytics-based recommendation that could be offered to students, faculty, or interested others.

Figure 4.6 Kernel Density Plots of Tutoring Intervention Differences for Four Students



To facilitate interpretation of these plots, note that such predictive modeling differs from typical higher education applications where predictions illustrate results for certain groups or types of students. Often predicted probability calculations employ the same data that created the model. This typical approach facilitates understanding overall effects, providing nuance in understanding how an effect may operate for different groups of students identified by covariates.

In contrast, the present prescriptive modeling application aims to understand predictions made on an individual level for students who were not included in the dataset that generated the model. The kernel density plots illustrate this alternate prescriptive modeling application by visualizing the two different simulated worlds. The expected outcome for an individual in each experimental scenario (i.e., T=0 or T=1) can be calculated in the corresponding simulated world and these means statistically compared to determine an overall treatment effect for that individual.

In this example, the effect of getting tutoring for all four focal students had similar effect sizes that were statistically significant at the $\alpha = 0.10$ level. The effect size for a paired sample as measured by Cohen's *d* was $d = 0.075$ for student one ($p = 0.094$), $d = 0.078$ for student two ($p = 0.082$), $d = 0.079$ for student three ($p = 0.076$), and $d = 0.077$ for student four ($p = 0.085$). These can be considered medium effect sizes for an educational intervention (Kraft, 2020).

4.7 Discussion

This study presents an exploratory investigation of the variation in the data about modality use and tutoring activity, describing clusters and patterns found in the data and illustrating how such information might be used in a simulation projecting predicted results within an educational system. Results suggested that while a number of students used both multiple modalities and tutoring, not all utilized both means of support simultaneously. Most students who received tutoring used multiple modalities but were not the heaviest users of multiple modalities. Students who were the heaviest users of multiple modalities were typically not those who spent the most time working in the adaptive system, and vice versa. Most tutees received a grade of A, B, or C on their weekly assignments. The other groups noted in the results appear to be mixed in terms of their outcomes, suggesting that additional tutoring may be beneficial for some of them. The existence of noticeable clusters suggests that further analysis drilling down into these patterns would be warranted in future research. This initial descriptive look suggests there were groups of students with similar patterns of modality use and tutoring activity that may warrant further investigation regarding the combinations that lead to greater student learning at particular points in the course. Four such groups stand out, including those who rely most heavily on tutoring, those who rely most heavily on repeating activities using different modalities, those who rely most heavily on spending time going through the material in the adaptive system slowly and those who spend a moderate amount of time getting tutoring combined with moderate use of multiple modalities.

Interestingly, use of multiple modalities and tutoring did not coincide as frequently as I initially anticipated. This suggests that some students may prefer repeating the activity using different modalities whereas other students may prefer getting help from a person they can talk to. I suggest that such descriptive patterns along with the predictive information about specific activities could be used to identify points in the course where students more frequently sought help through working with multiple modalities and tutoring. This in turn should enable deeper future exploration of whether particular combinations of modality use and tutoring for specific activities provide greater benefit for student grades for assignments for a given week, as well as their grade in the course overall. Future qualitative research could investigate the motivations students have for pursuing these different strategies. This would help distinguish the extent to which there may be additional predictive power when students use both strategies together as I initially thought, along with the extent to which each strategy (i.e., tutoring and use of multiple modalities) separately entails useful information about student confusion and difficulty with the material. My initial results suggest all of these strategies (i.e., tutoring, use of multiple modalities, and combining these strategies for seeking help) may provide beneficial information for future prescriptive modeling.

Each of the four students illustrated here had a prediction of what would happen in the simulated world where they received tutoring and the different simulated world where they did not receive tutoring. Each predicted point estimate in this experiment on a simulated educational system was drawn from a distribution of such possible estimates. These led to the reported treatment effect estimates for each student. In the real world, time limits our ability to explore potential outcomes that are not observed in real time. Here, simulating two potential future worlds for each student allowed exploration of both potential outcomes under the two scenarios explored through models developed from prior students' data.

The effects found for the four focal students in the example presented for the second activity in the adaptive system were reasonably sized for educational interventions (Kraft, 2020), but should be combined with the results of further analysis to determine what size effect would be considered actionable at this institution. As hypothesized, prescriptive analysis found that combining modality switches and tutoring should benefit some students. Additional research should be conducted to determine the extent of this benefit and determine the most opportune moments in the course to provide tutoring recommendations overall to help focus guidance to students. This initial look offers an example of the kind of analysis that could be conducted with more activities to evaluate various points in the course. The approach offers numerous avenues for potential future research to refine the modeling, improve targeted predictions, and identify practically useful prescriptive analytics. Future work could extend the current analysis to develop model-based predictions of success in future activities, weeks, and the course overall that could be based dynamically on data collected to date about each student at various time points within the course. The amount of data for any given activity-tutoring combination in the dataset analyzed is small, so collecting additional data in future semesters to add to the analysis would be beneficial. This is the kind of initiative that would be possible at an institutional level, particularly as data continue to be collected within the data warehouse across time.

4.8 Implications

The COVID-19 pandemic catapulted UDL design concerns into a broader spotlight as faculty rushed to shift their pedagogy to emergency remote delivery (Basham et al., 2020; Levey et al., 2021). The increasing awareness of accessibility issues has challenged faculty to change practice (Burgstahler, 2021; Cook et al., 2009; Izzo, 2012). The analysis presented here offers one avenue to extend that changing practice to further support students. UDL provides a framework for faculty to make instructional design decisions. Future research could look at discriminating between groups of students who exhibit particular learning behavior patterns early in the course so that appropriate recommendations might be made for them. This might also help instructors identify where to focus additional class time, for example.

Comparing simulated worlds in a hypothetical experiment builds on existing analytics approaches that focus on predictions for individual students by extending the timeline of prediction throughout a course. Utilizing a simulated worlds approach with a Bayesian network framing opens the possibility of a series of studies that integrate knowledge about causal effects into experiments on simulated educational systems. This chapter presents a first step toward combining knowledge about how various possible interventions might affect student learning in simulated worlds. Multiple interventions could be explored separately and simultaneously in future research to guide actions of students, faculty, and course designers. Here I demonstrated how knowledge about elements of learning could be combined in a network representing activity throughout a course, with downstream effects on learning outcomes, enabling questions to be asked about what would happen under different scenarios. The process I described projects predictions through the simulated educational system to identify estimates under different potential outcome scenarios given two possible treatment states in a hypothetical experiment. This process can be used as a tool to ask interventional questions that could be used by instructors advising students or revising curriculum, by students deciding what support to seek, or by academic advisors identifying how students they support might be helped. While the present study focused on a tutoring intervention, various interventions could be explored through this technique.

While Bayesian network-informed approaches have been utilized in adaptive learning systems, they are typically either still experimental or have been developed behind paywalls in proprietary systems (Kabudi et al., 2021). I argue that institutions could be developing similar capability themselves or working with vendors willing to make the steps of the predictive process transparent enough to enable augmented predictions utilizing aggregated data across multiple systems. This illustration shows how such an approach might work to benefit students, particularly when an institution can aggregate data across multiple systems from different vendors in a data warehouse.

If the results of this type of analysis were presented in a dashboard (e.g., perhaps one that might be integrated with another dashboard presentation developed by the institution), then the results could be more easily understood (Kitto et al., 2017). An instructor might utilize this kind of information when considering whether to reach out to a student to suggest tutoring. An automated system could be developed that presents such information on a dashboard for points in the course that challenged previous students in the course, or email could be sent to students with a personalized message at appropriate moments during the course (Pardo et al., 2019). Alternately, students or faculty could use a web interface developed to allow inquiry into the benefit predicted for a student receiving additional assistance at times of interest during the course. Particularly when tutoring support is limited, assistance identifying where such support might be most beneficial could help students (and institutions) most effectively utilize this resource. Importantly, however, educator and student perspectives about effective communication of results and associated recommendations would be key to any such implementation (Shibani et al., 2020). While I offer these speculative thoughts since such presentation issues would be highly relevant in real world applications, such presentation questions are beyond the scope of the present illustrative example.

The literature on adaptive learning has sought to identify benefit from using learning style information to tailor the adaptivity of such systems (Khamparia & Pandey, 2020), despite the hypothesis that matching material to learning style preferences has earned status as a "neuromyth" (Betts et al., 2019). For example, prior experimental research on adaptive learning across five days with 42 students suggested that students benefit from having the initially presented modality tailored to their learning style, although use of the multiple modalities provided was not explicitly studied even though they were provided (Mustafa & Sharif, 2011). For the present study, system administrators believe it is likely that the adaptive learning system did not have sufficient time to determine such preferences and change the default modality presented for many, if any, of the students studied. While students could have changed the default setting on their own, this possibility was not advertised to them, so the default initial presentation (typically textual) was likely enabled for almost all students. Future research could further investigate such possible benefit using data before and after the students had used the adaptive learning system enough for it to adjust the initial modality presented.

To focus attention on the structure of the analysis for this example, choices were made to simplify the model presented in Figure 4.1, but this illustration can be scaled up in numerous ways in practice to make the results increasingly realistic and useful. Future analysis extending this example could construct a more thorough model by adding variables representing possible alternative explanations, extending the model in time throughout the course, and/or connecting this course to other linked courses in the introductory English sequence. Only four students at one time point were focused on here, but this type of analysis could obviously be extended to across many students and different time points throughout the course could be compared to one another. Hopefully this illustration provides basic guidance for this kind of prescriptive analysis that will inspire future work extending this kind of simulation approach that allows comparison of outcome predictions within different potential worlds for a given student.

Additionally, there are numerous ways this preliminary research could be extended. For example, each activity throughout the course could be analyzed for the predicted effectiveness of a tutoring intervention coming after it. The effect on the overall course grade could also be predicted, and although more activities would be considered in the longer time frame, the logic needed would be similarly based on the conditional probabilities according to the assumptions of the Bayesian network. That is, an intervention point would be picked advantageously for analysis and existing knowledge about prior activity would be condensed into the knowledge state score at the beginning of the chosen activity. This would then become a parent variable involved in predicting values for the conditional probability distributions for subsequent variables within that activity to propagate through the network to the chosen outcome of interest (which could occur after the current activity). Future analysis could also extend this work to include analyzing interventions on other variables such as modality use, repetition of activity, or other possible explanations that could be added to the model as variables.

Given these results, it would be worth further investigating the type of material available at points in the course where students seek tutoring. This type of analysis might help faculty and others on a course development team address “pinch points” in courses

(Tobin & Behling, 2018). Perhaps the places where students are seeking tutoring are points in the course where fewer alternatives in different modalities are offered. If so, then this would have obvious implications for course redesign, as these points in the course would be strong candidates for redesign efforts involving modalities. Alternately, perhaps the points where students sought tutoring constitute harder segments of the course in general, and perhaps other UDL-based supports could be designed to engage students more effectively in those moments. Again alternately, perhaps these places are ones where students find it harder to demonstrate their knowledge through the multiple-choice questions asked by the adaptive learning system. If so, this suggests that a challenge faced by those developing course material for adaptive learning systems in general includes developing alternative means of formative assessment that are not multiple-choice-based techniques. Such possible alternative methods of addressing issues faced by students deserve further thought and investigation. Such directions could be beneficial to consider when adopting a “plus-one” approach to course development or when doing larger course revisions (Tobin & Behling, 2018). The present proof-of-concept analysis is intended to be suggestive of productive directions for future research and illustrate the kind of thinking that may offer institutions new possibilities for making use of the increasingly voluminous data being collected about student learning to positively benefit struggling students.

The idea of UDL encourages educators to ask questions about alternate perspectives on learning based on three brain networks known to be involved in learning, and leads to asking questions about alternate perspectives on teaching (Meyer et al., 2014). What happens if educators view their role as one of continually observing students and inquiring about alternate ways material might be presented in order to facilitate individual learning journeys by students of different backgrounds and abilities given the knowledge state that they demonstrate today (Montessori, 2014)? What happens when educators conceive of teaching as simultaneously about building relationships that support these learning journeys, identifying inspiring challenges appropriately tailored for individuals, and scaffolding content presentation in directed ways supportively intended to build knowledge (Wood & Wood, 1996)? What happens to student learning outcomes if we approach educational design in multiple ways that recognize a wide range of cognitive functioning along several dimensions pertinent to learning (Meyer et al., 2014)? The assumption that there are multiple valid approaches to learning constitutes a core tenet of universal design theory as applied to education. Further investigating combinations of practices connected to UDL that may provide benefit to students remains warranted.

4.9 Conclusion

This work augments prior analytics research by providing a proof-of-concept example that could be extended to other circumstances to investigate, predict, and present analysis results about the connection between elements of UDL and student success. It can serve as an example of the kind of prescriptive analytics that could be done by an institution that wished to utilize the student data collected in an adaptive learning system and online tutoring system in a data warehouse making these data available for analysis. I suggest that just as curb cuts provide an easily understood symbol of universal design in the physical, built environment, it may become the case that providing multiple modalities for learning content may come to be a similarly easily understood symbol of educational universal design. As chapters two and three have found, the practice of using multiple modalities offers benefits for a wide range of students. The idea holds face validity and is straightforwardly implemented (even if it takes resources to do so). Combining this knowledge with information about when students in prior sections of a course have received tutoring to beneficial effect holds potential to inform predictive modeling as discussed here. Such predictions may offer insight into when feedback to students about seeking tutoring might be most beneficial. This holds importance for students at this institution since there is only a limited amount of free online tutoring available to each student each semester. Such predictive support could aid students in determining optimal times to get additional support.

CHAPTER 5

INTEGRATED DISCUSSION AND IMPLICATIONS

> Don't rely on a single tool. If you can derive similar causal effects from different sets of assumptions, great. If they contradict each other, this is useful to know too. Make use of your background knowledge to disentangle the mess. (Silva, 2015)

This dissertation provides deeper understanding about a critical component of the Universal Design for Learning (UDL) framework. As illustrated in the ADDIE-based feedback loops in Figures 1.1 and 1.2, chapters two and three iteratively demonstrate the effectiveness of use of multiple modalities at the activity, week, and course levels, and chapter four illustrates how UDL-oriented data may be combined with other data to inform improved practice. This research utilized technological advances, including adaptively providing content in multiple modalities and collecting data at scale aggregated across multiple systems. Results support practical steps that can be taken to bolster student success based on deeper understanding of a component of UDL. The study in chapter four provides a higher education example of using Bayesian network intervention predictions to target improved student learning outcomes. This demonstrates how research addressing fundamental questions about what works from a theoretically informed standpoint may inform practice while also helping to identify needed future research directions to fill in causal understanding.

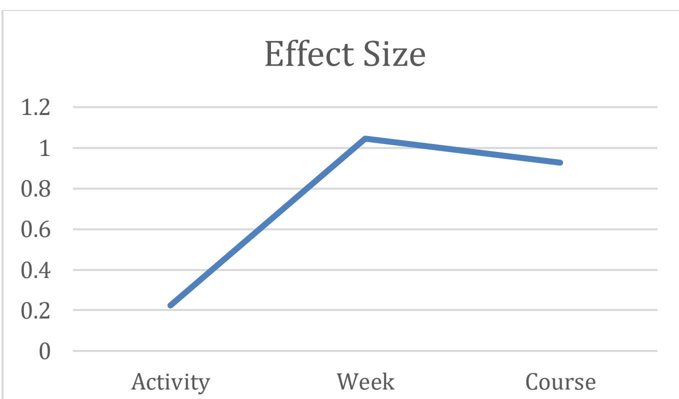
While chapters two through four each presented separate discussions of the results of chapter-specific analyses and offered implications of those results, this chapter takes a broader perspective. It synthesizes results from chapters two through four and considers the overall implications of the combined findings for theory, research, and practice.

5.1 Discussion

Research has been needed showing how specific UDL-guided practices, including making multiple modalities available, translate to learning success, and therefore to the outcomes of higher education's core educational mission. The lack of prior research about UDL's effectiveness motivated the present dissertation's investigation of using multiple modalities offered to nontraditional women undergraduates in an adaptive learning platform. The use of multiple modalities pertains to the variation in neural recognition networks that underlies the UDL principle of providing multiple means of representation. In my studies, I investigated the benefit UDL provides for performance on formative and summative learning measures including 20-minute learning activities, weekly assignment/quiz grades, and the final course grade.

The results of chapters two and three extend prior research by examining effects of use of multiple modalities on student knowledge scores and grades, seeking to confirm one aspect of UDL theory. The effect sizes for the main analyses as presented in the chapters are summarized in Figure 5.1.

Figure 5.1 Effect Size Summary for Activity, Week, and Course Level Analyses



As hypothesized, use of a second modality had a substantively important positive effect on learning gains, with greater than a 0.2 standard deviation effect size. Use of a second modality was found to have an even larger positive effect size of around one standard deviation for both grade outcomes. For the week grade, this corresponds to an improvement of 14 points out of 100, holding all other variables constant (see Table B.10). For the course grade, this corresponds to a positive effect of 0.75 points on a four-point scale, holding all other variables constant (Long, 1997). Both week and course treatment effects represent more than half a letter grade improvement as hypothesized.

To understand the meaning of these results, it is helpful to understand how effect sizes in education research are typically interpreted. For social science generally, Cohen (1977) considered 0.20 a small effect, 0.50 medium, and 0.80 large based on a broad sampling of social science research. Lipsey (1990) looked at 102 mean effect sizes from 186 meta-analyses of treatment effectiveness research in the behavioral sciences and, dividing the results into thirds, considered 0.00 to 0.32 small effects (midpoint 0.15), 0.33 to 0.55 medium effects (midpoint 0.45), and 0.56 to 1.20 large effects (midpoint 0.90). These recommendations are similar to Cohen’s advice, though somewhat more refined. However, the challenges of identifying policy-relevant interventions in educational research has led to development of more specific guidance about effect sizes for educational researchers. In higher education, Mayhew et al. (2016) recommend revising Cohen’s ideas for research on the impact of college based on the authors’ expert judgment, suggesting that 0.15 be considered small, 0.30 medium, and 0.50 large. Specifically looking at educational intervention research, based on the distribution of effects from over 700 K-12 RCTs with standardized test outcomes, Kraft (2020) suggests interventions with less than 0.05 be considered small effects, 0.05 up to 0.2 be considered medium effects, and effects over 0.2 be considered large. Considering a 0.2 effect size benchmark for educational research, all results in this dissertation would be interpreted as large effects for an educational intervention. However, even considering more conservative guidance, many results of this research would still be considered large.

Thus, analysis detected a medium to large positive effect (depending on which interpretive education-focused effect size guideline is considered) when understanding the knowledge gained across learning activities, and a very strong positive effect when those smaller effects were combined across a particular week for grades that closely pertain to the content (i.e., assignment and quiz grades). Results at the course grade level also suggest a large positive effect, although these results are somewhat less definitive since there is more uncertainty around the socioeconomic baseline equivalence for the groups compared and it is unclear how much this might have confounded the result. Although the sample size is small enough that results are just on the edge of being detected as significantly different, several sensitivity analyses also suggest the presence of an effect at the course level, even if difficult to detect.

Results revealed an effect both when including or not including zero grades in analysis. This means an effect is present even when considering students whose scores may have been impacted by other factors that affected their outcomes beyond their knowledge of the content, such as life events. This suggests making content available in multiple modalities encourages students to pursue learning strategies that help retain material.

Interestingly, models accounting for differences in first-generation status resulted in detecting a larger effect. Over three-quarters of students in the sample are first-generation, and such students are known to face challenges succeeding in college courses that may rest on cumulative disadvantages rather than innate ability (Bettencourt et al., 2020). Given this, results without accounting for this factor may underestimate the treatment effect. The amount of missingness for first generation status warrants some interpretive caution though, particularly if the assumption that missingness was random happens not to be true. However, the stronger result when including a first-generation indicator reinforces the importance of accounting for this aspect of socioeconomic status in obtaining an uncounfounded treatment effect estimate and suggests the listwise deleted results presented without this variable may represent an underestimate.

The two standard deviation effect size specifically for quizzes found in the chapter three week-level analysis is a very large effect. Although quizzes by themselves were not the primary analysis focus, this informal decomposition was done to investigate the supposition that grades on assignments closer to the content would be more affected by any treatment effect. My assumption was that quizzes would gauge knowledge of the content presented to a greater extent than general assignments. Given the strong result found for quizzes, this deserves further explicit research. While the result here provides initial evidence suggesting a strong effect of modality use on quizzes, I urge interpretive caution. Research suggesting large gains may get noticed even if it turns out later that another cause was driving the initial result, such as happened in the example where mastery learning was later found to explain more than half of the two sigma effect initially attributed to human tutoring (Bloom, 1984; VanLehn, 2011). As noted, the multiple modality effect on quizzes corresponds to a positive change of several grades, specifically, a 35-percentage point difference on quiz scores or an improvement index of +49 percentiles. This is large enough to invoke questions about what mechanisms may be at work underlying this result or whether there may be quiz-specific alternative explanations that should be considered. Given that the study design was not oriented around quizzes specifically, additional research should be done prior to drawing major conclusions. As an example of why caution is warranted, only a subset of the courses studied included quizzes, and it was deemed out of scope to study possible subject-related differences. Questions of baseline equivalence and attrition were also not specifically investigated for the subsample of quiz takers, but the sample was much smaller than the overall sample, so differences may exist. Thus, these results should cautiously be interpreted as suggestive of a very large effect that warrants verification by further focused research.

The prescriptive analytics example from chapter four suggests how institutions might build on knowledge about the effectiveness of using multiple modalities to make beneficial recommendations to students about actions they could take to boost their learning, such as getting tutoring. Given that the goal of achieving full universal design that works for all students is useful but often difficult to achieve, such understanding is crucial to educational practice. It becomes imperative to plan for the existence of courses that have been designed with UDL in mind but that still have gaps that have yet to be addressed. This practical reality makes it incumbent on institutions to identify those places where students are showing signs of struggling to learn the material and to offer timely recommendations for additional one-on-one academic support. The example illustrates how this might be done with a Bayesian network approach intended to serve as an inspiration for institutions to consider what would be possible to develop.

Some researchers argue UDL's components should all be investigated as a package because the UDL principles work together to create an effective learning environment (e.g., Chandler et al., 2017). This reasoning may have influenced UDL research to date, which has typically investigated all principles of UDL together (Mangiatordi & Serenelli, 2013; Orr & Hammig, 2009; Rao et al., 2014; Roberts et al., 2011). From this perspective, a central point of UDL is that the guidelines work in concert with one another to provide a holistic approach that benefits student learning, and I acknowledge this as a potential critique of my dissertation. However, in agreement with Crevecoeur and co-authors (2014), and following the typical theory-building process (Christensen, 2006), rigorous investigation of various tenets of UDL, both separately and together will be necessary for the field to gain a deep understanding of what aspects of UDL are key and why. Avoiding or limiting investigation of UDL's components will ultimately provide an incomplete picture of its effectiveness. Detailed investigation of isolated aspects of UDL, such as done here, will help illuminate their individual effects, facilitating more nuanced subsequent investigation of the extent to which integration of these aspects provides useful benefit beyond what each component provides on its own. These holistic and specific viewpoints complement each other. While I focus on representation, I acknowledge that good application of this one principle would not be sufficient to create a fully universally designed course environment. This focus represents a simultaneous limitation and strength of my research.

Overall, my results provide support for the position that UDL research and practice ought no longer to be a niche concern by those devoted to supporting students with disabilities. Instead, as Maria Montessori intuited a century ago (Kramer, 1988), and as UDL approaches today recognize (Rose & Meyer, 2002; Tobin & Behling, 2018), understanding how best to support students with disabilities and other underserved students opens a window into how to more effectively foster learning across the board for everyone. These results indicate that not only do options for perception help by providing essential access to students with disabilities, but they also provide benefit more broadly across students throughout the ability spectrum.

5.2 Limitations

While I have access to an unusually varied and robust institutional dataset for this dissertation, as with any analysis there are limitations. When I chose the institution to study, after talking with administrators, I expected I would have a reasonable number of students with disabilities in my sample. While in practice this is likely the case, I was surprised at how few students were recognized as needing accommodations for the courses I studied. Similar to other UDL research in an online setting which had no students with an identified disability (Rao & Tanners, 2011), if the students in my study did not seek any accommodations for these courses, even if they had sought accommodations for other courses in the past, my data would not distinguish this. Possibly, the already flexible nature of the online medium in which the courses were offered, plus the attention paid to quality online course development that included UDL principles, meant that surprisingly few students were officially recognized by the disability services office on campus as needing an accommodation in the courses studied.

The rate was two orders of magnitude smaller than national statistics would suggest is typical for the number of students with disabilities at an undergraduate institution in the U.S (i.e., 0.6% with course accommodations in chapter two compared with a 19% overall national rate; NCES, 2021). This obviously presents a major limitation for research about a framework intended at its heart to support students with disabilities. However, it is very likely that a larger number of enrolled students actually had a disability compared with the number of students officially recognized. This presumed discrepancy, based on anecdotal evidence from administrators and faculty conversations with students, is an example of why UDL principles are important to implement.

Utilizing UDL in course design means that fewer students may need official accommodation to succeed in their courses. It also may mean that students with disabilities who choose not to disclose for fear of stigma or other reasons will still get support they need even if they have not taken the (potentially expensive) step of being formally diagnosed in order to have their case evaluated for academic accommodations. In this sense, while future research with a larger number of students with known disabilities would be warranted, the present research offers an important contribution by revealing a possible practical impact UDL may have for an institution enrolling students who may be reluctant to disclose a disability. This may offer an example of the low number of accommodation requests that might be seen elsewhere by consistently implementing UDL principles along with other high quality online course practices, although such a tantalizing claim should be substantiated with further evidence.

There are several further limitations to note due to the sample I employed. My dissertation investigated only non-traditional students, only courses taught in a fully online environment, and only one, women-only institution. It remains unclear how generalizable these results might be to traditional students, to courses taught in a hybrid format or one primarily on campus with only supplemental online instructional materials, or to other institutions or institutional types. These areas offer opportunities for future investigation. My sample selection was intentional since this institution allowed me to investigate the usefulness of UDL's principle of offering students multiple means of representation. Notably, within the institution studied, multiple academic subjects were analyzed, warranting generalized conclusions across multiple subjects for this institution. It can be difficult to study multiple academic subjects at once in educational research and being able to do so is a strength of my dissertation. However, since I did not employ random sampling at the institutional level, my results should not be generalized to other student populations. Since I only investigated a single institution with a gender-restricted student population, the results are applicable to that institutional setting but may or may not generalize to other institutional settings or institutional circumstances. Given the lack of empirical research about UDL overall though, my dissertation represents an important contribution that can also guide future empirical research.

In particular, it remains unknown whether gender differences exist in the effect of using multiple modalities since no men were present at the institution studied and gender identity information was not available. In general, UDL research has not investigated gender specifically except in the context of sexual orientation (Couillard & Higbee, 2018). However, college outcomes are known to differ for males and females, particularly in STEM fields (Manly et al., 2018). As reported in Izzo's (2012) investigation of UDL implementation, the effectiveness of educational practices aligned with UDL's call for flexibility, such as use of interactive clickers, have been found to differ by gender. Thus, while there are no known reasons why the effect found here may differ for males, future research interrogating this further would be warranted to verify an assumption of no difference.

Measurement error was a possible threat to the validity of statistical conclusions (Shadish et al., 2002). In particular, the coding for the treatment studied here relied upon the creation of alternative activities in different modalities and modality coding of these activities by the instructional design team at the institution. Specifically, that team was not focused on designing learning activities that would optimize study of the construct of interest in this research. They were instead focused on development of content they thought would benefit students generally. So, for example, the "mixed" category was created because they sometimes chose not to separate content in different modalities, but this category was not always synonymous with multimedia. The consistency of coding for the activities across the members of the instructional design team is unknown, and thus may present an unknown source of error in the results. For example, in looking at the activities in one of the courses, I noted that content that might reasonably have been called "video" content was labeled "mixed." Presumably this was because of the titling and brief introductory technical information about the video at the beginning of the page where the video was embedded. Information about the consistency of the design team coding of such pages was not available, but it seems possible that without clear communication between the coders, such pages may have been labeled as "video" content by some coders instead. If such inconsistency existed, it would add noise to the data specifying particular modality sequences. Minimizing the effect of possible measurement error was a reason the present analyses were focused primarily on investigation of any additional modalities rather than discriminating patterns of particular modality sequences across all courses. Future research could investigate such patterns further, simultaneously verifying the consistency of the coding across different courses by different coders and clearly assessing the interrater reliability of that coding.

Finally, my setting is U.S.-based, as has been predominantly the case in prior UDL research (Mangiatordi & Serenelli, 2013). There might be important regional or cultural diversity relevant to fully designing courses from a universal standpoint that has not typically been represented or studied in the research literature, and such cultural differences might hold relevance if UDL implementation practice around presenting content differs culturally. This could be explored in future research.

5.3 Implications for Theory

Comparatively little universal design research before now has specifically targeted UDL's component of content representation, and within that, providing options for perception (Roberts et al., 2011). Although the research evidence underlying the ideas encapsulated in the UDL guidelines is robust (CAST, 2011), that research orientation has not carried forward to investigation of the framework itself (Murphy, 2021). While some past research has begun to provide empirical support for the concepts of universal design overall (e.g., Davies et al., 2013; Izzo et al., 2008; Schelly et al., 2011; Smith, 2012; Street et al., 2012), that support has not included rigorous effectiveness studies about student learning with control groups or within-person designs (Boysen, 2021). My dissertation contributes to building this research base by confirming the effectiveness of one aspect of UDL with such designs, advancing the higher education field's understanding of this framework.

My work begins addressing the need to more deeply understand theoretically posited effects in isolation (Crevecoeur et al., 2014). This dissertation focuses on understanding the effect of providing content through multiple modalities, particularly in the online context studied. Specifically, the analyses presented in chapters two and three verify the theoretical proposition that providing options for perception should benefit students, showing that a positive impact of use of multiple modalities occurs at scale across multiple subjects. Given that improving learning outcomes at scale remains challenging across higher education institutions (Fulcher & Prendergast, 2021), my results are particularly relevant because they represent evidence of successful practice across the online undergraduate curriculum at the institution studied. Additional evidence remains needed regarding the effectiveness of other aspects of UDL stated in the framework to expand this work (McGuire, 2014).

The fact that this benefit was found across all women taking courses throughout the curriculum at this institution across a full academic year empirically verifies the common sense proposition that the principles of universal design would show benefits broadly in education, even beyond the students with disabilities for whom the framework was designed (Tobin & Behling, 2018). This is in line with Roberts et al. (2011) who called for more studies about the impact of universal design approaches "on the outcomes of postsecondary education students with and without disabilities" (p. 5). The present results from two strong research designs begin to address the paucity of effectiveness research on UDL in higher education, a notable lack in the field that has recently begun to be noted more stridently and critically (Boysen, 2021; Murphy, 2021). This dissertation provides clear evidence of positive impact on learning outcomes for students who make use of the availability of options for learning content in different modalities, offering the kind of evidence for this aspect of UDL sought by policy advocates (Murphy, 2021). It also provides examples of the kind of research designs that are needed by more researchers seeking to investigate UDL in postsecondary settings (Crevecoeur et al., 2014; Smith et al., 2019). Additional research is needed to extend these findings to validate other elements of this complex framework by also seeking evidence of effectiveness for student learning.

This validation of the UDL guideline for providing multiple means of representation for students beyond those with disabilities suggests wide applicability of this approach in practice. This has implications for UDL theory and research about UDL overall since most empirical work to date has focused on students with disabilities. The framework's implications are broader than just this student subpopulation, however. Given the hidden and institutionally unidentified nature of disabilities for many undergraduates (Newman & Madaus, 2015), design anticipating the presence of a range of abilities, whether or not that full range has been specifically identified for the students at a particular institution, becomes imperative from a social justice standpoint (Bradshaw, 2020). The present research supports the argument that the framework has widespread applicability as an approach to help all students, particularly those struggling to learn for a variety of reasons, of which disability might be one.

This dissertation also has implications for theoretical directions regarding connections between the concept of learning styles and UDL. Recent critique of the UDL literature has noted that at times, UDL advocacy mirrors language and arguments used to promote matching instruction to learning styles (Boysen, 2021). Numerous learning style theories have been advanced over multiple decades positing different ways to categorize learners, and a vast educational advocacy and research literature has incorporated them (Cuevas, 2015; De Bello, 1990; Dunn & Dunn, 2005; Felder & Silverman, 1988). This learning styles research has typically focused on the idea that meshing a student's preferred learning style to the way content is presented to them in a one-to-one match will be beneficial for students (Pashler et al., 2009). Multiple careful examinations of numerous studies about matching students and material based on learning styles have concluded that this matching hypothesis lacks consistent supporting evidence (Aslaksen & Lorås, 2018; Cuevas, 2015; Pashler et al., 2009; Riener & Willingham, 2010) and should be considered a "neuromyth" despite it being widely believed by educators (Betts et al., 2019). Care needs to be taken to construct arguments about UDL that do not overlap with debunked ideas from the learning styles literature (Boysen, 2021). The entrenched belief in hypothesized benefits from matching in educational training and practice despite elusive evidence makes such caution particularly important.

The approach used in this dissertation goes beyond the basic matching hypothesis as considered in learning styles research, complicating thinking about what questions should be posed and how to conduct relevant research. It does not start from a premise that one-to-one matching of content and preferences is necessary or necessarily desirable. As described in chapter two, my research questions and hypotheses are informed by the cognitive science and multimedia literature (Clark & Paivio, 1991; Mayer, 2001; Mealor et al., 2016). Additionally, contrary to the learning styles research direction which argues for separating students into groups based on learning styles (Pashler et al., 2009), my research removed differences based on learning styles in chapter two by using a within-subjects design and in chapter three by conditioning on four aspects of a student's learning style (i.e., visual, verbal, aural, and physical styles determined by self-report).

The positive findings suggest more nuanced hypotheses beyond those that suppose matching or simplistic "best" ways of approaching material should be considered when investigating UDL. These results suggest that the multiplicity inherent in the UDL approach may be productive to investigate, perhaps more so than attempts at simplification. That is, rather than seeking averages that work best for everyone (reifying the evaluative value placed on the statistical normal curve that is problematically based in historical eugenics-related efforts), UDL encourages educators to become informed about the full range of possibilities that are needed for different individuals. Diversity of perception and diversity of processing become recognized as core values instead, and the question challenging theorists, researchers, and practitioners becomes one of understanding how best to incorporate that diversity into our understanding and planning rather than smoothing it out and problematically focusing on the "normal" center. While support for extremes of these ranges are needed for some students who cannot hear or see at all, for example, these functional extremes only become disabilities in the context of educational environments that do not anticipate and support them. It may also be the case that engaging different perceptual and processing capabilities provides benefit for all students, in contrast to the matching idea that would seek to limit engagement with different formats. A primary benefit of UDL's framing over the concept of learning styles as traditionally applied may be its focus on design. That is, understanding the variety of perceptual and processing capabilities that are inherently highlighted in UDL's framing may be highly important for instructional design by helping ensure that a range of options are consistently provided. This may hold more importance than identifying particular options as best in certain contexts. Such possibilities deserve additional investigation.

The present research suggests value in shifting the focus of research questions away from matching toward multiplicity, particularly in technologically adaptive settings. Presenting content in multiple ways within adaptive systems that limit what is initially shown to students would be in line with disability-focused research arguing that presenting too many options for content simultaneously may be overwhelming for some students (Beacham & Alty, 2006). This presents an intriguing, potentially important nuance to the idea that learning style informed teaching constitutes a neuromyth (Betts et al., 2019). The circumstances of application may matter. That is, although it may be a myth to think that identifying and matching singly presented content with a student's dominant preference is beneficial, perhaps there is more that should be learned about whether sequences or combinations might be fruitfully informed by learning style information.

Numerous researchers have posited that learning style information may be useful in customizing complicated technological systems that offer students multiple content options for learning presented in adaptive learning systems (Khamparia & Pandey, 2020). Such research typically has occurred independently of the research communities in which the matching hypothesis has been investigated in more traditional learning settings and debunked as a myth (Riener & Willingham, 2010). While often unstated in the computer science-oriented literature, the hypothesis that ordering information based on learning styles would be helpful appears to undergird a fair amount of research that has occurred with adaptive learning systems, though some research does recognize the limitations of this assumption (Costa et al., 2020). As one example, a small-scale experiment using an adaptive learning system where options for modality switching were available found several small, positive effects on learning in the context of a system tailoring presentation of initial material to a student's preferred learning style. While the authors did not explore this, given the kind of system they were studying, their results suggest that perhaps the order of modality use may matter (Mustafa & Sharif, 2011). It is unclear whether the results may have been influenced by students in that study using multiple modalities, which appear to have been available to them. The current study used a similar adaptive learning setup (although with insufficient time for the system to learn to tailor initial presentation mode), finding that combining learning activity in different modalities benefitted students. These potentially complementary results suggest that researching the ordering of modality combinations would be warranted. This may be particularly relevant for adaptive settings where technology facilitates presentation of material in different modalities rather than simply matching a single modality, particularly in terms of streamlining the content presented. Certain types of content known to be inaccessible to students who are blind or D/deaf, for example, could be blocked for those students and only accessible alternatives presented. The ordering of other modality types presented could be adjusted with the aim of identifying optimal sequences, perhaps considering customization based on both the student and the content. This type of ordered approach would go beyond simplistic, singular matching.

Such a direction would be consistent with group comparison research that found support for dual coding (i.e., visual and verbal) in brain pathways involved in sensing and processing, but not for the learning styles matching hypothesis (Cuevas & Dawson, 2018). In fact, in that study, recall was slightly better for learners whose learning style preference was crossed with their auditory or visual study condition. Their results indicate that making use of dual processing capabilities benefits learning beyond a one-to-one matching scenario. The results of this dissertation suggest benefit in going beyond crossing to deliberately making use of these dual pathways for encoding via multiple, differently formatted encounters with content to reinforce learning. This suggests we need more nuanced theory about the potential benefits of understanding learning styles in the context recommended by UDL where multiple options would be made available, also taking into consideration methods of content presentation and instruction that do not overtax cognitive load.

It might prove worthwhile to investigate using cognitive styles and learning styles information to inform our development of different types of content as well as our presentation ordering of the alternatives for that content, while also encouraging students to use more than one modality if they show signs of struggling to learn the content after their first attempt to grasp the material. The circumstances under which we might build hypotheses based on such scenarios might seem complex at first glance, but such a circumstance was straightforwardly found in the present dissertation. This research assumed that students may benefit from access to content presented in different ways and utilized technology that streamlined initial content presentation while not limiting a student to just the modality matching their preferred style the best. This implies that as our technologically based educational systems gain in complexity, our hypotheses about what works can still be straightforward when we place some of that complexity into the context of the study, as in my research. All the courses I studied existed within an educational ecosystem which utilized an adaptive learning system, so my hypotheses focused on one aspect of educational practice within such a seemingly complicated environment. The system was designed to reduce complexity for students though, hiding that complexity in the adaptive engine undergirding the system. My results suggest that expanding our thinking beyond what has traditionally been possible for a single instructor in a single classroom may correspondingly expand both our understanding of UDL practice and our understanding of what constitutes effective practices encouraging learning. This may help educators and course developers better understand the perspectives of those who perceive and process information differently than they do.

The results of this dissertation intriguingly suggest that additional research connected to actual student outcomes is needed to augment the intersection of UDL, adaptive learning system development, cognitive science, and learning styles. The causal graphical modeling approach discussed earlier offers an approach for structuring thinking about how current knowledge can inform future research possibilities targeted at intentionally increasing knowledge from a causal perspective. Next steps in this direction would involve considering and constructing a more thorough model than used here that represents additional possible explanations for effects on learning outcomes and what is known about directed connections between the posited factors. This could inform future research questions within a systematic research agenda that would extend beyond simplistic, unverified hypotheses about matching from the learning styles literature that seem to be seeping into discussion of UDL (Boysen, 2021) and could serve to develop working hypotheses about UDL that could inform needed effectiveness research (Crevecoeur et al., 2014; Smith et al., 2019). Thinking through possible alternatives and associated hypotheses would be worthwhile. The adaptive approach potentially offers students options when they need them without either limiting or overloading them. The present results suggest this area remains undertheorized and under-researched.

5.4 Implications for Research

Two choices intended to limit the scope of the current research have immediate implications for future research. Specifically, while neither looking at dosage effects nor variation by subject were investigated here, both areas provide logical directions for extension of this work.

Regarding dosage, the treatment studied here was a student's utilization of more than one modality while learning material from an activity designed to take the student about 20 minutes to learn (with appropriate extensions of the number of times multiple modalities were used for the week and course level analyses). In addition to this single dichotomous operationalization of whether multiple modalities were used, the effect of using two, three, or more modalities while learning the content, known as a dosage effect, could also be investigated with these data. This would determine whether a treatment effect of larger magnitude may exist at higher doses of different modalities, for example. However, the beneficial effect may not be strictly linear and it remains unclear where there might be a downturn in effectiveness. There may also be differences in how many uses of multiple modalities help the most for particular activities. Presumably there is a limit to how much using multiple modalities might provide benefit over the span of a week or the entire course. Future research could investigate the point at which such a relationship between modality use and grades might turn, the form of that relationship, and whether it might differ for students with and without certain types of disabilities. Knowing whether the effect is beneficial at different doses could be helpful when guiding students toward best practices for studying. Another approach to a dosage analysis would be to extend the preliminary look at the number of times multiple modalities were used during each week or throughout the course to a more formal analysis rather than the informal analysis conducted here.

Considering variation by subject, while the present research sought to identify an overall effect, and thus averaged across a variety of fields, patterns of modality use could also be disaggregated by subject or course in future research. This would allow more nuanced recommendations for future course development efforts. This kind of subject-based look at effects could help identify subjects and courses where the option to use different modalities appears particularly beneficial for students. As with the analysis in chapter four, this kind of information could help guide resource allocation, by individual faculty reviewing the courses they teach, by departments reviewing curriculum, and by instructional design teams assisting with curriculum revisions.

The analysis in chapter four offers additional implications for future research. The kind of Bayesian network technique employed in this analysis could be straightforwardly extended in multiple ways, with both theoretical and practical implications. For example, predictions of points at which tutoring may be beneficial could be extended throughout the entire course instead of restricting analysis to a given week. Additionally, some courses taught at this institution are sequenced together, one being a prerequisite for another, while others stand alone. In future analyses, it would be possible to map connections between courses when applicable. For example, there is a prior introductory English course that is sequenced with the one studied here and some of the learning activities between these courses connect to each other. Data from these courses could be modeled and analyzed together to improve predictive ability. Such integrated analysis extending the internal predictive ability of the adaptive system could be extended throughout an entire program of study to help identify interconnected points of difficulty whereby students who struggle learning particular concepts in prerequisite courses later struggle with content in a subsequent disciplinary course. This type of network analysis incorporating data beyond the adaptive system holds the intriguing possibility that prescriptive analytics could be used to tailor help in the current course based on difficult learning concepts in a previous course where the student moved on prior to fully mastering important material.

As with other learning analytics research, the results of such prescriptive analysis should be integrated into an easily digestible presentation for students (Clow, 2012), which would be a logical extension of the current research. For example, this could be done through a dashboard display that succinctly indicates when tutoring is predicted to be most beneficial throughout the course to assist informed learning decision making by either students or faulty (Jivet et al., 2018). Such work could be extended further by investigating communication practices that might follow tutoring recommendations presented to students (Kitto et al., 2017), such as academic coaching interventions supporting dashboard or email-based analytics information with a human connection.

Another direction for research given the limitation posed by the small number of known students with disabilities in this study would be to extend this research at this institution to determine the extent to which students self-identify as disabled (Fox et al., 2021). Conducting survey research among the students could reveal the extent students might have had accommodations in the past or might encounter difficulties they did not know they could get help with. Additionally, such a survey could ask them how the adaptive learning system helps them learn in noticeable ways. If they self-identified as disabled, more could be asked about the extent to which their specific educational needs are being met by the adaptive system. Such research could offer insight into the context studied here that is centrally relevant to the UDL framework investigated.

Looking beyond this institution, although the present research did not include many students with known disabilities, the panel data analysis technique used here could be employed with data from other institutions using adaptive learning systems with larger populations of students with known disabilities to confirm the magnitude of the effect for these students in other contexts. Such research can extend what is known about causal effects of UDL, a direction that is needed overall in the UD research field (Rao et al., 2014).

5.5 Implications for Analytical Practice

My dissertation utilizes data and online technological advances in the service of supporting traditionally underrepresented students' success. My analyses demonstrate the potential of using very large datasets, data integration across campus systems, and advanced statistical techniques augmenting the descriptive dashboard analytics presentations that are becoming increasingly available through technology vendors.

Chapter two's study design merges a UDL framing with rigorous statistical analysis using data from multiple sources, generating a larger dataset (almost 200,000 observations) than is typical for much educational research. When initially working with these data, I encountered processing and memory limitations despite using a high-end personal computer with Stata MP. Thus, while these data would not be categorized as "big data" by data scientists (Sin & Muthu, 2015), they necessitated approaching data processing from a thoughtful perspective that took processing capacity into account when cleaning and formatting the data for the analyses. My dissertation presents an example of using a large, integrated dataset to conduct multiple analyses to sharpen inference, improving the internal validity, precision, and accuracy of findings.

Although much continues to be learned about how to improve instructional design to facilitate learning, future research efforts in this area could be bolstered by utilizing data increasingly made available by educational technology systems (Smith et al., 2019). Learning management systems have become ubiquitous across higher education (Dahlstrom et al., 2014), and online presentation of content has been actively developed by publishers, institutions, and learning-oriented technology vendors. Adaptive learning systems are less broadly implemented, but their use is growing. However, to date, the field has generally lacked appropriate data about students' use of multiple modalities to improve learning, limiting researchers' ability to conduct causally oriented investigations. The same has been true for tying other aspects of UDL to student outcomes. Although some prior studies have investigated samples of hundreds of college students (Mangiatordi & Serenelli, 2013), challenges associated with obtaining data appropriate for investigating tenets of UDL theory have led to recommendations to pursue action research at a smaller scale as one way to increase information about UDL's efficacy (Smith et al., 2019). Instead, my dissertation provides an example of scaling up effectiveness research about UDL.

The research presented in both chapters two and four utilizes data that would potentially be widely available to institutions using adaptive learning systems assuming modality information could be coded for the activities. As institutions increasingly collect data from multiple systems in what have been termed data warehouses and data lakes, more sophisticated analyses drilling down into aspects of learning activity such as have been presented in this dissertation will become increasingly possible at this and other institutions. The type of analysis in chapter four would be feasible when learning activities have been clearly articulated and mapped as they typically would be when populating an adaptive system. This means that many institutions have the potential to analyze data from their own online course material as has been illustrated here. Since adaptive systems inherently gather data about student learning activity, figuring out how to make ethical use of such data simultaneously poses an opportunity, a challenge, and an imperative given that thoughtful use of such systems' data can clearly provide beneficial information for and about students who may typically struggle.

For the analysis conducted in chapter two, the causal graphical modeling work suggested that future research could begin to investigate mechanisms involved in the operation of the effect of the use of multiple modalities, including both the effect of time on task and the effect of repetition of material. While full mediation analysis was beyond the scope of this dissertation, the magnitude of such potential alternative explanations could be explored in future research. Similarly, other potential alternative explanations for the effect as explicated in earlier chapters could be explicitly modeled to facilitate identifying which relationships would warrant future causally oriented research.

Given numerous challenges faced when conducting research about UDL (Rao et al., 2020; Smith et al., 2019), this kind of graphical modeling approach could help coordinate further UDL effectiveness studies to clarify currently obscure connections. This approach has seen only sparse use to date in higher education. That use has been more in areas such as educational technology (Xenos, 2004) or intelligent tutoring systems (Pardos et al., 2006) than directly in the higher education field's literature. The research design employed here presents a higher education example of using causal graphical models with directed acyclic graphs in causally oriented modeling, particularly for fully online courses and for courses that include a significant technological component that leaves data traces. My research offers a guide for empirical research analyzing the effectiveness of use of multiple modalities as well as other course-level UDL design factors in the future, both at this institution and in other contexts. Combining results from similar studies at other institutions could lead toward more generally applicable knowledge about UDL as future research continues in this area.

This approach to causal analysis could be developed as part of a future research agenda oriented toward investigating additional aspects of UDL (Smith et al., 2019). It would facilitate identifying which of several clearly articulated possible alternative models is more likely given what we have observed. By using it with a larger network of variables than explored here, the approach can also help identify fruitful areas for future experimental or quasi-experimental research where relationships could be further clarified. Learning models from data with these variables could assist in identifying which relationships can be directionally determined and which need more data or different research designs to determine causal directions. Specifically, targeting unclear areas where current results do not allow explicit distinction between models would help identify appropriate hypotheses to test with future research designed for that purpose.

By exploring inferential insight about student progress, I target institutional stakeholders' ability to improve course delivery and intervene meaningfully for students when and where it matters most. Findings offer an example of using a learning analytics perspective that could inform recommendations to students for tutoring and suggest that such information could inform faculty and instructional designers when targeting future course development. The challenges associated with designing high quality, causally-oriented quantitative studies of educational outcomes (see Song & Herman, 2010) make this study particularly useful as a guide for analytical practice in future research efforts.

5.6 Implications for Institutional Practice

Due to the historically individual nature of the postsecondary teaching field and the tradition of maintaining a high level of faculty control over individual classes, research on teaching practices at an institutional level typically have faced significant practical challenges. Based on their review of the literature, Lawrie and colleagues (2017) called for institution-level research of the sort conducted here about inclusive strategies to foster holistic change:

> In line with Hockings' recommendations from 2010, additional scholarship focusing on institution-level initiatives is still merited, as work focused in this way remained relatively sparse in the materials reviewed for this research. While several authors have offered perspectives on multiple meanings and dimensions of inclusive learning and teaching, examples of whole of institution approaches remain rare. (p. 9)

The institution studied in the current research offered an all-too-rare opportunity to analyze a significant fraction of the courses offered across the online undergraduate curriculum and report the results beyond institutional stakeholders in institutional research, academic planning and assessment, and academic management. As analytics initiatives grow across higher education, such institution-wide research at the course level would ideally become more commonly practiced. Institutional-wide research within courses, ethically conducted, would extend the research base for the developing national UDL research practices and agenda (Rao et al., 2020; Smith et al., 2019). The present study illustrates the power that institutional resources can have to move knowledge in the field forward. The high rigor level of this study would not have been possible without significant resources devoted by the institution (aided by federal funding through the FIPSE grant), including widespread interest in and support for analytics and a commitment to making use of data to inform practice. It also illustrates a potential direction for learning analytics practice that makes greater use of combined data warehoused from across multiple systems. This dissertation models within course analyses that could be conducted at other institutions devoting resources to large scale course design and analytics.

I can confidently state that the aspect of UDL encouraging options for presentation benefits students at this institution with respect to their demonstration of content mastery later in the week across many types of courses, and this has implications for faculty development practice. Teaching through different modalities makes a noticeable difference for students. As found in chapter two, an improvement index of +9 above the median in formative knowledge gain came from facilitating student use of modality alternatives like video or interactive exercises based on Open Education Resources (OER) that faculty and other course developers at any institution can access (Porcello & Hsi, 2013). While it takes time to design and build out alternate course material utilizing OER, this type of design improvement can be straightforward to implement and has been shown to be an effective way to provide content in different modalities (Navarro et al., 2016). The research reported here suggests this type of effective course design is worth incorporating into faculty and course development efforts.

The overall results about the effectiveness for learning can therefore also inform instructional design (Tobin & Behling, 2018). The UDL aspect of offering options for perception now constitutes a recommended practice with empirical support for its effectiveness for student learning. Although it is not yet clear how much of the two standard deviation quiz effect found in chapter two may be due to idiosyncratic institutional factors that would not translate easily to other contexts, the combined effect for assignments and quizzes was large and seems likely to apply in other contexts as well given the research design. Since the present study investigated many different course subjects, the effect seems to be present across multiple areas. This provides encouragement about possible broader applicability of these findings, since it is already known that the average effect is meaningfully large beyond just a single course context. Thus, this UDL-based approach seems worth incorporating into instructional design activities across subjects at an institutional level.

Building on the knowledge from chapters two and three that using multiple modalities benefits students in their learning, chapter four illustrated how such knowledge could lead to recommendations regarding additional tutoring assistance. Additionally, as illustrated in Figure 1.1, these results hold potential to inform an additional course-specific feedback loop for faculty, course developers, and instructional designers working on course revisions. As noted by Gagliardi, "The use of analytics typically focuses on intervening on behalf of students who need support or are at risk, but more power lies in using analytics to identify structural flaws in programs…" (Gagliardi, 2018). In addition to implications for students, the information about difficulty points identified by data patterns and Bayesian network simulated intervention analysis holds implications for course development modifications. At these points, students show signs of needing additional assistance. In the moment, tutoring may be one effective means of providing support. However, beyond the current semester, identifying these points holds potential to flag where prioritizing content revision may be beneficial. This could include revising existing material to improve the quality or clarify the explanations, adding new content material in different modalities, or considering other aspects of UDL that might aid students at that point in the course.

At an institutional level, figuring out more successful and cost-effective approaches to meeting academic needs of currently underserved students offers institutions significant growth potential (Christensen & Eyring, 2011). This poses an increasing concern for higher education leaders given shrinking state support for higher education and increasingly unignorable demands for affordability by students and their families (Bears, 2018; Heller, 2001). Attention to effective learning strategies for students that improve their learning outcomes and reduce their overall costs and time-to-degree continue to be sorely needed. Institutions who serve needs of students with a wide range of ability levels in a cost-effective manner stand to grow significantly (Christensen & Eyring, 2011). Effectively utilizing adaptive learning and analytics in the context of an institution committed to a high level of student support holds potential to position an institution competitively in the market compared to rival institutions who have not yet developed such resources and processes given values supporting such innovation (Christensen et al., 2004). UDL challenges institutions to design learning experiences intentionally including multiple means of achieving key elements for facilitating learning (Meyer et al., 2014). By contributing to our understanding of the effectiveness of the use of multiple modalities, this dissertation makes clear that paying more rigorous attention to research design strategies can help verify noticeable improvements to our students' learning success in innovative ways.

5.7 Conclusion

This dissertation research suggests that even people who do not have diagnosed disabilities requiring accommodations benefit from the practical implications of course design offering multiple means of presenting content. This practice perhaps provides the field with a practical educational analogue of the physical curb cut. When considering the practical implications of disability-related adjustments to the physical environment, the example of curb cuts being useful for parents with strollers, travelers rolling luggage, or deliveries being rolled to stores from trucks is easy to understand and iconic. The argument for the widespread usefulness of this adjustment inspired by the physical requirements of wheelchair users has face validity as something that holds potential to benefit everyone. That is, curb cuts provide essential access to some people with disabilities while providing benefit to all.

In a similar manner, the present research corroborates the intuition that making material available for learning in multiple formats holds the potential to benefit all learners. I acknowledge the possible critique of the curb cut analogy potentially hiding disability under ableist norms that do not acknowledge disability. However, I believe the analogy provides benefit by increasing disability awareness, simultaneously broadening the argument for implementing practices that are widely beneficial in a way that can be successfully used to support people with disabilities (Rao & Tanners, 2011). The educational practice of offering options for content via different modalities provides essential access for those students who have functional sensory or cognitive impairments who need content presented in alternate formats while providing benefit to all. In fact, students from a wide array of backgrounds and abilities can benefit from this practice. Now it is possible to better quantify the benefit obtained for all when resources are put toward this practice. Doing so just makes sense.

This research suggests that processing content information in more than one modality provides learning benefit. Such an effect goes beyond simply providing access to content to actual improvement in learning. While it is difficult to say conclusively what the benefit is for students with disabilities from these results due to the tiny number of students with disability accommodations in these courses, given the assumption that students with disabilities exist in the sample even though they are unidentified, it seems plausible that students across the ability spectrum benefit from this practice designed for students with disabilities.

Additionally, since this study was conducted at a women-only institution, gender is an inherent aspect of the study, though not from a comparative standpoint. Given the possible low need for accommodations, it may be the case that such practice helps to offset other disadvantages students may face due to other aspects of systemic inequality, including the intersection of ability with other identities possessing additional oppressions, such as gender (Bradshaw, 2020). Such a supposition deserves further inquiry since it remains inconclusive from the present evidence though the results are suggestively intriguing. If so, the implications of such practice may be far reaching. Just like curb cuts have become a ubiquitous aspect of the urban landscape, having alternatives for content available may one day become standard educational practice facilitated by technology. Adaptive technology offers the possibility of designing such options in ways that provide sufficient support without overwhelming students, only revealing options when students' demonstrated struggle indicates that they need additional support.

By investigating a key aspect of UDL theory's tenet of providing for multiple means of representation, this dissertation offers a needed contribution to both the UDL and higher education literature. Empirical research about how to support the academic success of students with disabilities continues to warrant extension in higher education (Kimball et al., 2016). This dissertation contributes to knowledge about one aspect of a framework intended to support these students, substantiated by data showing a widespread benefit among all students. My research averaged over several dimensions, such as subjects and courses, as well as (unknown) student ability, seeking an overall effect. Future research should expand this to a distributional view, quantifying the magnitude of this benefit for students along the broad spectrum of functional ability. Although arguments for application of UDL have been more numerous than careful research of its components' effectiveness by the higher education community (Edyburn, 2010), the studies presented here are part of an ongoing effort by the UDL community to change this research dynamic (Rao et al., 2020; Smith et al., 2019). By isolating one aspect of UDL in my research, I advance understanding of a practice that has the potential to help not just students with disabilities, but all students.

Connecting the usage of multiple content representations to student learning outcomes and course success, this study's analyses searched for multiple indications of a possible causal relationship at both formative and summative levels of analysis. Using data from online courses in multiple fields at a single institution, along with abundant data from connected student systems, I employed several methods to bolster the internal validity of my findings. These included a panel data analysis and two instrumental variables analyses which found statistically significant and substantively important effects of the use of multiple modalities for learning content on student learning outcomes. These effects included knowledge gained across a learning activity, average weekly assignment and quiz grades, and likely final course grade as well. I also investigated clusters combining modality switches and tutoring, and presented an example Bayesian network analysis that connected these combinations to course activities, providing an example for identifying points in a course that are key for subsequent student success. This type of analysis offers the possibility of indicating both where course revision may be warranted and when intervening to recommend tutoring might be particularly beneficial for students given the existing presentation of course content. This would be equivalent to gathering data from streets in a city that still is full of curbed sidewalks that would help urban planners identify where installing curb cuts would be most immediately beneficial in a practical sense for citizens, on the way to installing curb cuts everywhere throughout a city. The goal here may be to offer options for perception throughout a course in a way that is responsive to the needs of different students, but a plan for improvement is needed to get there from where we are. The type of network analysis approach offered here holds potential to be developed further to guide such practice, both to identify support for students in an immediate sense and to provide direction for longer term systemic improvements through course redesign efforts.

To summarize, the results from this study provide support for UDL's proposition that providing multiple means of representation through options for perception is beneficial for student learning. This positive effect is seen for formative learning activities as well as weekly assignment and quiz grades; it is likely present for course grades too. The analysis also provides a proof-of-concept demonstration that a Bayesian network approach holds potential to assist course developers in targeting course material revision where those efforts have the potential to make the biggest improvements in student success as well as to identify where tutoring assistance for students can be most effectively targeted in the meantime until sufficient resources can be directed toward implementing such systemic improvements.

APPENDIX A

ADDITIONAL MATERIAL FOR CHAPTER 2

Additional details pertaining to the analysis in chapter two are provided below.

A.1 Variable Description

Table A.1 Variable Operationalization

Variable	Operational Notes
Knowledge state gain across activity	Continuous. Students' knowledge of the material was assessed at the beginning of each week, and a starting knowledge score was assigned. As students went through the content for the week, at the end of each learning activity (each designed to take about 20 minutes), students were given a short (e.g., 5 question) assessment. Their knowledge score was updated after each activity based on their performance. The value of this variable is the difference between starting and ending knowledge scores across a learning activity.
Use of >1 modality	Binary. Each learning activity was offered in a variety of formats, including text, video, audio, interactive, and mixed modalities. Course designers identified Open Educational Resources (OER) in as many of these modalities for each activity as possible. Each time a student accessed material for a given activity, the modality accessed was logged. 1 = student used more than one modality while working on a particular activity; 0 = used only one modality.
Hours spent on activity	Continuous. The time a student spent on an activity across all modalities was logged. To keep the model variables' order of magnitude similar for ease of interpretation, this variable represents the time spent in fractions of an hour. (While intended to take 20 minutes, students spent a mean time of 0.124 hours = 7 minutes per activity.) Time was top coded at 10 hours for 62 outlier cases (with time spans over 10 hours up to 13.6 days) in line with adjustments made automatically for 7 cases by the adaptive learning system. These adjustments captured students who left their computer with an activity open, for example. Cases shorter than 2 seconds were dropped under the assumption that they were not very meaningful for this analysis and do not represent students engaging the material.
Session	Binary. Classes across two course sessions in Spring 2018 were analyzed, and each was given a separate indicator value.

Table A.2 Correlation Matrix for Analysis of Knowledge State Gain Across Activity

	(1)	(2)	(3)	(4)	(5)
(1) Knowledge state gain across activity	1.000				
(2) Use of >1 modality?	0.153	1.000			
(3) Hours spent on activity	0.039	0.109	1.000		
(4) Fall 2018	-0.001	-0.002	0.022	1.000	
(5) Spring 2019	0.001	0.002	-0.022	-1.000	1.000

A.2 Additional Methodological Considerations

This section explains two methodological considerations that may be of interest to technically oriented readers. These include the low R^2 value and the possibility of studying a hierarchical linear model.

As shown in Table 2.3, the R^2 value ranged from 0.009 for the panel data analysis, to 0.023 for the regression analysis taking clustering by student into account. When including time spent on the activity and the semester in the fixed effects panel analysis (results not shown), the R^2 improved from 0.009 to 0.023, like the ANOVA and regression analyses, but the coefficient for use of more than one modality was unchanged. These very small R^2 values suggest other factors, such as course material quality or activity difficulty level, have much larger influences on learning gains but were unmeasured and not included in the model. Additionally, it is unclear how other non-student-level factors that are known to influence student achievement, such as teacher involvement (Ayllón et al., 2019), operate in adaptive learning contexts on the scale of single learning activities that are not designed to have direct instructor contact while the student completed them, and were not part of this study. However, the low R^2 values are not a problem for the present analysis, which is focused on capturing the treatment effect, rather than explaining variance in the outcome. The panel data approach used here isolates the variance in the treatment due to non-student-level factors. This means that research aiming to explain non-student-level influences on learning gains should investigate factors beyond the treatment studied here. However, for the purpose of this study's investigation of the effect of use of multiple modalities, the model was deemed sufficient despite the low R^2. For this type of treatment effectiveness research, the effect size of the treatment of interest is more important than the amount of outcome variance explained.

Given the clustering that occurred by person, activity, week, and course in these data, I considered whether a hierarchical linear model (HLM), also known as a multilevel model or mixed effects model, might be appropriate (Raudenbush & Bryk, 2002). HLM would be useful when one expects that important variation may have occurred across groups (i.e., there was sufficient intraclass correlation). While HLM could potentially be applied in several ways to this project, I have not chosen to do so at this time for the following two reasons.

First, although these data could be viewed from a growth curve model perspective, where multiple observations across time were contained within an individual, this presupposes that the variation I attempt to model was meaningfully differentiated by individual. For the analysis presented here, I viewed differences across individuals as important heterogeneity across which I wished to average. That is, I sought an overall effect of use of multiple modalities net of unobserved person-level characteristics rather than seeking to model an effect that differed by individual. However, I acknowledge that this alternate approach could provide an interesting avenue for future exploration with a slightly different research purpose.

Secondly, while HLM could be considered for application to these data to model clustering at the activity, week, or course levels, in this instance, these were fixed characteristics of the learning environment encountered by a given student. That is, sampling variability occurred only at the student level. Assignment to course and instructor were fixed once a student registered, and instructors typically only taught one course a given student was enrolled in during a particular subterm. Assignment to modules occurred in a fixed sequence per week. This meant it was more appropriate to consider week and course as variables for inclusion in the model than to conduct a multilevel analysis. Given that my primary interest in this study was investigating an overall average effect across multiple fields, rather than identifying the effect within particular courses or weeks within courses, I chose not to include these variables in the model. Additionally, results from a sensitivity analysis that clustered by the activity being completed along with accounting for the student were substantively similar to the results presented for clustering within student alone. Thus, while I accounted for clustering in several ways to compare results from different techniques, I chose not to use a multilevel model, although I acknowledge that future researchers may choose to investigate this phenomenon from a different perspective.

A.3 Sensitivity Analyses

I conducted additional analyses to probe how sensitive my results were to choices about treatment operationalization, covariates, missing data, and modeling approach.

I checked two different operationalizations of the treatment variable to verify the influence of the “mixed” category on the results. The mixed category consisted of a combination of the other modes within a single activity part, and I treated this as a separate modality in my main analysis given that multimedia is known to have an effect on learning that is separate from single modality learning (Mayer, 2001). In the situation studied, the mixed nature often consisted of material in several modalities presented as parts of the same activity, and my assumption is that this material will be processed by students slightly differently than material presented in a single modality, and thus it would be relevant to analyze the mixed category separately. To probe this assumption, the additional analyses used the same dataset in the same way as my primary analysis, with the difference that the first alternate analysis considered the mixed category as inherent use of multiple modalities instead of its own category as one of five possible modalities, and the second dropped the mixed category from the data prior to analysis. The first alternative had a slightly lower effect (g = 0.21), while the second had a slightly higher effect (0.25), but both led to substantively similar conclusions to the main result presented (0.22).

Given that task repetition might be an alternate explanation relating to the effect of the use of multiple modalities, I conducted a Bayesian network analysis to learn the structure of a five variable model including the four variables in Figure 1 plus a dichotomously coded variable indicating whether the activity was repeated by the student. Results were very similar in nature to the results from the model in Figure 1, suggesting that repetition may also partially mediate the effect of the use of multiple modalities and that time may also mediate the effect of repetition. Given that explaining the mechanisms at work through a mediation analysis is not my aim here, it was sufficient for the purpose of this study to confirm that the model learned still contained a directed arrow from treatment to outcome even considering the presence of a possible additional alternative explanation.

It is worth noting that top students not receiving a recommendation to use multiple modalities by the adaptive learning system who wanted to improve their already high score further might have chosen to go through the material again using a different modality. This seems unlikely to have been an issue for treatment assignment in the panel analysis since such a factor would have been innate to the student and thus accounted for in the study design. However, this possible source of unwanted treatment assignment variation was investigated further for the regression results using a latent factor of student motivation to achieve a high grade. This factor was based on data from the first two weeks of the session and was used to condition a clustered regression model for data from the last four weeks of the session. Indicators included means across the two weeks (standardized against the class average) of completed classwork grades, prior on-time assignment submission, mean number and length of weekly discussion posts, and mean time to complete course activities in the adaptive system ($\alpha = 0.7$). Results for use of multiple modalities were similar to the primary regression-based result, suggesting that such grade motivation may play a minimal role in treatment assignment and is likely not a large concern for this analysis. However, I acknowledge that this was likely an imperfect measure of students' desire for high grades and future research could investigate this issue further.

In addition to this grade motivation variable, I included six other demographic and prior education variables in the OLS regression sensitivity analysis. These included the number of failures or withdrawals in the previous semester, overall grade point average at the semester's start, number of credits transferred in upon entry to the institution, race/ethnicity, age, and Pell grant status. However, since the inclusion of these covariates did not substantially alter the results, I only present the more parsimonious OLS results without these covariates.

As a sensitivity check on the impact of dropping the cases missing on the dependent variable, the missing knowledge state scores were multiply imputed and the gain subsequently calculated (Enders, 2010). This allowed incorporation of the uncertainty that exists due to the missing data while using all available information without dropping cases, lending interpretive confidence to the results (Manly & Wells, 2015). To improve the common assumption that the sources of missingness were due only to observed variables and thus Missing at Random (MAR; McKnight et al., 2007), demographic and educational covariates were included in the imputation process, including race (2.2% missing), age (1.2% missing), Pell grant status (as a proxy for socioeconomic status; 0% missing), number of credits transferred upon entry to the institution (2.6% missing), number of prior semester course withdrawals and failures (0% missing), career grade point average prior to the course (0% missing), and whether the student was in a science, technology, engineering, or math (STEM) major (0% missing). Including these conceptually logical observed variables in the imputation process meant the data were dealt with reasonably for the MAR assumption. For this analysis of sensitivity to the presence of missing data on the dependent variable, since the imputation model included a mix of continuous and categorical variables, the chained equations approach for MI was utilized (van Buuren, 2012) using Stata's *mi impute chained* command with M=40 imputations since the largest fraction of missing information across the analyzed imputations was *FMI*=0.39 (Graham et al., 2007). As is typical in MI analysis, Rubin's (1987) rules were used to pool results across all imputations. Observed and imputed values were checked and compared reasonably, and analysis results were checked for substantive discrepancies between MI and LD. While results indicated a slightly larger effect (0.063***, Hedges' g = 0.286 and improvement index of +12.9), substantively similar conclusions would be drawn. While I focus on listwise deleted results for ease of testing and interpretation, the multiply imputed results support the findings presented.

As a final sensitivity check, I altered the model chosen slightly, running an OLS analysis that absorbed the student terms (which is mathematically equivalent to adjusting the analysis with a dummy variable for student) as well as adjusting for clustering by activity, using Stata's *areg , absorb(id_student) vce(cluster id_activity)*. Since the results did not provide new information not already provided by other models, and were almost identical to the final panel model, these results were omitted in favor of the conceptually simpler model and for simplicity of presentation.

APPENDIX B

ADDITIONAL MATERIAL FOR CHAPTER 3

Additional details pertaining to the analysis in chapter three are provided below.

B.1 Variable Description

Table B.1 Variable Operationalization

Table Abbreviation	Variable	Operational Notes
Week grade	Weekly assignment and quiz grade	Percentage. All grades earned by a student in each week on LMS gradebook entries labeled assignments, homework, problem sets, labs, comprehensive problems, and quizzes were summed. This sum was divided by the number of points possible for these gradebook items for that week to get the student's mean grade for the week.
Course grade	Course grade	4.0 scale. Official course grade earned by the student.
≥# modalities	Use of >1 modality for more than a specified number of times during the week or course	Binary. Each time a student accessed material for a given activity in the adaptive learning system, the modality was logged. These modality uses were summed across each week and across the entire course. Different treatment variables were created with a specified number of modality uses as a minimum threshold. The primary week level treatment was seven uses of multiple modalities, so any student whose count of use of multiple modalities on activities during a given week was seven or greater had a value of one for this variable. The primary course level treatment was 38 uses, so students with a sum of 38 or more had a one.
Race-White Race-Black Race-Hispanic Race-Other	Race/ethnicity	Categorical split into binary indicators.
Age	Age	Continuous. In years.

Table Abbreviation	Variable	Operational Notes
First gen	First-generation student	Binary. Self-identified as a first-generation college student in the institutional administrative database.
Pell	Pell grant eligible	Binary. Student is eligible for a federal Pell grant.
Xfer credits	Number of credits transferred at entry	Continuous. Official number of academic credits earned at other postsecondary institutions upon entry to this one.
Base Pell	Mean baseline Pell by cohort	Percentage. Mean baseline of the Pell status variable calculated by RCT cohort. Calculated separately for each dataset analyzed.
Base credits	Mean baseline credits transferred by cohort	Continuous. Mean number of transfer credits calculated by RCT cohort. Calculated separately for each dataset analyzed.
WF count	Count of prior semester W and F grades	Categorical.
GPA	Career GPA	4.0 scale. Official GPA at the start of the session.
Grade motiv.	Motivation to achieve a high grade score	Continuous. Latent factor score derived from a principal components factor analysis of two indicator variables (discussed below).
Verbal LS	Verbal learning style score	Percentage. Score from the SmarterMeasures survey administered when the student entered the RCT.
Visual LS	Visual learning style score	(Same as above.)
Aural LS	Aural learning style score	(Same as above.)
Physical LS	Physical learning style score	(Same as above.)
Life score	Life factors score	(Same as above.)
Tech score	Technical competency score	(Same as above.)
Session	Session	Binary. Classes across two course sessions in Spring 2018 were analyzed, and each was given a separate indicator value.
Cohort 1	RCT Cohort Fall 2015	Binary. RCT starting cohort term 1 (Fall 2015)
Cohort 2	RCT Cohort Spring 2016	Binary. RCT starting cohort term 2 (Spring 2016)

Table Abbreviation	Variable	Operational Notes
Cohort 3	RCT Cohort Fall 2016	Binary. RCT starting cohort term 3 (Fall 2016)
Cohort 4	RCT Cohort Spring 2017	Binary. RCT starting cohort term 4 (Spring 2017)
Cohort 5	RCT Cohort Fall 2017	Binary. RCT starting cohort term 5 (Fall 2017)
Cohort 6	RCT Cohort Spring 2018	Binary. RCT starting cohort term 6 (Spring 2018)
RCT group	RCT group status	Binary. Randomized into RCT treatment group=1.

B.1.1 Dependent Variable Sensitivity Analysis

The outcomes studied included two summative measures of the student's understanding of the course material: a) mean grade on the week's assignments and quizzes, and b) overall course grade. Course design included assignments relevant to each week's content due at the end of the week. The weekly grade outcome included the mean of grades on assignments, problem sets, labs, and quizzes. Other graded work was not included (such as discussions, response papers, presentations, mid-term exams, final exams, and final projects) because either the nature of the work was conversational, or the timing was not week-based. Under the supposition that grades on weekly quizzes may have been a more focused measure of content mastery than grades on other types of assignments, I investigated the sensitivity of my results to the choice of outcome, decomposing the week outcome into quiz grades and other assignment grades.

Students who received zero points for a week's assignments or who failed the course (earning zero credit) may have encountered external circumstances potentially confounding the effect of the treatment of interest. Given that such students' performance may be qualitatively different than students who received a grade, I focused on students who received a non-zero grade to clarify the treatment effect for students who did not encounter extenuating circumstances. However, I also conducted a sensitivity analysis including these zero grades for comparison at both the week and course level to investigate the extent to which the treatment may assist students in overcoming external barriers to success.

B.1.2 Independent Conditioning Variables

In causally oriented analysis, ideally one would condition on a pre-treatment measure of the outcome, but this was not possible for the grade outcomes used here. Instead, as a reasonable alternative per What Works Clearinghouse (WWC; 2017) recommendations, I aimed for treatment and comparison groups similar on measures of socioeconomic status and prior education. Accordingly, federal Pell grant status was used as a proxy for low socioeconomic status and the number of credits transferred from another institution upon entry to this institution was used as a proxy for prior education. The equivalence of these variables between treatment and comparison groups at baseline prior to treatment was evaluated using WWC criteria. The necessary statistical adjustment was achieved by including the mean of both Pell status and transfer credits per RCT cohort (to capture any cohort effects), as well as their individual values. Indicators of RCT cohort were also included as covariates.

Analyses were run with this minimal set of covariates as well as a full set of independent conditioning variables, which also included the session in which the course was taught, prior educational information, student demographics, and personal characteristics. Variables related to a student’s prior education included the number of failures and withdrawals in the previous session, as well as overall grade point average.

Relevant demographics included race/ethnicity and age. Personal characteristics included several scale measures from the SmarterMeasure survey that students filled out at the beginning of the RCT (SmarterServices, 2021). These standard scales, intended to gauge readiness for online learning, included four learning styles measures gauging preference for verbal, visual, aural, and physical learning. While in an adaptive system, material can be made available to students in a variety of ways designed to enhance comprehension, incorporating demonstrated learning preferences (Kang et al., 2017), here, learning style information was conditioned on rather than used to jump-start adaptivity (Mustafa & Sharif, 2011). The SmarterMeasure scales also included a measure reflecting life factors impacting the ability to do well in an online course, and a measure of technological competence relating to skills that might impact ability to succeed online. These covariates were included with the intent of increasing the precision of the estimates.

A score representing a latent variable was constructed from two indicators of motivation to achieve a high grade using a principal components factor analysis (*factor, pcf* in Stata). Indicators included: a) the student's grade on all first week graded work, standardized against the class; and b) the number of on time classwork submissions during the first week, standardized against the class. Three measures of reliability were checked and found to be reasonable, including Cronbach's alpha ($\alpha = 0.76$), composite reliability (CR = 0.7), and average variance extracted (AVE = 0.5). This variable can be interpreted as gauging how a student compares to her classmates in terms of motivation to do well in the course. Additional potential indicators that might have been used such as number and length of discussion posts during week one were unavailable for these RCT data for this analysis.

B.1.3 Missing Data

In general, missing data were not considered a major threat to the validity of results for this study, as most variables had little to no missing data (1.3% missing for race/ethnicity, 3.6% for visual/verbal/aural/physical learning orientation, 3.3% for the life factors score, 4.5% for the technical competency score, and none missing for others). Missing data were handled as described below to check their influence on results.

Some missing data posed no problem because they were missing legitimately or randomly. Given that course assignment structure differed, some courses had no grades for certain types of assignments in given weeks, and so only cases for the assignment types of interest were analyzed. Thus, there were no missing data on the outcome. Additionally, students dropping the course would result in attrition across the study, and since they legitimately earned null grades after dropping, such students were not analyzed. However, overall attrition was analyzed separately to gauge whether this was a concern when interpreting results. While students persisting in the course would most likely have completed at least some work within each week, some students might skip a week and then return to the course. Such sources of missing data are most likely due to life events outside of the course which were assumed to be either missing completely at random (MCAR) and therefore not of concern for analysis, or missing at random (MAR) once covariates were controlled for (McKnight et al., 2007; Rubin, 1976), since such life events may be correlated with other personal factors such as income level, race/ethnicity, or the life factors scale score. It was further assumed that other sources of missingness were also missing at random (MAR) once covariates were controlled for, and while this assumption was untestable, it was deemed sufficiently reasonable here for handling missing data by multiple imputation.

Analyses investigating the sensitivity of results to the choice of handling missing data were conducted, comparing results by multiple imputation to those by listwise deletion (Manly & Wells, 2015; Schafer & Graham, 2002). Listwise deletion was considered reasonable given the generally small amount of missing data, and after checking various tests for endogeneity and the strength of the instrument. However, missing data were primarily handled through multiple imputation to allow the inclusion of whether the student was the first in her family to attend college as a covariate in the analysis. This was a desirable variable to include since it represented an otherwise uncaptured element of socioeconomic status, a background characteristic the WWC recommends including for analyses of this type (What Works Clearinghouse, 2017). Many students did not know whether their parents went to college, however, and so the first-generation status variable had 80.8% missing data. Analysis without this first-generation variable resulted in a fraction of missing information (FMI) of 0.076, which is fairly low. However, FMI=0.662 in an additional analysis including first-generation status. This informed the choice to impute 100 datasets for analysis using the Stata 16 *mi impute chained* command (Graham et al., 2007). Convergence and imputed values were deemed reasonable after visual inspection and analyses were pooled according to Rubin's (1987) rules. Given the amount of missing data on first-generation status, sensitivity analyses were conducted both with and without this variable.

B.2 Sample

Both the week level and course level datasets used for the instrumental variables analyses started from the same RCT dataset. Cases were dropped from the final listwise deleted analysis datasets as illustrated in Tables B.2 and B.3 below. These drops were done with the intent of improving the ability to detect the effect of interest. Several additional analyses were performed to gauge the sensitivity of results to these sampling choices and the assumptions underlying them.

Table B.2 Week level dataset dropped cases

	N	Courses	Control	Treatment	Explanation
Initial	2322	26	21	22	RCT dataset 2017 Spring II&III
Drop 1	2202	26	21	22	Students who withdrew
Drop 2	1835	26	21	22	Only analyzing weeks 2-6
Drop 3	1470	17	17	17	Only matched C-T courses
Drop 4	1375	17	17	17	No zeros on outcome
Drop 5	1002	15	15	15	Drop missing on outcome
Drop 6	945	15	15	15	Drop missing on covariates

Table B.3 Course level dataset dropped cases

	N	Courses	Control	Treatment	Explanation
Initial	2322	26	21	22	RCT dataset 2017 Spring II&III
Drop 1	2202	26	21	22	Students who withdrew
Drop 2	367	26	21	22	Only one case per student
Drop 3	294	17	17	17	Only matched C-T courses
Drop 4	267	17	17	17	No zeros on outcome
Drop 5	252	17	17	17	Drop missing on covariates

One sensitivity analysis utilized all available RCT data instead of just matched RCT data. It was assumed that the matched course design whereby each course was taught in treatment and control sections resulted in a stronger design that represented a more accurate estimation of the treatment effect. All such matched course sections were included in the analysis. This resulted in an odd number of course sections because one course had three sections of data available. The additional course data that were available from the RCT included courses that were matched to other courses that occurred prior to the spring 2018 sessions when the modality information key to this analysis began to be available. Thus, while these courses and instructors were matched overall in the RCT, they were not matched for the primary sample analyzed for the present study. Although the stronger matched design was the analysis emphasized when presenting results, an additional analysis utilizing all available data was also performed. It was expected that the effect might be more difficult to detect in the alternate analysis.

For some drops, such as zeros on the outcome and missing data, sensitivity analyses were performed that included these data instead. For both the weekly assignment and quiz grade as well as the course grade, it was assumed that students who fail were substantively different than students who pass. That is, students who choose not to do the work in a given week or who give up doing the work of the course were not directly comparable to students who were actively attempting to learn the material and who earned a non-failing grade. In other words, it was assumed that the outcome distribution would best be characterized by a bimodal distribution that accounted for zeros differently than for other grades. Because of this, the choice was made to focus on what effect existed for students who were doing work demonstrating some understanding of the content. This obviously introduceed a form of selection bias since students who failed were not included. To address this, sensitivity analyses were conducted that included these zeros as well. However, given that the impact of use of multiple modalities was primarily expected to be on learning outcomes and not a mechanism for helping non-traditional students manage other aspects of their lives that end up leading to failure, it was deemed to be reasonable for the purpose of this study to focus on the non-zero part of the outcome distribution. Additionally, to better model the full range of knowledge, which extended beyond what was represented in the data as grades from A to F, a limited regression model was utilized. As is described in more detail in the tobit modeling section below, this censored regression model appropriately took the limited nature of the grade variables into consideration. Although this represented an improvement upon simple regression, it was still deemed worthwhile to compare results with and without zero values included.

Due to the difficulty of conducting appropriate postestimation tests for an instrumental variables analysis with multiply imputed data, such tests were conducted using listwise deletion. Tests were also conducted individually on each multiply imputed dataset and the distribution of resulting test statistics across the imputations was summarized using the minimum, median, and maximum. Once appropriate tests were conducted and the appropriateness of the instrumental variables technique for the data was determined, it was assumed that the multiply imputed results presented a more robust understanding of the treatment effect since this approach utilized all available information, while appropriately taking relevant uncertainty into account.

I will note that some courses did not have an appropriate week level outcome variable because they did not include the type of weekly assignment or quiz grades being incorporated into the outcome as might be expected to show an impact of the use of multiple modalities to learn content. It was assumed that there would be an effect of using multiple modalities on assessments closer to the content, such as quizzes, rather than assessments tailored toward other aspects of the course, such as reflection papers. Thus, if a course did not include an appropriate weekly outcome measure, that course was dropped from analysis. This resulted in one course with two sections being dropped.

Analysis for the week level was also initially done with all six weeks of data, rather than just weeks two through six for the week level analysis. That analysis with all six weeks did not include the grade motivation latent variable that used the first week of data as presented in the final week level results. It was determined that a stronger study design could be achieved by attempting to address possible confounding of the treatment due to some students' strong motivation to achieve a higher grade which would lead them to take additional action to boost that grade, including using multiple modalities even though they demonstrated they already knew the material quite well. Once the means to generate a variable from indicators of this latent construct was determined, analysis was switched to this stronger design. Conclusions drawn from the earlier analysis with all six weeks (not presented) were substantively similar. That is, the effect at the week level was strongly significant and substantively important. The choice was made to go with the stronger analytical strategy, which involved dropping the week one data from the full analysis in order to utilize these week one data to determine a score representing a latent grade motivation construct from the first week of the class.

B.3 Attrition

Since only the final two sessions of the RCT were analyzed here, about 30% of students who agreed to participate in the RCT ended up in the analysis sample. The RCT itself had low attrition over its three years.

For the analysis of weekly assignment and quiz grade with all courses in the RCT (not just matched courses), there was very low differential attrition between the treatment and comparison groups (0.015), and low (1.95%) overall attrition, so this study was considered to have low attrition by WWC standards using the cautious boundary for assumptions (What Works Clearinghouse, 2020b). Similarly for the analysis of course grade, a combination of very low differential attrition (0.036) and low overall attrition (3.24%) resulted in a low attrition rate for the course outcome in this study under the WWC's cautious boundary for assumptions. The optimistic boundary would have been appropriate for studying postsecondary academic achievement outcomes according to the relevant WWC domain guidance, but all attrition analyses in this study met the stricter cautious boundary guidelines (What Works Clearinghouse, 2017).

B.4 Baseline Equivalence

The WWC domain guidance for postsecondary studies of technology and student learning recommend establishing baseline equivalence for two measures since a pre-test measure was not available for the outcomes studied (What Works Clearinghouse, 2017). The WWC recommends that the first be a continuous measure of prior educational achievement. This study used the number of undergraduate credits the student transferred to the institution when starting their degree as the first baseline measure. The second gauged socioeconomic status, here via Pell grant eligibility. Tables B.4 and B.5 show the comparison group and treatment group count, mean outcome, and standard deviation, as well as the simple baseline difference and standardized baseline difference as per WWC guidelines (What Works Clearinghouse, 2020b).

Table B.4 Baseline Equivalence for Pell Grant Status

Analysis	N_C	$\overline{Y}_C$	N_T	$\overline{Y}_T$	$\hat{\beta}$	$\lvert d_{Cox} \rvert$
Week-no 0[a]	752	0.654	193	0.701	0.046	0.212
Week-w/0[b]	817	0.654	204	0.704	0.050	0.231
Course-no 0[a]	203	0.626	49	0.762	0.136	0.648
Course-w/0[b]	223	0.637	52	0.766	0.129	0.620

[a] Sample does not include zeros (i.e., failing grades) on dependent variable for subsequent analysis.
[b] Sample includes zeros (i.e., failing grades) on dependent variable for subsequent analysis.

Table B.5 Baseline Equivalence for Academic Credits Transferred Upon College Entry

Analysis	N_C	$\overline{Y}_C$	SD_C	N_T	$\overline{Y}_T$	SD_C	*Simple Baseline Diff.*	*|Std. Baseline Diff.|*
Week-no 0[a]	752	26.597	24.046	193	27.330	26.041	1.102	0.030
Week-w/0[b]	817	26.149	23.820	204	27.548	25.966	0.757	0.058
Course-no 0[a]	203	28.113	24.424	49	24.331	25.958	-2.643	0.152
Course-w/0[b]	223	27.825	24.251	52	23.858	25.685	-3.210	0.161

[a] Sample does not include zeros (i.e., failing grades) on dependent variable for subsequent analysis.
[b] Sample includes zeros (i.e., failing grades) on dependent variable for subsequent analysis.

Baseline equivalence for week level Pell status was 0.21, calculated as the difference between treatment and control groups at baseline; this required statistical adjustment per WWC standards since it was within the 0.05-0.25 range. Baseline equivalence for credits transferred was 0.03 for the week level, showing better balance between groups. For the course level, baseline equivalence for credits transferred was 0.15, also requiring adjustment. However, the course baseline equivalence for Pell status was beyond the desired maximum, even with adjustment. Thus, baseline equivalence can be met for these variables with adjustment for all but Pell status for the course level analysis since the rest have a standardized baseline difference absolute value of less than 0.25.

Additionally, not all of the four race/ethnicity groups met the WWC's baseline equivalence recommendations in both the week and course data. Specifically, only the Other race/ethnicity category in the week data was under the desired 0.25, as were the White and Black categories in the course data. So, the samples did not achieve recommended equivalence on this additional race/ethnicity metric overall. Since the sample was all female, checking baseline equivalence on gender was not relevant. Although the WWC also recommended checking the percentage of students who speak English as a second language (ESL), this information was not available. However, while these ancillary results might suggest some caution in interpretation, baseline differences in race/ethnicity or ESL are not expected to be a major threat to validity for a study of this type. Overall, baseline equivalence was deemed reasonable for the week level analysis after adjustment, with some caution to be noted for the course level analysis.

Although baseline equivalence was not achieved on one element of socioeconomic status for the course outcome, it is notable when interpreting the course level results that other aspects of this research offer strong study design. The week level results and robust evidence presented in chapter two of a positive effect on formative learning activities bolster plausibility for an effect on course grades. However, given the small sample size, other elements of the study are key to detecting this effect. The quality of the data collection, low attrition, matched course design, matched instructor design, use of quality metrics in course development providing consistency across courses, principled method of handling missing data to retain information, strength of the RCT instrument, and appropriate modeling of the dependent variable all played important roles. While lack of socioeconomic baseline equivalence may pose a threat to external validity, Pell status and first-generation status were included as covariates to partially address this issue to the extent possible. Additionally, motivation to achieve a high grade may be an important unobserved confounder in the course analysis that may have led to students already doing well in the course pursuing multiple modalities to boost already high grades. However, while the relationships seen at the week level suggest such a factor may indeed be strong enough to influence results, it seems unlikely to erase them since conditioning on a measure of such grade motivation for the week analysis did not explain away the effect of interest. It would be worth addressing such issues in future research, including identifying and ruling out other possible confounders. Thus, while future research would be warranted to confirm the effect at the course level, the results here are highly suggestive that an effect may exist at the course level as well as the week level.

B.5 Model

The model analyzed for both week and course level outcomes was an instrumental variables tobit model run in Stata 16 using the *ivtobit* command (StataCorp, 2019). A tobit model was appropriate since both dependent variables were censored, having both lower and upper limits. An instrumental variables analysis allowed for the endogenous treatment regressor to be predicted by the exogenous RCT participation instrument. Additional covariates were included to increase precision. The model was

$$y_{1i}^{*} = y_{2i}\beta + \boldsymbol{x_{1i}}\boldsymbol{\gamma} + \upsilon_i \qquad (1)$$

$$y_{2i} = \boldsymbol{x_{1i}}\boldsymbol{\Pi_1} + x_{2i}\Pi_2 + \nu_i \qquad (2)$$

assuming $(\upsilon_i, \nu_i) \sim N(0, \Sigma)$, that is, the two error terms υ_i and ν_i were multivariate normal distributed with mean 0. The assumed correlation between the two error terms resulted in the endogeneity of y_{2i}. It was also assumed that the pair of errors (υ_i, ν_i) were independent and identically distributed across all i; where $i = 1,\ldots,N$ cases. The structural equation (1) includeed parameters β and $\boldsymbol{\gamma}$. y_{2i} was the endogenous treatment, so equation (2) gave the first stage equation necessary for calculating estimates via instrumental variables (i.e., Wald estimate when there are no other covariates beyond the single instrument). Π_1 and Π_2 were the reduced-form parameters. $\boldsymbol{x}_{1i}$ were the exogenous variables (i.e., model covariates) and x_{2i} was the primary instrument (i.e., RCT group). y_{1i}^* was a latent variable for which observations are censored at 0 and 1 for week grade and at 0 and 4 for course grade. Specifically,

$$\text{Week: } y_{1i} = \begin{cases} 0 & y_{1i}^* < 0 \\ y_{1i}^* & 0 \le y_{1i}^* \le 1 \\ 1 & y_{1i}^* > 1 \end{cases} \qquad \text{Course: } y_{1i} = \begin{cases} 0 & y_{1i}^* < 0 \\ y_{1i}^* & 0 \le y_{1i}^* \le 4 \\ 4 & y_{1i}^* > 4 \end{cases}$$

B.6 Descriptive Results

Means and standard errors are shown in Tables B.6 (week-level) and B.7 (course-level). No variables were so highly correlated as to cause concern (see Tables B.8 and B.9). Since this analysis used RCT data, treatment and control course split was checked for students, courses, instructors, and field of study.

As anticipated, these women were typically older (average age of 36 years), low-income students (66.0% were federal Pell grant recipients), who were first-generation students (76.4%) coming back to complete their degrees (on average, they transferred in almost 27 credits). While slightly over half were White (56.0%), there were significant numbers of Black (19.6%) and Hispanic (17.7%) students as well. Life factors external to college were an important influence for many of these students (average scale score of 77.2%), and almost all possessed basic technological competency (average scale score of 94.6%). They exhibited a range of learning styles (with scale scores ranging from 73.1% to 80.4% on verbal, visual, aural, and physical approaches. Most were doing well academically (average GPA of 3.48) and had few prior withdrawals or failures in their courses (2.4%).

The students were spread out across the sample in terms of time and courses. The students were roughly split between the two sessions (92 in one and 95 in the other). The number of students per course analyzed ranged from four (in upper-level psychology and business courses) to 36 (in an introductory English course), with a mean of 16. For the 267 student cases analyzed in the course level analysis under multiple imputation, mean course grade was 3.35 on a 4.0 scale and 19.5% received the treatment.

In the week dataset, there were 37 instances of 15 courses, 18 treatment (96 students) and 19 comparison courses (81 students). 15 instructors taught matched treatment and control sections, while the matches for 5 others occurred outside the sample studied. The number of students in treatment and comparison courses by field of study was fairly balanced, including 35/31 in humanities, 31/30 in professional studies, and 47/33 in social sciences. The course dataset was likewise spread across RCT treatment and comparison groups on such factors.

Table B.6 Estimated Means and Standard Errors for Week Level Analyses

	(1) Matched, MI, no 0 grades		(2) Matched, MI, includes 0 grades		(3) Matched, LD, no 0 grades		(4) Matched, LD, includes 0 grades	
Variables	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Weekly grade[a]	0.881	(0.004)	0.805	(0.008)	0.885	(0.004)	0.819	(0.008)
≥7 multi-modal uses[b]	0.207	(0.013)	0.200	(0.012)	0.204	(0.013)	0.200	(0.013)
Race/ethnicity								
White	0.560	(0.016)	0.544	(0.015)	0.559	(0.016)	0.541	(0.016)
Black	0.196	(0.013)	0.195	(0.012)	0.197	(0.013)	0.198	(0.012)
Hispanic	0.177	(0.012)	0.188	(0.012)	0.175	(0.012)	0.188	(0.012)
Other	0.067	(0.008)	0.073	(0.008)	0.070	(0.008)	0.073	(0.008)
Age	36.177	(0.289)	35.813	(0.276)	36.219	(0.298)	35.909	(0.284)
First-generation	0.764	(0.027)	0.782	(0.023)				
Pell	0.660	(0.015)	0.653	(0.014)	0.663	(0.015)	0.663	(0.015)
Transfer credits	26.986	(0.763)	26.668	(0.728)	26.822	(0.796)	26.301	(0.759)
Base Pell	0.660	(0.004)	0.653	(0.004)	0.663	(0.005)	0.663	(0.004)
Base credits	26.948	(0.252)	26.555	(0.221)	26.822	(0.259)	26.301	(0.232)
WF count	0.024	(0.005)	0.029	(0.006)	0.025	(0.006)	0.031	(0.006)
GPA	3.475	(0.018)	3.319	(0.025)	3.496	(0.018)	3.376	(0.023)
Grade motivation	0.197	(0.024)	0.097	(0.027)	0.214	(0.025)	0.144	(0.026)
Verbal LS	0.731	(0.003)	0.728	(0.003)	0.731	(0.003)	0.730	(0.003)
Visual LS	0.742	(0.003)	0.740	(0.003)	0.741	(0.003)	0.740	(0.003)
Aural LS	0.753	(0.005)	0.749	(0.005)	0.753	(0.006)	0.749	(0.005)
Physical LS	0.804	(0.002)	0.802	(0.002)	0.804	(0.002)	0.803	(0.002)
Life score	0.772	(0.002)	0.770	(0.002)	0.772	(0.002)	0.770	(0.002)
Tech score	0.946	(0.003)	0.945	(0.002)	0.948	(0.003)	0.947	(0.002)
Session	0.502	(0.016)	0.506	(0.015)	0.505	(0.016)	0.509	(0.016)
RCT Cohort 1	0.038	(0.006)	0.035	(0.006)	0.035	(0.006)	0.032	(0.006)
RCT Cohort 2	0.051	(0.007)	0.048	(0.006)	0.046	(0.007)	0.044	(0.006)
RCT Cohort 3	0.106	(0.010)	0.109	(0.009)	0.112	(0.010)	0.118	(0.010)
RCT Cohort 4	0.142	(0.011)	0.138	(0.010)	0.133	(0.011)	0.129	(0.011)
RCT Cohort 5	0.361	(0.015)	0.361	(0.015)	0.372	(0.016)	0.373	(0.015)
RCT Cohort 6	0.302	(0.015)	0.309	(0.014)	0.302	(0.015)	0.304	(0.014)
RCT treatment	0.589	(0.016)	0.595	(0.015)	0.589	(0.016)	0.598	(0.015)
Observations	1,002		1,097		945		1,021	

Note: (1) and (2) included matched treatment-comparison courses from the RCT with multiply imputed data. (3) and (4) included matched courses with listwise deleted data.

[a] Weekly mean grade on assignments and quizzes. [b] Use of >1 modality ≥7 times/week.

Table B.7 Estimated Means and Standard Errors for Course Level Analyses

	(1) Matched, MI, no 0 grades		(2) Matched, MI, includes 0 grades		(3) Matched, LD, no 0 grades		(4) Matched, LD, includes 0 grades	
Variables	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Course grade	3.345	(0.047)	3.037	(0.071)	3.370	(0.048)	3.088	(0.071)
≥38 multi-modal uses[a]	0.195	(0.024)	0.187	(0.023)	0.194	(0.025)	0.189	(0.024)
Race/ethnicity								
White	0.565	(0.031)	0.548	(0.029)	0.567	(0.031)	0.549	(0.030)
Black	0.185	(0.024)	0.182	(0.023)	0.187	(0.025)	0.185	(0.023)
Hispanic	0.184	(0.024)	0.200	(0.024)	0.179	(0.024)	0.196	(0.024)
Other	0.065	(0.015)	0.070	(0.015)	0.067	(0.016)	0.069	(0.015)
Age	36.000	(0.543)	35.680	(0.520)	36.052	(0.558)	35.782	(0.536)
First-generation	0.705	(0.050)	0.733	(0.049)				
Pell	0.644	(0.029)	0.646	(0.028)	0.647	(0.030)	0.655	(0.029)
Transfer credits	27.820	(1.488)	27.556	(1.417)	27.599	(1.556)	27.218	(1.478)
Base Pell	0.644	(0.009)	0.646	(0.007)	0.647	(0.010)	0.655	(0.008)
Base credits	27.820	(0.379)	27.469	(0.323)	27.599	(0.384)	27.218	(0.337)
WF count	0.019	(0.010)	0.027	(0.011)	0.020	(0.010)	0.029	(0.011)
GPA	3.507	(0.030)	3.328	(0.046)	3.532	(0.030)	3.377	(0.044)
Verbal LS	0.731	(0.007)	0.729	(0.006)	0.731	(0.007)	0.730	(0.006)
Visual LS	0.740	(0.006)	0.739	(0.006)	0.740	(0.006)	0.739	(0.006)
Aural LS	0.750	(0.011)	0.746	(0.010)	0.750	(0.011)	0.746	(0.010)
Physical LS	0.801	(0.004)	0.800	(0.004)	0.802	(0.004)	0.801	(0.004)
Life score	0.768	(0.004)	0.768	(0.004)	0.769	(0.005)	0.768	(0.004)
Tech score	0.943	(0.005)	0.943	(0.005)	0.944	(0.005)	0.944	(0.005)
Session	0.487	(0.031)	0.493	(0.029)	0.488	(0.032)	0.495	(0.030)
RCT Cohort 1	0.034	(0.011)	0.031	(0.010)	0.032	(0.011)	0.029	(0.010)
RCT Cohort 2	0.056	(0.014)	0.051	(0.013)	0.052	(0.014)	0.047	(0.013)
RCT Cohort 3	0.105	(0.019)	0.109	(0.018)	0.111	(0.020)	0.116	(0.019)
RCT Cohort 4	0.154	(0.022)	0.153	(0.021)	0.143	(0.022)	0.142	(0.021)
RCT Cohort 5	0.352	(0.029)	0.350	(0.028)	0.365	(0.030)	0.364	(0.029)
RCT Cohort 6	0.300	(0.028)	0.306	(0.027)	0.298	(0.029)	0.302	(0.028)
RCT treatment	0.577	(0.030)	0.585	(0.029)	0.575	(0.031)	0.585	(0.030)
Observations	267		294		252		275	

Note: (1) and (2) included matched treatment-comparison courses from the RCT with multiply imputed data. (3) and (4) included matched courses with listwise deleted data.

[a] Use of >1 modality ≥38 times/week.

Table B.8 Correlation Matrix for Analysis of Weekly Assignment and Quiz Grade

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) Assign/Quiz	1.0000												
(2) Modality 7/wk	0.0907	1.0000											
(3) White	0.1341	0.0643	1.0000										
(4) Black	-0.0050	-0.0791	-0.5570	1.0000									
(5) Hispanic	-0.1853	-0.0187	-0.5175	-0.2277	1.0000								
(6) Other	0.0225	0.0260	-0.3083	-0.1356	-0.1260	1.0000							
(7) Age	0.0791	-0.1033	0.1549	-0.0444	-0.1076	-0.0723	1.0000						
(8) Pell	-0.1137	0.0386	-0.2811	0.1779	0.1506	0.0458	-0.2069	1.0000					
(9) Xfer Entry	0.0138	0.0182	0.0593	-0.0989	0.0883	-0.0928	0.0735	0.0177	1.0000				
(10) Base Pell	0.0198	0.0230	-0.1761	0.0937	0.0847	0.0708	-0.0660	0.3084	0.0673	1.0000			
(11) Base Xfer	0.0527	0.0263	-0.0425	0.0446	0.0406	-0.0473	-0.1165	0.0638	0.3251	0.2069	1.0000		
(12) Prior WF	-0.0342	-0.0283	0.0313	0.0494	-0.0662	-0.0395	0.0988	-0.1260	0.0187	0.0001	-0.0885	1.0000	
(13) GPA	0.4762	0.0844	0.3323	-0.0854	-0.3059	-0.0584	0.2218	-0.1560	-0.0299	-0.0565	-0.0808	-0.2001	1.0000
(14) AcadMotiv	0.4017	0.0128	0.2838	-0.1260	-0.2811	0.0623	0.2008	-0.2318	0.0058	-0.0704	-0.0029	-0.1481	0.5068
(15) Verbal	0.0344	-0.0121	0.0812	-0.1545	-0.0169	0.1081	0.2533	-0.1114	0.0064	0.0324	-0.0769	0.0080	0.1498
(16) Visual	0.0221	-0.0239	0.1195	-0.0504	-0.0931	-0.0156	0.2065	-0.0055	-0.0636	0.0324	0.0277	-0.0892	0.1460
(17) Aural	0.0435	-0.1582	-0.0008	-0.0285	-0.0329	0.0950	0.2066	0.0643	0.0671	0.0612	-0.1186	0.0685	0.0182
(18) Physical	0.0450	-0.0365	0.0281	-0.0571	-0.0245	0.0708	0.2652	-0.0974	-0.1380	0.0662	-0.1412	-0.0673	0.1641
(19) LifeFactor	0.0141	-0.0441	-0.1570	0.0314	0.0728	0.1486	0.0583	-0.0505	-0.0835	0.0275	-0.0229	0.0035	0.0503
(20) TechComp	0.0489	-0.0740	0.0111	0.0165	-0.0785	0.0696	0.0938	0.0438	0.0336	0.1523	0.0468	-0.0579	0.1320
(21) Session	0.0223	0.2551	0.0490	-0.0207	-0.0127	-0.0441	0.0339	-0.0918	0.0124	-0.1170	-0.1358	0.1186	0.0115
(22) Cohort 1	-0.0511	0.0323	0.0530	0.0073	-0.0419	-0.0521	0.0320	0.0379	-0.1255	0.1228	-0.3860	0.1033	0.0018
(23) Cohort 2	-0.0403	0.0279	0.0713	0.0069	-0.1004	-0.0001	0.0752	-0.0057	-0.0508	-0.0185	-0.1561	-0.0314	-0.0402
(24) Cohort 3	-0.0273	-0.0387	-0.0826	-0.1085	0.1015	0.1789	0.0457	0.0970	-0.0703	0.3145	-0.2161	0.0439	-0.0606
(25) Cohort 4	-0.0114	0.0407	0.0226	-0.0767	0.0902	-0.0586	-0.0451	-0.0567	0.0256	-0.1837	0.0789	-0.0388	0.0301
(26) Cohort 5	-0.0211	0.0549	0.0147	0.0095	-0.0315	0.0036	0.0213	-0.1461	-0.1436	-0.4738	-0.4416	0.0380	0.0472
(27) Cohort 6	0.0882	-0.0870	-0.0290	0.1154	-0.0411	-0.0625	-0.0674	0.1166	0.2538	0.3782	0.7806	-0.0685	-0.0128
(28) RCT	0.0797	0.4228	-0.0659	-0.0359	0.0779	0.0683	-0.1112	0.0475	0.0916	0.1540	0.2818	0.0226	-0.0416

	(14)	(15)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)
(14) AcadMotiv	1.0000												
(15) Verbal	0.0977	1.0000											
(16) Visual	0.0617	0.1851	1.0000										
(17) Aural	0.1520	0.2276	0.0490	1.0000									
(18) Physical	0.1520	0.6256	0.5343	0.3714	1.0000								
(19) LifeFactor	0.0617	0.3047	0.0204	0.3689	0.4274	1.0000							
(20) TechComp	0.1154	-0.0137	0.0234	0.0349	0.0791	0.0171	1.0000						
(21) Session	-0.0319	0.0375	-0.0388	0.0042	-0.0097	-0.0401	0.0095	1.0000					
(22) Cohort 1	0.0198	-0.0138	0.1048	0.0318	0.1170	-0.0559	0.0019	0.0040	1.0000				
(23) Cohort 2	0.1483	-0.0151	0.0703	0.0622	-0.0158	0.0005	0.1018	-0.0884	-0.0415	1.0000			
(24) Cohort 3	-0.1452	0.0158	-0.0465	0.1449	0.1130	0.1444	0.0887	-0.0302	-0.0676	-0.0776	1.0000		
(25) Cohort 4	-0.1061	-0.0572	-0.0108	-0.2224	-0.0815	-0.0869	0.0080	0.0772	-0.0746	-0.0856	-0.1394	1.0000	
(26) Cohort 5	0.0247	-0.0024	-0.1446	0.0129	-0.0196	0.0011	-0.1716	0.1897	-0.1466	-0.1682	-0.2739	-0.3022	1.0000
(27) Cohort 6	0.0771	0.0464	0.1184	0.0105	-0.0364	-0.0140	0.0670	-0.1976	-0.1250	-0.1435	-0.2336	-0.2577	-0.5063
(28) RCT	-0.0075	-0.1161	-0.0588	-0.1753	-0.1240	0.0234	-0.0289	-0.0308	-0.0639	0.0584	-0.0237	-0.0460	0.0602

Table B.9 Correlation Matrix for Analysis of Course Grade

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) Course Grade	1.0000												
(2)Modal. 38/crs	0.0416	1.0000											
(3) White	0.2839	-0.0365	1.0000										
(4) Black	-0.1393	-0.0036	-0.5484	1.0000									
(5) Hispanic	-0.1820	0.0065	-0.5340	-0.2233	1.0000								
(6) Other	-0.0664	0.0677	-0.3081	-0.1288	-0.1254	1.0000							
(7) Age	0.2382	-0.1616	0.1935	-0.0742	-0.1351	-0.0606	1.0000						
(8) Pell	-0.1119	0.0903	-0.2429	0.2046	0.0627	0.0663	-0.1741	1.0000					
(9) Xfer Entry	0.1204	-0.0424	0.0791	-0.0922	0.0416	-0.0765	0.0816	0.0341	1.0000				
(10) Base Pell	-0.0433	0.0072	-0.1592	0.1210	0.0666	0.0249	-0.0574	0.3172	0.0396	1.0000			
(11) Base Xfer	0.0677	0.0199	-0.0354	0.0414	0.0437	-0.0611	-0.1021	0.0508	0.2470	0.1603	1.0000		
(12) Prior WF	-0.1813	-0.0589	0.0079	0.0657	-0.0559	-0.0322	0.1322	-0.1121	-0.0233	-0.0153	-0.1294	1.0000	
(13) GPA	0.7513	-0.0614	0.3611	-0.1371	-0.2758	-0.0793	0.2434	-0.1320	-0.0266	-0.1233	-0.0207	-0.1309	1.0000
(14) Verbal	0.0736	-0.0598	0.0414	-0.0948	-0.0200	0.0960	0.2337	-0.0899	-0.0091	0.1137	-0.0379	-0.0210	0.1272
(15) Visual	0.0846	-0.1394	0.1000	-0.0425	-0.0516	-0.0527	0.1837	-0.0040	-0.1002	0.0699	0.0371	-0.0926	0.1103
(16) Aural	0.0069	-0.1433	0.0355	-0.0352	-0.0732	0.0962	0.2089	0.0696	0.0656	0.0999	-0.1955	0.0828	0.0595
(17) Physical	0.1141	-0.1242	0.0126	-0.0292	-0.0245	0.0579	0.2672	-0.0785	-0.1740	0.1273	-0.1130	-0.0682	0.1892
(18) LifeFactor	0.0240	-0.0373	-0.1618	0.0403	0.0682	0.1527	0.0535	-0.0708	-0.1205	0.0635	-0.0548	0.0153	0.0830
(19) TechComp	0.0892	-0.1148	0.0193	0.0541	-0.1247	0.0682	0.0873	0.0501	-0.0236	0.1259	0.0285	-0.0366	0.1837
(20) Session	0.0779	0.1621	0.0353	-0.0192	-0.0200	-0.0094	0.0473	-0.0923	0.0523	-0.1751	-0.0678	0.1228	0.0046
(21) Cohort 1	0.0041	0.0254	0.0210	0.0295	-0.0253	-0.0487	0.0450	0.0391	-0.1128	0.1232	-0.4565	0.1151	-0.0163
(22) Cohort 2	-0.0357	-0.0693	0.0588	-0.0196	-0.1087	0.0803	0.0615	-0.0153	-0.0421	-0.0484	-0.1703	-0.0280	-0.0465
(23) Cohort 3	-0.0891	0.0177	-0.0736	-0.1045	0.0989	0.1566	0.0565	0.0763	-0.0721	0.2406	-0.2920	0.0339	-0.1005
(24) Cohort 4	0.0741	0.0860	-0.0098	-0.0208	0.0761	-0.0646	-0.0370	-0.1017	0.0839	-0.3206	0.3398	-0.0489	-0.0037
(25) Cohort 5	-0.0527	0.0648	0.0465	-0.0245	-0.0307	-0.0068	-0.0231	-0.1295	-0.1131	-0.4082	-0.4578	0.0585	0.0800
(26) Cohort 6	0.0757	-0.1224	-0.0273	0.1117	-0.0316	-0.0713	-0.0332	0.1541	0.1680	0.4860	0.6801	-0.0780	0.0164
(27) RCT	0.0012	0.4220	-0.0532	-0.0421	0.0861	0.0390	-0.1339	0.0539	0.0525	0.1700	0.2125	0.0545	-0.0742

	(14)	(15)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)
(14) Verbal	1.0000												
(15) Visual	0.1686	1.0000											
(16) Aural	0.2061	0.0376	1.0000										
(17) Physical	0.6134	0.5455	0.3525	1.0000									
(18) LifeFactor	0.3109	0.0364	0.3326	0.4168	1.0000								
(19) TechComp	0.0097	-0.0015	0.0499	0.0873	0.0426	1.0000							
(20) Session	0.0122	-0.0540	0.0091	-0.0272	-0.0482	0.0038	1.0000						
(21) Cohort 1	0.0060	0.0876	0.0446	0.1133	-0.0479	0.0137	0.0043	1.0000					
(22) Cohort 2	-0.0003	0.0335	0.0619	-0.0002	0.0212	0.0973	-0.0483	-0.0422	1.0000				
(23) Cohort 3	0.0417	-0.0131	0.1245	0.1200	0.1541	0.0908	-0.0674	-0.0640	-0.0825	1.0000			
(24) Cohort 4	-0.0607	-0.0030	-0.2683	-0.0893	-0.1192	-0.0187	0.1458	-0.0739	-0.0952	-0.1443	1.0000		
(25) Cohort 5	-0.0534	-0.1573	0.0704	-0.0569	0.0130	-0.1646	0.1500	-0.1373	-0.1769	-0.2681	-0.3096	1.0000	
(26) Cohort 6	0.0719	0.1271	-0.0014	0.0025	-0.0203	0.0730	-0.2015	-0.1179	-0.1518	-0.2301	-0.2657	-0.4936	1.0000
(27) RCT	-0.1180	-0.0576	-0.1781	-0.1335	0.0159	-0.0484	0.0036	-0.0734	0.0189	-0.0539	0.0295	0.0844	-0.0554

B.7 Instrumental Variables Tobit Regression Tables

Table B.10 Instrumental Variables Tobit Analysis second stage, Weekly Assignment and Course Grade Outcomes, MI and LD

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Weekly Grade				Course Grade			
Variables	Matched, MI, no 0s	Matched, MI , includes 0s	Matched, LD, no 0s	Matched, LD, includes 0s	Matched, MI, no 0s	Matched, MI , includes 0s	Matched, LD, no 0s	Matched, LD, includes 0s
≥7 or 38 modalities	0.139***	0.102+	0.136***	0.111*	0.754*	0.893+	0.618*	0.592
	(0.032)	(0.054)	(0.032)	(0.054)	(0.359)	(0.493)	(0.309)	(0.404)
Race/ethnicity								
Black	0.031*	0.058*	0.033*	0.056*	-0.109	0.077	-0.100	0.065
	(0.014)	(0.023)	(0.013)	(0.024)	(0.134)	(0.154)	(0.125)	(0.150)
Hispanic	0.029*	0.084***	0.022	0.068***	0.124	0.279+	0.081	0.235
	(0.015)	(0.025)	(0.015)	(0.025)	(0.138)	(0.157)	(0.132)	(0.155)
Other	0.034+	0.070*	0.025	0.074*	0.054	0.130	0.041	0.186
	(0.020)	(0.033)	(0.020)	(0.034)	(0.206)	(0.232)	(0.187)	(0.225)
Age	0.000	0.001	0.000	0.001	0.017***	0.014*	0.017***	0.014*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.006)	(0.007)	(0.006)	(0.007)
Pell	-0.021+	-0.004	-0.018	-0.009	-0.182	-0.337***	-0.130	-0.281*
	(0.011)	(0.019)	(0.011)	(0.020)	(0.111)	(0.130)	(0.107)	(0.130)
Xfer. credits	0.000	0.001*	0.000	0.001*	0.005*	0.006*	0.005*	0.006*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)	(0.002)	(0.002)
Base Pell	-0.019	0.010	-0.006	0.018	0.382	0.634	0.376	0.831
	(0.040)	(0.073)	(0.042)	(0.078)	(0.411)	(0.541)	(0.411)	(0.570)
Base credits	-0.003+	-0.001	-0.002	-0.003	-0.049	-0.088+	-0.029	-0.046
	(0.002)	(0.003)	(0.001)	(0.003)	(0.034)	(0.051)	(0.023)	(0.035)
WF count	0.089***	0.075+	0.083***	0.058	-0.526+	-0.582+	-0.457+	-0.574+
	(0.029)	(0.044)	(0.028)	(0.044)	(0.300)	(0.305)	(0.266)	(0.296)
GPA	0.103***	0.303***	0.100***	0.289***	1.483***	1.993***	1.490***	2.046***
	(0.011)	(0.015)	(0.011)	(0.016)	(0.109)	(0.110)	(0.108)	(0.116)
Grade motiv.	0.040***	0.039***	0.045***	0.039***				
	(0.007)	(0.012)	(0.008)	(0.013)				
Verbal LS	-0.110+	-0.339***	-0.108+	-0.383***	-0.754	-1.009	-0.823	-1.059
	(0.064)	(0.109)	(0.063)	(0.110)	(0.607)	(0.722)	(0.571)	(0.699)
Visual LS	-0.090	-0.314***	-0.102	-0.325***	-0.219	-0.552	-0.357	-0.674
	(0.068)	(0.118)	(0.066)	(0.115)	(0.634)	(0.742)	(0.583)	(0.702)
Aural LS	0.095***	0.113+	0.093***	0.119+	0.235	0.445	0.194	0.415
	(0.036)	(0.063)	(0.036)	(0.062)	(0.349)	(0.408)	(0.323)	(0.389)
Physical LS	0.086	0.418+	0.082	0.433+	0.618	0.543	1.012	0.785
	(0.145)	(0.253)	(0.142)	(0.249)	(1.387)	(1.657)	(1.296)	(1.592)
Life score	-0.137+	-0.155	-0.113	-0.116	-0.058	-0.277	0.008	-0.185
	(0.077)	(0.133)	(0.075)	(0.133)	(0.795)	(0.941)	(0.749)	(0.919)

Variables	(1) Matched, MI, no 0s	(2) Matched, MI , includes 0s	(3) Matched, LD, no 0s	(4) Matched, LD, includes 0s	(5) Matched, MI, no 0s	(6) Matched, MI , includes 0s	(7) Matched, LD, no 0s	(8) Matched, LD, includes 0s
	Weekly Grade				Course Grade			
Tech score	0.020	-0.226*	-0.019	-0.245*	0.131	-0.368	-0.065	-0.679
	(0.066)	(0.112)	(0.065)	(0.111)	(0.653)	(0.775)	(0.598)	(0.719)
Session	-0.015	-0.000	-0.010	0.004	0.067	0.028	0.109	0.101
	(0.012)	(0.020)	(0.012)	(0.020)	(0.107)	(0.128)	(0.100)	(0.121)
Cohort 1	-0.189***	-0.090	-0.151***	-0.130	-1.117	-1.863	-0.668	-0.883
	(0.052)	(0.089)	(0.046)	(0.085)	(0.821)	(1.187)	(0.581)	(0.838)
Cohort 2	-0.145***	-0.083	-0.121***	-0.099	-0.720+	-0.942+	-0.357	-0.337
	(0.034)	(0.058)	(0.035)	(0.063)	(0.398)	(0.545)	(0.333)	(0.450)
Cohort 3	-0.066*	-0.074	-0.050+	-0.102*	-0.815+	-1.300*	-0.514	-0.740+
	(0.030)	(0.054)	(0.027)	(0.051)	(0.469)	(0.638)	(0.326)	(0.441)
Cohort 4	-0.058***	-0.050	-0.048*	-0.062	-0.141	-0.485	-0.064	-0.152
	(0.022)	(0.039)	(0.021)	(0.039)	(0.198)	(0.320)	(0.188)	(0.256)
Cohort 5	-0.100***	-0.093*	-0.083***	-0.115*	-0.886*	-1.381*	-0.587*	-0.813*
	(0.029)	(0.046)	(0.025)	(0.046)	(0.414)	(0.566)	(0.282)	(0.389)
Constant	0.758***	0.174	0.753***	0.285	-0.880	-0.416	-1.634	-2.079
	(0.113)	(0.199)	(0.111)	(0.203)	(1.353)	(1.800)	(1.146)	(1.495)
1st Stage RCT treatment	0.393***	0.378***	0.393***	0.378***	0.371***	0.338***	0.385***	0.352***
	(0.030)	(0.027)	(0.030)	(0.027)	(0.061)	(0.062)	(0.059)	(0.055)
Observations	1,002	1,097	945	1,021	267	294	252	275

Note: Standard errors in parentheses. *** p<0.01, ** p<0.01, * p<0.05, + p<0.10

Full instrumental variables tobit regression results for the second stage are shown in Table B.10. These results correspond to the effect sizes presented in Table 3.1.

Additional sensitivity analyses to model type were conducted, comparing ordinary least squares (OLS) to a regression-based instrumental variables analysis and a tobit-based instrumental variables analysis (see Table B.11). These IV analyses accounted for the endogenous nature of the instrument in the regression, using that to partition the variance, allowing estimation of a treatment effect. To check the influence on precision of including a host of person-centered covariates, a minimal model was compared (see Table B.11). Additionally, the main results presented assume homoscedastic errors, so a model assuming heteroskedastic errors accounted for using a robust approach to calculating standard errors investigated this assumption (using *vce(robust)*). As seen in Table B.11, the *ivtobit* model, which more accurately reflected the underlying nature of the dependent variable, identified the largest estimate of the three modeling methods. Interestingly, the minimal model has a larger effect, so perhaps the more parsimonious model would be beneficial. However, the robust approach had almost identical results to the main results, leading to the conclusion that the results without relaxing the homoscedastic assumption sufficed.

Table B.11 Instrumental Variables Alternate Models, Weekly Grade Outcome, Matched RCT Courses, No Zeros on Grades, Listwise Deletion for Missing Data

Variables	(1) OLS	(2) IV Regress	(3) IV Tobit (main model)	(4) Minimal model	(5) Robust errors
≥7 modalities	0.022*	0.106***	0.136***	0.161***	0.136***
	(0.010)	(0.027)	(0.032)	(0.035)	(0.032)
Race–Black	0.022*	0.029***	0.033*		0.033*
	(0.011)	(0.011)	(0.013)		(0.014)
Race–Hispanic	0.007	0.011	0.022		0.022
	(0.011)	(0.012)	(0.015)		(0.014)
Race-Other	0.024	0.020	0.025		0.025
	(0.016)	(0.016)	(0.020)		(0.018)
Age	-0.001	-0.000	0.000		0.000
	(0.000)	(0.000)	(0.001)		(0.001)
Pell	-0.011	-0.017+	-0.018	-0.046***	-0.018
	(0.009)	(0.009)	(0.011)	(0.012)	(0.011)
Transfer credits	-0.000	-0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Base Pell	0.021	-0.005	-0.006	0.000	-0.006
	(0.033)	(0.035)	(0.042)	(0.048)	(0.041)
Base credits	-0.000	-0.002	-0.002	-0.006***	-0.002
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
WF count	0.062***	0.064***	0.083***		0.083*
	(0.022)	(0.023)	(0.028)		(0.039)

Variables	(1) OLS	(2) IV Regress	(3) IV Tobit (main model)	(4) Minimal model	(5) Robust errors
GPA	0.094***	0.085***	0.100***		0.100***
	(0.008)	(0.009)	(0.011)		(0.014)
Grade motivation	0.040***	0.041***	0.045***		0.045***
	(0.006)	(0.006)	(0.008)		(0.009)
Verbal LS	-0.053	-0.056	-0.108+		-0.108+
	(0.050)	(0.052)	(0.063)		(0.063)
Visual LS	-0.037	-0.023	-0.102		-0.102
	(0.052)	(0.054)	(0.066)		(0.066)
Aural LS	0.029	0.063*	0.093***		0.093***
	(0.026)	(0.029)	(0.036)		(0.036)
Physical LS	0.017	-0.022	0.082		0.082
	(0.111)	(0.116)	(0.142)		(0.147)
Life score	-0.065	-0.068	-0.113		-0.113+
	(0.059)	(0.062)	(0.075)		(0.068)
Tech score	-0.054	-0.007	-0.019		-0.019
	(0.050)	(0.053)	(0.065)		(0.073)
Session	0.004	-0.014	-0.010		-0.010
	(0.008)	(0.010)	(0.012)		(0.012)
Cohort 1	-0.066*	-0.118***	-0.151***	-0.232***	-0.151***
	(0.034)	(0.038)	(0.046)	(0.051)	(0.045)
Cohort 2	-0.049*	-0.093***	-0.121***	-0.169***	-0.121***
	(0.025)	(0.029)	(0.035)	(0.037)	(0.035)
Cohort3	-0.009	-0.033	-0.050+	-0.116***	-0.050*
	(0.020)	(0.022)	(0.027)	(0.029)	(0.025)
Cohort 4	-0.015	-0.032+	-0.048*	-0.098***	-0.048*
	(0.016)	(0.018)	(0.021)	(0.024)	(0.020)
Cohort 5	-0.029	-0.059***	-0.083***	-0.133***	-0.083***
	(0.018)	(0.021)	(0.025)	(0.029)	(0.025)
1st stage RCT T			0.393***	0.401***	0.393***
			(0.030)	(0.032)	(0.027)
Constant	0.699***	0.753***	0.753***	1.145***	0.753***
	(0.086)	(0.091)	(0.111)	(0.069)	(0.124)
Observations	945	945	945	945	945
R-squared	0.297	0.241			

Note: Standard errors in parentheses. *** p<0.01, ** p<0.01, * p<0.05, + p<0.10

B.8 Effect Size Tables

The following tables of effect sizes present some additional detail and sensitivity analyses. Table B.12 includes the full effect size table recommended by the WWC, including group standard deviations, which space did not permit including in the main manuscript. Next, Table B.13 presents results like the main table in the text, but for missing data that was listwise deleted. Following that, Table B.14 shows effects across a range of treatment values. Finally, Table B.15 shows all available RCT courses (rather than just the matched treatment-comparison courses used in the other analyses).

Regarding Table B.15, the team who designed the RCT identified courses to be taught in both treatment and comparison sections. Instructors were also identified who were willing to teach in both treatment and control conditions. By having the same teachers teach and the same courses taught in both treatment and control conditions, it was expected that potential variation due to instructor or course effects would be removed from estimates by design. For this study, given that I was only analyzing the final two sessions of the RCT (because modality information was not available before then), not every course taught during those two sessions had a match during those sessions. The matched course or matched instructor instance may have taken place prior to this study. While most of the courses and instructors studied here were matched, a few were not. A matched design was determined to be stronger than an unmatched design, however. Even though dropping unmatched courses meant reducing sample size, a power analysis (using G*Power) suggested that the study had sufficient power to detect moderate to large effects even when dropping some data. (Note: This power analysis also informed another design-based choice to drop week one cases from the week level analysis and use those cases to develop a latent score for grade motivation, discussed elsewhere.)

Unmatched instructors were deemed to be of less importance than unmatched courses in terms of learning content in the adaptive learning system, so while unmatched courses were dropped, the five unmatched instructors (out of twenty total) were retained. Thus, instructor effects, while possible with my slightly unbalanced sample of instructors, were deemed to likely have minimal impact on the treatment of interest, particularly given that fifteen were matched.

Table B.12 Additional Detail for Hedges' g Standardized Effect Sizes (ES) for Use of Multiple Modalities, Matched RCT Data, Using Multiple Imputation for Missing Data

Analysis	N_C	$\overline{Y}_C$	SD_C	N_T	$\overline{Y}_T$	SD_T	Impact Est.	SE	*p*	Std. ES
<u>Week</u>										
Assign/quiz-w/0 & first gen.	878	0.794	0.289	219	0.943	0.228	0.150	0.065	0.020	0.539
Assign/quiz-w/0	878	0.794	0.289	219	0.896	0.228	0.102	0.054	0.061	0.366
Assign-no 0	674	0.900	0.104	191	0.991	0.087	0.091	0.024	0.000	0.903
Quiz-no 0	270	0.862	0.149	41	1.207	0.127	0.345	0.115	0.003	2.355
Assign/quiz-no 0 & first gen.	795	0.876	0.139	207	1.047	0.104	0.170	0.038	0.000	1.283
Assign/quiz-no 0	795	0.876	0.139	207	1.015	0.104	0.139	0.032	0.000	1.045
<u>Course</u>										
Grade-w/0 & first gen.	239	2.995	1.248	55	3.992	1.043	0.998	0.541	0.065	0.821
Grade-w/0	239	2.995	1.248	55	3.888	1.043	0.893	0.493	0.070	0.735
Grade-no 0 & first gen.	215	3.329	0.787	52	4.089	0.719	0.760	0.382	0.047	0.979
Grade-no 0	215	3.329	0.787	52	4.048	0.719	0.719	0.344	0.037	0.926

Note: Impact Est.=impact estimate; Std. ES=WWC percentile standardized effect size.

Table B.13 Effect Sizes for Use of Multiple Modalities for Week and Course Level Outcomes, Matched RCT Data, Using Listwise Deletion for Missing Data

Analysis	N_C	$\overline{Y}_C$	SD_C	N_T	$\overline{Y}_T$	SD_T	Impact Est.	SE	p	Std. ES
Week										
Assign/quiz-w/0	817	0.809	0.273	204	0.920	0.224	0.111	0.054	0.039	0.421
Assign-no 0	635	0.902	0.104	179	1.001	0.083	0.099	0.024	0.000	0.989
Quiz-no 0	261	0.866	0.147	36	1.408	0.128	0.542	0.259	0.036	3.739
Assign/quiz-no 0	752	0.879	0.138	193	1.015	0.091	0.136	0.032	0.000	1.044
Course										
Grade-w/0	223	3.054	1.210	52	3.646	1.073	0.592	0.404	0.143	0.498
Grade-no 0	203	3.355	0.771	49	3.973	0.727	0.618	0.309	0.045	0.808

Note: Impact Est.=impact estimate; Std. ES=WWC percentile standardized effect size.

Table B.14 Effect Sizes for Various Treatment Dosage Values for Week and Course Level Outcomes, All (Unmatched) RCT Data, Using Listwise Deletion for Missing Data

Analysis	N_C	$\overline{Y}_C$	SD_C	N_T	$\overline{Y}_T$	SD_T	Impact Est.	SE	p	Std. ES
Week-w/0										
≥1 modality use	622	0.793	0.286	399	0.854	0.220	0.061	0.029	0.038	0.232
≥5 modality use	767	0.807	0.275	254	0.898	0.223	0.092	0.044	0.038	0.348
≥7 modality use	817	0.809	0.273	204	0.920	0.224	0.111	0.054	0.039	0.421
≥9 modality use	865	0.812	0.271	156	0.960	0.220	0.148	0.072	0.041	0.562
Week-no 0										
≥1 modality use	564	0.875	0.138	381	0.949	0.118	0.074	0.017	0.000	0.571
≥5 modality use	704	0.879	0.138	241	0.991	0.103	0.112	0.026	0.000	0.862
≥7 modality use	752	0.879	0.138	193	1.015	0.091	0.136	0.032	0.000	1.044
≥9 modality use	797	0.881	0.136	148	1.064	0.092	0.183	0.045	0.000	1.403
Course-w/0										
≥1 modality use	165	3.135	1.122	110	3.476	1.277	0.341	0.231	0.139	0.287
≥23 modality use	197	3.066	1.205	78	3.496	1.140	0.430	0.292	0.142	0.361
≥34 modality use	216	3.040	1.223	59	3.544	1.027	0.503	0.345	0.144	0.424
≥38 modality use	223	3.054	1.210	52	3.646	1.073	0.592	0.404	0.143	0.498
≥68 modality use	258	3.084	1.189	17	5.154	1.168	2.071	1.452	0.154	1.739
Course-no 0										
≥1 modality use	154	3.359	0.769	98	3.741	0.754	0.382	0.187	0.041	0.498
≥23 modality use	180	3.356	0.783	72	3.820	0.710	0.464	0.228	0.042	0.607
≥34 modality use	196	3.351	0.778	56	3.882	0.707	0.531	0.265	0.045	0.695
≥38 modality use	203	3.355	0.771	49	3.973	0.727	0.618	0.309	0.045	0.808
≥68 modality use	236	3.371	0.757	16	5.316	0.865	1.945	1.063	0.067	2.539

Note: Impact Est.=impact estimate; Std. ES=WWC percentile standardized effect size.

Table B.15 Effect Sizes for Use of Multiple Modalities for Week and Course Level Outcomes, All (Unmatched) RCT Data, Using Listwise Deletion for Missing Data

Analysis	N_C	$\overline{Y}_C$	SD_C	N_T	$\overline{Y}_T$	SD_T	Impact Est.	SE	p	Std. ES
Week										
Assign/quiz-w/0	982	0.814	0.271	296	0.857	0.234	0.043	0.059	0.466	0.163
Assign/quiz-no 0	906	0.882	0.140	282	0.938	0.140	0.056	0.033	0.094	0.400
Course										
Grade-w/0	260	3.095	1.170	80	3.285	1.113	0.191	0.347	0.583	0.164
Grade-no 0	239	3.367	0.756	75	3.750	0.768	0.383	0.253	0.130	0.504

Note: Impact Est.=impact estimate; Std. ES=WWC percentile standardized effect size.

APPENDIX C

ADDITIONAL MATERIAL FOR CHAPTER 4

Additional details pertaining to the analysis in chapter four are provided below.

C.1 Variable Description

Table C.1 Variable Operationalization

Variable	Operational Notes
Week grade	Binary. Indicator of whether the student earned an A, B, or C (1) versus a D or F (0) on the mean grade for assignments and quizzes in week one of the course.
Knowledge state score	Continuous. Score earned by the student on the material covered in the adaptive learning system, updated at the beginning and end of an activity based on their performance on short formative assessments. (Also see description in Table A.1.)
Use of >1 modality	Binary. Each learning activity was offered in a variety of formats, including text, video, audio, interactive, and mixed modalities. Course designers identified Open Educational Resources (OER) in as many of these modalities for each activity as possible. Each time a student accessed material for a given activity, the modality accessed was logged.
Activity repetition	Binary. Indicator of whether the activity was repeated more than once.
Tutoring	Binary. Indicator of whether tutoring was obtained after the previous activity had started and before the next activity begins.
Race/ethnicity	Categorical split into binary indicators for White, Black, Hispanic, and Other.
Age	Continuous. In years.
Pell grant eligible	Binary. Student is eligible for a federal Pell grant.
Number of credits transferred at entry	Continuous. Official number of academic credits earned at other postsecondary institutions upon entry to this one.

C.2 Additional Descriptive Results

Figures C.1 and C.2 show the same data presented in Figure 4.2 split into the three subterms used for training the Bayesian network model (in C.1) and the other three subterms used for testing (in C.2). As with Figure 4.2, the two activities that occurred during the first week appear under the Week 1 heading, while the six activities that occurred during the second week appear under the Week 2 heading. Figure C.3 shows the full dataset through the entire the course, including all five weeks of activity in the adaptive learning system.

Ten-fold cross validation was used to validate the predictive ability of the models as described in section C.4 below but was not used for the simulated intervention predictions. Future analysis could be done using a 10-fold cross validation approach for the entire analysis where a different 10% of the dataset is held out each time for simulation testing after the model is trained on the rest of the data to determine the model parameters.

Figures C.1 and C.2 reveal some differences in the patterns of modality use and tutoring between the groups of students in the training and testing data. Fewer students in the test data opted to receive tutoring during the first two weeks of the course. Test students repeated activities and used multiple modalities within the adaptive system to a greater extent than students in the training data. Students in both groups spent roughly similar amounts of time in the adaptive learning system.

After visual inspection, week one was chosen for the Bayseian network analysis example since week one had the simplest activity structure and included both use of multiple modalities and tutoring activity.

Figure C.1 Patterns of Modality Use and Tutoring, Training Data for First Two Weeks

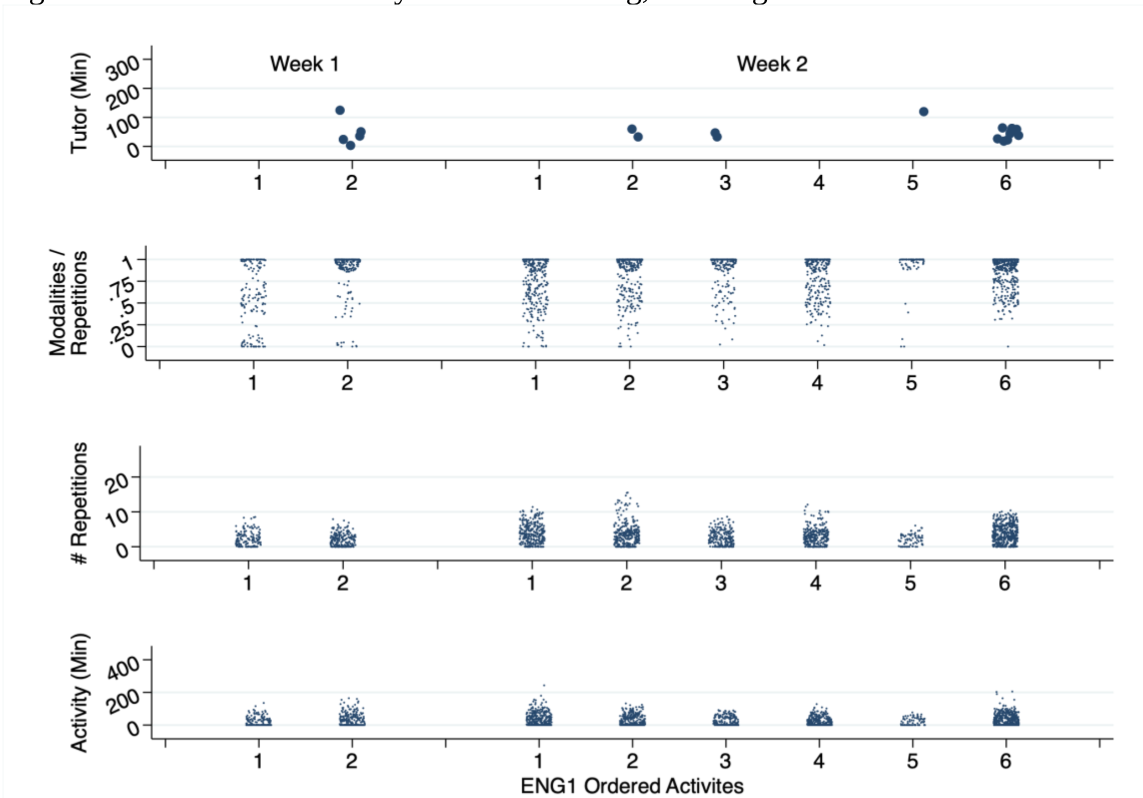


Figure C.2 Patterns of Modality Use and Tutoring, Testing Data for First Two Weeks

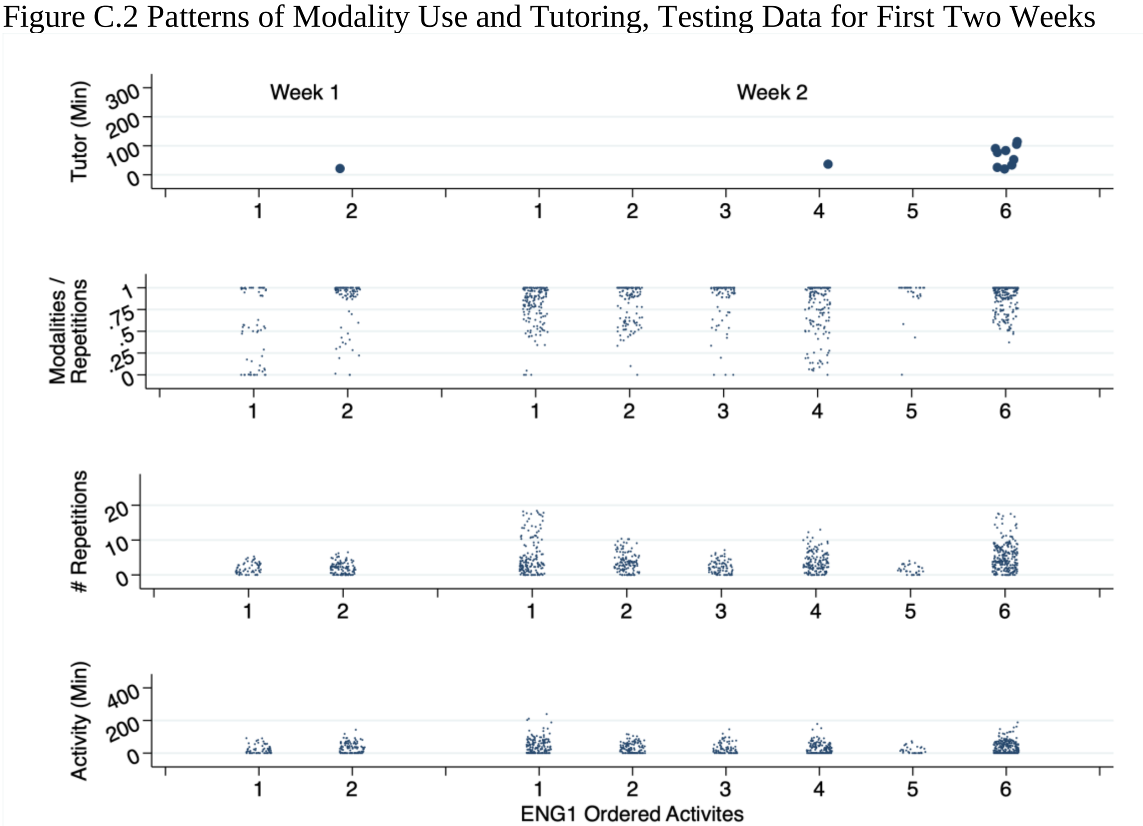
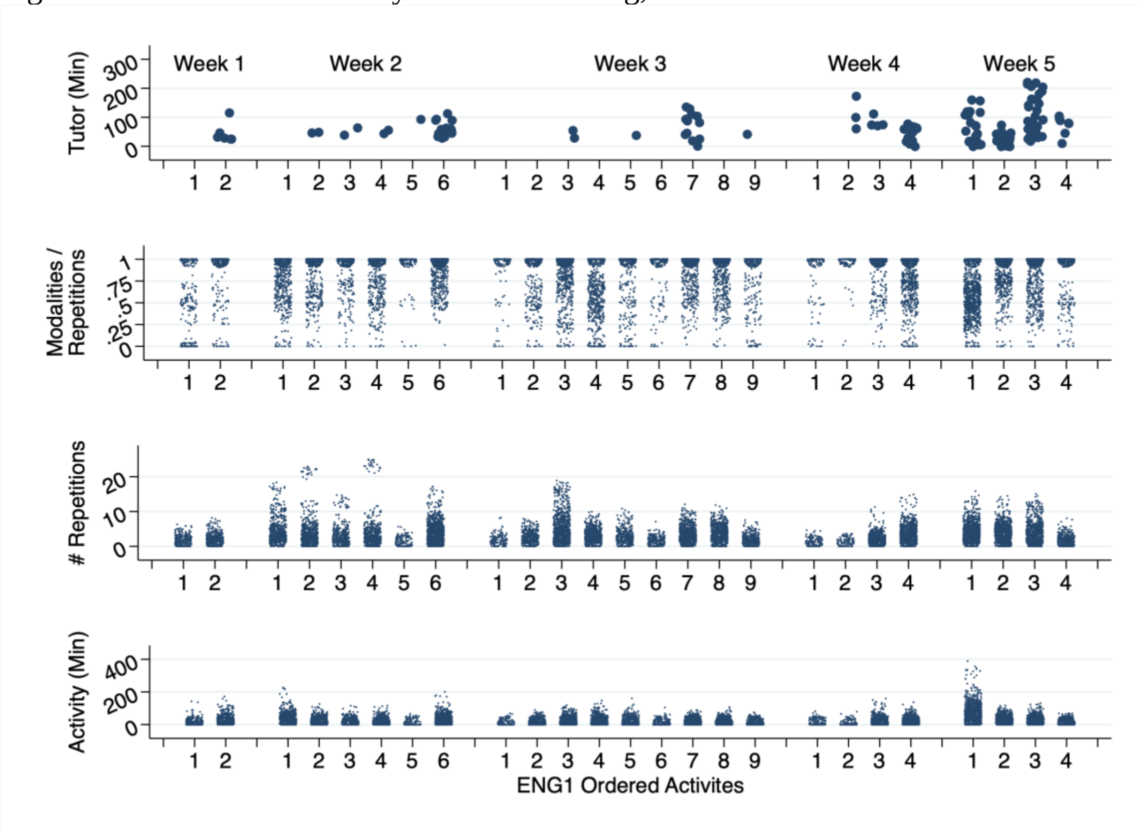


Figure C.3 Patterns of Modality Use and Tutoring, All Data for All Five Weeks



Note for Figures C.1 through C.3: Points jittered, and tutoring points enlarged for visibility. Activities shown in sequence for each week displayed. The four rows of plots from top to bottom in each figure show: 1) the amount of time each student spent receiving tutoring after beginning to work on the activity (zero tutoring times not displayed), 2) the ratio of the number of times multiple modalities were used when working on the activity to the number of repetitions of that activity overall by each student, 3) the number of times each student repeated the activity, and 4) the amount of time each student spent working on the activity overall.

C.3 Bayesian Network Simulation Analysis Approach

What follows offers additional information about using a Bayesian network technique to develop recommendations utilizing tutoring information along with system log information from the adaptive tutoring system and background information from the student information system. While several simplifying assumptions have been made to facilitate this analysis, the example illustrates possible predictive analysis that could feed prescriptive analytic information presented to students.

Two important characteristics that define a Bayesian network are that it: a) assumes the conditional independence of nodes beyond a variable's parent nodes (i.e. connected by directed arrows into that variable from another node called a parent), child nodes (i.e., connected by directed arrows out of that variable to another node called a child), and parents of child nodes; and b) assumes that the directed arrows represent causal influences. Under these assumptions, model equations corresponding to the graphical model in Figure 4.1 that describe the joint distribution of the Bayesian network include:

$Y_{11} given$

$R_1 = \beta_{10} + \beta_{11}Y_{11} + \beta_{12}D_1 + \boldsymbol{\beta_{1X}X} + \varepsilon_{1R_1}$

$D_1 = \beta_{20} + \beta_{21}Y_{11} + \boldsymbol{\beta_{2X}X} + \varepsilon_{2D_1}$

$T_1 = \beta_{30} + \beta_{31}Y_{11} + \beta_{32}D_1 + \boldsymbol{\beta_{3X}X} + \varepsilon_{3T_1}$

$Y_{12} = \beta_{40} + \beta_{41}Y_{11} + \beta_{42}R_1 + \beta_{43}D_1 + \beta_{44}T_1 + \boldsymbol{\beta_{4X}X} + \varepsilon_{4Y_{12}}$

$Y_{21} = \beta_{50} + \beta_{51}Y_{12} + \boldsymbol{\beta_{5X}X} + \varepsilon_{5Y_{21}}$

$R_2 = \beta_{60} + \beta_{61}Y_{21} + \beta_{62}D_2 + \boldsymbol{\beta_{6X}X} + \varepsilon_{6R_2}$

$D_2 = \beta_{70} + \beta_{71}Y_{21} + \boldsymbol{\beta_{7X}X} + \varepsilon_{7D_2}$

$T_2 = \beta_{80} + \beta_{81}Y_{21} + \beta_{82}D_2 + \boldsymbol{\beta_{8X}X} + \varepsilon_{8T_2}$

$Y_{22} = \beta_{90} + \beta_{91}Y_{21} + \beta_{92}R_2 + \beta_{93}D_2 + \beta_{94}T_1 + \boldsymbol{\beta_{9X}X} + \varepsilon_{9Y_{22}}$

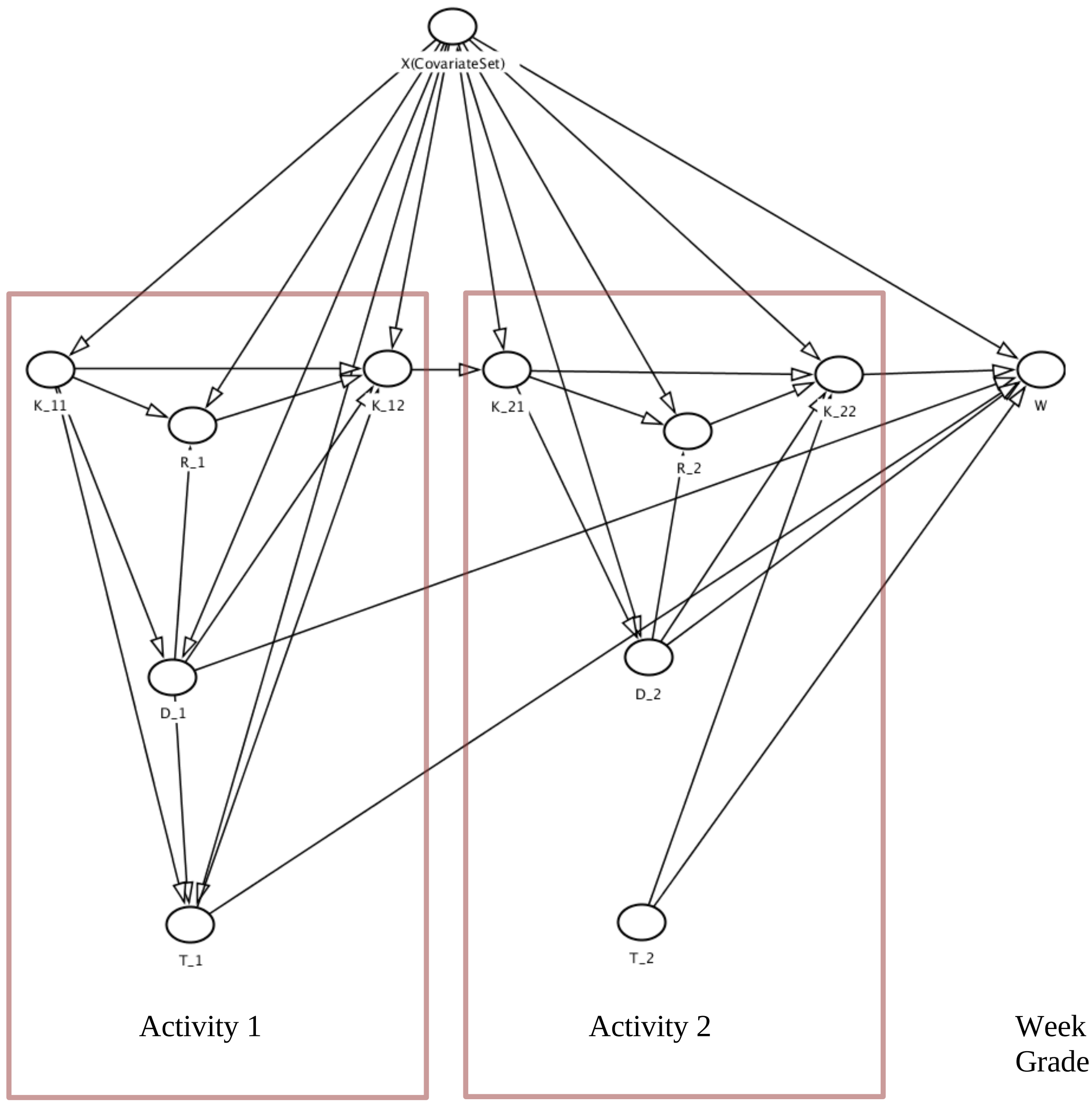
$W = \beta_{1'0} + \beta_{1'1}Y_{22} + \boldsymbol{\beta_{1'X}X} + \varepsilon_{1'W}$

Analysis proceeded in two overarching steps as illustrated in Figure 4.2 and explained further below. In the first step, parameters for the model equations were determined from the training data. In the second step, a simulation approach allowed evaluation of predicted student outcomes using the testing data under a pair of hypothetical tutoring interventions where students would either do tutoring or not do tutoring after the adaptive learning activity selected for analysis.

For step one, the model equations were evaluated using parametric assumptions about the functional form of the relationships between variables, and their parameters were stored for use in the subsequent simulation step. *Y* was evaluated using ordinary least squares regression. Since *R*, *D*, *T*, and *W* were binary, they were evaluated using logistic regression. Due to the small amount of tutoring (*T*) data available, when *T* was used as a predictor, inverse propensity score weighting was used to determine the conditional average treatment effect (CATE) aggregated across all students in the course. This weight was then used in the two simulation models where tutoring was a predictor variable (i.e., ending knowledge state and week grade). For each variable, the standard deviation of residuals given by the root mean square error term of an ordinary least squares regression (or *Root MSE* in Stata) was calculated and stored for use when predicting the corresponding error term in step two.

For step two, the submodel corresponding to Figure 4.1 for the tutoring intervention $do(T_2 = [0,1])$ is presented in Figure C.4. The equations for the intervention are the same as presented above except that either $T_2 = 0$ or $T_2 = 1$ depending on the scenario.

Figure C.4 Interventional Submodel for Week One of ENG1



In step two, with only two adaptive activities during week one of the course, the choice of the time point for a hypothetical tutoring intervention was straightforward for this example, though future investigation looking across more weeks of this course could investigate additional time points. The start of the second activity was selected for simulated evaluation and the variable at that point was identified (i.e., Y_{21}). Values for this (parent) variable were taken from the data to start the simulated dataset. The model parameters determined in step one were then used along with these actual data to predict values for all variables from the intervention time point forward. Additionally, the root mean square error values determined in step one for each variable were used to specify the distribution from which to draw values for the random error term that was added to these predictions, $N \sim (0, RMSE)$. Thus, each of the two intervention scenario value sets calculated for a given student by the simulation approximate the distributions of expected outcome values under that scenario given the error in the models.

Running this simulation 500 times produced a distribution of predicted values for the week grade for each student. The means of week grade for the treatment group that received tutoring under the simulated intervention and the comparison group that did not receive tutoring in the simulated intervention were compared visually by plotting kernel densities for both groups, statistically by conducting a t-test, and substantively by calculating the effect size using Cohen's *d*. The distance between these means offers an indication of whether to recommend tutoring to a given student. Where the prescriptive analytics indicate tutoring may be helpful, in addition to offering recommendations to students about this optimal choice, the activity identifiers could be stored for later use by faculty and other course developers. Such collections of identifiers could inform course revisions by illuminating activities where students showed the most signs of struggling.

C.4 Bayesian Network Model Validation

The five models used in the intervention simulation to predict the values of the variables in the network were each evaluated using typical goodness of fit statistics within a 10-fold cross validation procedure to determine predictive ability. Reported statistics include the adjusted R^2 (for logistic regression), R^2 (for OLS regression), and the

Bayesian information criteria (BIC) for the main model. The range of the root mean square error of the model (RMSE) across each 10% subset of the dataset in the cross validation is also reported. The logistic regression models were additionally evaluated for their classification ability using the receiver operating curve (ROC). The area under the ROC curve (AUC) gauges the predictive ability of the model to correctly classify both positive and negative outcomes across different values of the logistic classifier cutoff (which was set to the typical 0.5 when evaluating the models to determine the parameters). Results are presented in Table C.2. While there were a few higher root mean square error values in the cross-validation results, they were not typical except for the model for repetition, which is the model that performed the least well overall. However, AUC results combined with the other goodness of fit metrics suggested all five models were reasonable to use for prediction in the simulation.

Table C.2 Evaluation Statistics for Simulation Variable Models

	Use of multiple modalities (D_2)	Repetition (R_2)	Tutoring (T_2)	Ending knowledge state (K_{22})	Week grade (W)
R^2 / Adj. R^2	0.629	0.068	0.403	0.412	0.225
BIC	38.970	208.387	48.868	-70.423	36.849
RMSE	[0.019,0.329]	[0.427,0.527]	[0.015,0.236]	[0.108,0.147]	[0.016,0.316]
AUC	0.969	0.685	0.944	--	0.833

BIBLIOGRAPHY

Abell, M. M., Jung, E., & Taylor, M. (2011). Students' perceptions of classroom instructional environments in the context of "Universal Design for Learning." *Learning Environments Research*, *14*(2), 171–185. https://doi.org/10.1007/s10984-011-9090-2

Abrams, H. G., & Jernigan, L. P. (1984). Academic support services and the success of high-risk college students. *American Educational Research Journal*, *21*(2), 261–274. https://doi.org/10.2307/1162443

Alkış, N., & Temizel, T. T. (2018). The impact of motivation and personality on academic performance in online and blended learning environments. *Journal of Educational Technology & Society*, *21*(3), 35–47.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Archambault, M. (2016). *The diffusion of Universal Design for Instruction in post-secondary institutions* [Ed.D. dissertation, University of Hartford]. https://search.proquest.com/pqdtglobal/docview/1783127057/abstract/363278AEC9FF4642PQ/64

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 267–270. https://doi.org/10.1145/2330601.2330666

Aslaksen, K., & Lorås, H. (2018). The modality-specific learning style hypothesis: A mini-review. *Frontiers in Psychology*, *9*, 1538. https://doi.org/10.3389/fpsyg.2018.01538

Association on Higher Education and Disability. (2021). *AHEAD Statement on Language*. https://www.ahead.org/professional-resources/accommodations/statement-on-language

Austin, A. E., & Sorcinelli, M. D. (2013). The future of faculty development: Where are we going? *New Directions for Teaching and Learning*, *2013*(133), 85–97. https://doi.org/10.1002/tl.20048

Ayllón, S., Alsina, Á., & Colomer, J. (2019). Teachers' involvement and students' self-efficacy: Keys to achievement in higher education. *PLoS ONE*, *14*(5), 1–11. https://doi.org/10.1371/journal.pone.0216865

Barnard-Brak, L., Paton, V., & Sulak, T. (2012). The relationship of institutional distance education goals and students' requests for accommodations. *Journal of Postsecondary Education and Disability*, *25*(1), 5–19.

Basham, J. D., Blackorby, J., & Marino, M. T. (2020). Opportunity in crisis: The role of Universal Design for Learning in educational redesign. *Learning Disabilities: A Contemporary Journal*, *18*(1), 71–91.

Baumann, C., & Harvey, M. (2021). What is unique about high performing students? Exploring personality, motivation and competitiveness. *Assessment & Evaluation in Higher Education*, *0*(0), 1–13. https://doi.org/10.1080/02602938.2020.1870930

Beacham, N. A., & Alty, J. L. (2006). An investigation into the effects that digital media can have on the learning outcomes of individuals who have dyslexia. *Computers & Education*, *47*(1), 74–93. https://doi.org/10.1016/j.compedu.2004.10.006

Bears, Z. (2018, March 7). *Students demand #DebtFreeFuture from Beacon Hill lawmakers*. http://phenomonline.org/demand-a-debt-free-future/

Ben-Moshe, L., Cory, R. C., Feldbaum, M., & Sagendorf, K. (Eds.). (2005). *Building pedagogical curb cuts: Incorporating disability in the university classroom and curriculum*. Syracuse University Graduate School.

Bettencourt, G. M., Kimball, E. W., & Wells, R. S. (2018). Disability in postsecondary STEM learning environments: What faculty focus groups reveal about definitions and obstacles to effective support. *Journal of Postsecondary Education and Disability*, *31*(4), 387–400.

Bettencourt, G. M., Manly, C. A., Kimball, E., & Wells, R. S. (2020). STEM degree completion and first-generation college students: A cumulative disadvantage approach to the outcomes gap. *Review of Higher Education*, *43*(3), 753–779.

Betts, K., Miller, M., Tokuhama-Espinosa, T., Shewokis, P. A., Anderson, A., Borja, C., Galoyan, T., Delaney, B., Eigenauer, J. D., & Dekker. (2019). *International report: Neuromyths and evidence-based practices in higher education*. Online Learning Consortium. https://www.academia.edu/41867738/International_Report_Neuromyths_and_Evidence_Based_Practices_in_Higher_Education

Bittinger, J. (2016). *Intersection of the costs of disability and postsecondary education*. Association for the Study of Higher Education Annual Meeting, Columbus, OH.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, *13*(6), 4–16. https://doi.org/10.2307/1175554

Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, *10*(4), 405–418. https://doi.org/10.1109/TLT.2017.2740172

Bohman, P. (2000, August 24). *Universal design and disability access to the web*. https://web.archive.org/web/20000824092739/http://www.webaim.org/articles/webnet2000

Bongey, S. B., Cizadlo, G., & Kalnbach, L. (2010). Blended solutions: Using a supplemental online course site to deliver Universal Design for Learning (UDL). *Campus - Wide Information Systems*, *27*(1), 4–16. http://dx.doi.org.silk.library.umass.edu/10.1108/10650741011011246

Bowe, F. (2000). *Universal design in education: Teaching nontraditional students*. Bergin & Garvey.

Boysen, G. A. (2021). Lessons (not) learned: The troubling similarities between learning styles and Universal Design for Learning. *Scholarship of Teaching and Learning in Psychology*. https://doi.org/10.1037/stl0000280

Bradshaw, D. G. (2019). Proactive design and inclusive practices: Universal Design for Learning in higher education [Ed.D., California State University, Los Angeles]. In *ProQuest Dissertations and Theses*. http://search.proquest.com/docview/2457368678/abstract/1E24ECC00E6C4BC6PQ/1

Bradshaw, D. G. (2020). Examining beliefs and practices of students with hidden disabilities and Universal Design for Learning in institutions of higher education. *Journal of Higher Education Theory and Practice*, *20*(15), 12–20.

Bruns, C., Herrmann, T., Böckmann-Barthel, M., Rothkötter, H.-J., Bernarding, J., & Plaumann, M. (2021). IT support in emergency remote teaching in response to COVID-19. *GMS Journal for Medical Education*, *38*(1). https://doi.org/10.3205/zma001412

Burgstahler, S. E. (2006). The development of accessibility indicators for distance learning programs. *ALT-J: Research in Learning Technology*, *14*(1), 79–102. https://doi.org/10.1080/09687760500479753

Burgstahler, S. E. (2015). *Universal design in higher education: From principles to practice* (2nd ed.). Harvard Education Press.

Burgstahler, S. E. (2021). A universal design framework for addressing diversity, equity, and inclusion on postsecondary campuses. *Review of Disability Studies: An International Journal*, *17*(3), Article 3. https://rdsjournal.org

Burgstahler, S. E., & Cory, R. (Eds.). (2008). *Universal design in higher education: From principles to practice*. Harvard Education Press.

Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using Stata* (Revised). Stata Press.

Capp, M. J. (2017). The effectiveness of Universal Design for Learning: A meta-analysis of literature between 2013 and 2016. *International Journal of Inclusive Education*, *21*(8), 791–807. https://doi.org/10.1080/13603116.2017.1325074

CAST. (2011). *UDL guidelines—Version 2.0: Research evidence*. http://www.udlcenter.org/research/researchevidence

CAST. (2014). *UDL guidelines—Version 2.0: Principle I. Provide multiple means of representation*. Universal Design for Learning Guidelines Version 2.0. http://www.udlcenter.org/aboutudl/udlguidelines/principle1#principle1_g1

CAST. (2018). *Universal Design for Learning guidelines version 2.2*. https://udlguidelines.cast.org/

Cavanagh, T., Chen, B., Lahcen, R. A. M., & Paradiso, J. R. (2020). Constructing a design framework and pedagogical approach for adaptive learning in higher education: A practitioner's perspective. *International Review of Research in Open and Distributed Learning*, *21*(1), 173–197. https://doi.org/10.19173/irrodl.v21i1.4557

Chandler, R., Zaloudek, J. A., & Carlson, K. (2017). How do you intentionally design to maximize success in the academically diverse classroom? *New Directions for Teaching and Learning*, *2017*(151), 151–169. https://doi.org/10.1002/tl.20254

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *American Association of Higher Education Bulletin*, *39*(7), 3–7.

Christensen, C. M. (2006). The ongoing process of building a theory of disruption. *Journal of Product Innovation Management*, *23*(1), 39–55. https://doi.org/10.1111/j.1540-5885.2005.00180.x

Christensen, C. M., Anthony, S. D., & Roth, E. A. (2004). *Seeing what's next: Using the theories of innovation to predict industry change*. Harvard Business Press.

Christensen, C. M., & Eyring, H. J. (2011). *The innovative university: Changing the DNA of higher education from the inside out.* Jossey-Bass.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210.

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 134–138. https://doi.org/10.1145/2330601.2330636

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Elsevier Science & Technology.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Routledge.

Cook, L., Rumrill, P. D., & Tankersley, M. (2009). Priorities and understanding of faculty members regarding college students with disabilities. *International Journal of Teaching and Learning in Higher Education*, *21*(1), 84–96.

Cooksy, L. J., Gill, P., & Kelly, P. A. (2001). The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and Program Planning*, *24*(2), 119–128. https://doi.org/10.1016/S0149-7189(01)00003-9

Costa, R. D., Souza, G. F., Valentim, R. A. M., & Castro, T. B. (2020). The theory of learning styles applied to distance learning. *Cognitive Systems Research*, *64*, 134–145. https://doi.org/10.1016/j.cogsys.2020.08.004

Couillard, E., & Higbee, J. L. (2018). Expanding the scope of universal design: Implications for gender identity and sexual orientation. *Education Sciences*, *8*. https://eric.ed.gov/?id=EJ1200238

Crevecoeur, Y., Sorenson, S., Mayorga, V., & Gonzalez, A. (2014). Universal Design for Learning in K-12 educational settings: A review of group comparison and single-subject intervention studies. *The Journal of Special Education Apprenticeship*, *3*(2). http://scholarworks.lib.csusb.edu/josea/vol3/iss2/1

Cuevas, J. (2015). Is learning styles-based instruction effective? A comprehensive analysis of recent research on learning styles. *Theory and Research in Education*, *13*(3), 308–333. https://doi.org/10.1177/1477878515606621

Cuevas, J., & Dawson, B. L. (2018). A test of two alternative cognitive processing models: Learning styles and dual coding. *Theory and Research in Education*, *16*(1), 40–64. https://doi.org/10.1177/1477878517731450

Cumming, T. M., & Rose, M. C. (2021). Exploring Universal Design for Learning as an accessibility tool in higher education: A review of the current literature. *Australian Educational Researcher*, 1–19. https://doi.org/10.1007/s13384-021-00471-7

DaDeppo, L. M. W. (2009). Integration factors related to the academic success and intent to persist of college students with learning disabilities. *Learning Disabilities Research & Practice*, *24*(3), 122–131. https://doi.org/10.1111/j.1540-5826.2009.00286.x

Dahlstrom, E., Brooks, D. C., & Bichsel, J. (2014). *The current ecosystem of learning management systems in higher education: Student, faculty, and IT perspectives*. EDUCAUSE Center for Analysis and Research. http://www.educause.edu/ecar

Darwiche, A. (2009). *Modeling and reasoning with Bayesian networks*. Cambridge University Press.

Davies, P. L., Schelly, C. L., & Spooner, C. L. (2013). Measuring the effectiveness of Universal Design for Learning intervention in postsecondary education. *Journal of Postsecondary Education and Disability*, *26*(3), 195–220.

Dawson, S., Gašević, D., Siemens, G., & Joksimovic, S. (2014). Current state and future trends: A citation network analysis of the learning analytics field. *Proceedings of the 4th International Conference on Learning Analytics And Knowledge*, 231–240. https://doi.org/10.1145/2567574.2567585

Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019). Increasing the impact of learning analytics. *Proceedings of the 9th International Conference on Learning Analytics and Knowledge*, 446–455. https://doi.org/10.1145/3303772.3303784

De Bello, T. C. (1990). Comparison of eleven major learning styles models: Variables, appropriate populations, validity of instrumentation, and the research behind them. *Journal of Reading, Writing, and Learning Disabilities International*, *6*(3), 203–222. https://doi.org/10.1080/0748763900060302

de Freitas, S., Gibson, D., Du Plessis, C., Halloran, P., Williams, E., Ambrose, M., Dunwell, I., & Arnab, S. (2015). Foundations of dynamic learning analytics: Using university student data to increase retention. *British Journal of Educational Technology*, *46*(6), 1175–1188. https://doi.org/10.1111/bjet.12212

Dell, C. A., Dell, T. F., & Blackwell, T. L. (2015). Applying Universal Design for Learning in online courses: Pedagogical and practical considerations. *Journal of Educators Online*, *12*(2), 166–192.

DeSilva, E., Nemeroff, A., & Lopez, P. (2017). Igniting a universal design mindset on campus. *EDUCAUSE Review*. https://er.educause.edu/articles/2017/12/igniting-a-universal-design-mindset-on-campus

Dorn, E., Hancock, B., Sarakatsannis, J., & Viruleg, E. (2020, June 1). *COVID-19 and student learning in the United States: The hurt could last a lifetime*. https://www.mckinsey.com/industries/public-sector/our-insights/covid-19-and-student-learning-in-the-united-states-the-hurt-could-last-a-lifetime

Dunn, R., & Dunn, K. (2005). Thirty-five years of research on perceptual strengths: Essential strategies to promote learning. *Clearing House*, *78*(6), 1–6.

Edyburn, D. L. (2010). Would you recognize universal design for learning if you saw it? Ten propositions for new directions for the second decade of UDL. *Learning Disability Quarterly*, *33*(1), 33–41. https://doi.org/10.1177/073194871003300103

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Etzkowitz, H., Kemelgor, C., & Uzzi, B. (2000). *Athena unbound: The advancement of women in science and technology*. Cambridge University Press.

Evans, N. J., Broido, E. M., Brown, K. R., & Wilke, A. K. (2017). *Disability in higher education: A social justice approach*. Jossey-Bass.

Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering Education*, *78*(7), 674–681.

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T., & Vuorikari, R. (2016). *Research evidence on the use of learning analytics: Implications for education policy* (EUR 28294 EN; Joint Research Centre Science for Policy Report). European Union. http://oro.open.ac.uk/48173/

Fidalgo, P., & Thormann, J. (2017). Reaching students in online courses using alternative formats. *International Review of Research in Open and Distance Learning*, *18*(2). https://search.proquest.com/socscicoll/docview/1931686015/abstract/B78DE55451F4459EPQ/104

Finlay, K., & Magnusson, L. M. (2009). Implementing weak-instrument robust tests for a general class of instrumental-variables models. *The Stata Journal*, *9*(3), 398–421. https://doi.org/10.1177/1536867X0900900304

Fishkin, J. (2014). *Bottlenecks: A new theory of equal opportunity*. Oxford University Press.

Fovet, F., & Mole, H. (2013). UDL-from disabilities office to mainstream class: How the tools of a minority address the aspirations of the student body at large. *Collected Essays on Learning and Teaching*, *6*, 121–126.

Fox, A., Hedayet, M., Mansour, K. E., Kommers, S., & Wells, R. (2021). College students with disabilities experiences with financial, social, and emotional costs on campus in the United States. *International Journal of Disability, Development & Education*, 1–15. https://doi.org/10.1080/1034912x.2021.1966758

Frank, K. A., Maroulis, S. J., Duong, M. Q., & Kelcey, B. M. (2013). What would it take to change an inference? Using Rubin’s causal model to interpret the robustness of causal inferences. *Educational Evaluation and Policy Analysis*, *35*(4), 437–460.

Frazzetto, D., Nielsen, T. D., Pedersen, T. B., & Šikšnys, L. (2019). Prescriptive analytics: A survey of emerging trends and technologies. *The VLDB Journal*, *28*(4), 575–595. https://doi.org/10.1007/s00778-019-00539-y

Fulcher, K. H., & Prendergast, C. (2021). *Improving student learning at scale: A how-to guide for higher education*. Stylus Publishing, LLC.

Gagliardi, J. S. (2018). Unpacking the messiness of harnessing the analytics revolution. In J. S. Gagliardi, A. Parnell, & J. Carpenter-Hubin (Eds.), *The Analytics Revolution in Higher Education: Big Data, Organizational Learning, and Student Success* (pp. 189–200). Stylus Publishing, LLC.

Gagliardi, J. S., Parnell, A., & Carpenter-Hubin, J. (Eds.). (2018). *The analytics revolution in higher education: Big data, organizational learning, and student success*. Stylus Publishing, LLC.

Gierdowski, D. C. (2021). Accessibility hazards on the road to college. *EDUCAUSE Review*. https://er.educause.edu/blogs/2020/11/accessibility-hazards-on-the-road-to-college

Gladhart, M. A. (2010). Determining faculty needs for delivering accessible electronically delivered instruction in higher education. *Journal of Postsecondary Education and Disability*, *22*(3), 185–196.

Glass, T. (2013). *Creating learning environments for disengaged boys: Bridging the gender gap with Universal Design for Learning* [M.Ed. thesis, University of Manitoba]. https://www.proquest.com/docview/1490996489/abstract/4546C50FE6F4EB2PQ/1

Gordon, E. E., Morgan, R. R., O'Malley, C. J., & Ponticell, J. (2007). *The tutoring revolution: Applying research for best practices, policy implications, and student achievement*. Rowman & Littlefield Education.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*, 206–213.

Greenwald, A. G., & Gillmore, G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, *89*(4), 743–751. https://doi.org/10.1037/0022-0663.89.4.743

Greer, D. L., Crutchfield, S. A., & Woods, K. L. (2013). Cognitive theory of multimedia learning, instructional design principles, and students with learning disabilities in computer-based and online learning environments. *Journal of Education*, *193*(2), 41–50. https://doi.org/10.1177/002205741319300205

Hall, T. E., Cohen, N., Vue, G., & Ganley, P. (2015). Addressing learning disabilities with UDL and technology: Strategic Reader. *Learning Disability Quarterly*, *38*(2), 72–83. https://doi.org/10.1177/0731948714544375

Hamraie, A. (2017). *Building access: Universal design and the politics of disability*. University of Minnesota Press.

Hattie, J. (2015). The applicability of visible learning to higher education. *Scholarship of Teaching and Learning in Psychology*, *1*(1), 79–91. https://doi.org/10.1037/stl0000021

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), 60–87.

Heller, D. E. (Ed.). (2001). *The states and public higher education policy: Affordability, access, and accountability*. Johns Hopkins University Press.

Hendrickson, J. M., Therrien, W. J., Weeden, D. D., Pascarella, E., & Hosp, J. L. (2015). Engagement among students with intellectual disabilities and first year students: A comparison. *Journal of Student Affairs Research and Practice*, *52*(2), 204–219. https://doi.org/10.1080/19496591.2015.1041872

Herold, D. S., & Chen, T. (2021). Switching from face-to-face to online instruction midsemester: Implications for student learning. *Journal of Teaching and Learning with Technology*, *10*, 321–336.

Higbee, J. L., Bruch, P. L., & Siaka, K. (2008). Disability and diversity: Results from the Multicultural Awareness Project for Institutional Transformation. In J. L. Higbee & E. Goff (Eds.), *Pedagogy and student services for institutional transformation: Implementing Universal Design in higher education* (pp. 405–417). Center for Research on Developmental Education and Urban Literacy, University of Minnesota.

Higbee, J. L., & Goff, E. (2008). *Pedagogy and student services for institutional transformation: Implementing universal design in higher education*. Center for Research on Developmental Education and Urban Literacy. https://eric.ed.gov/?id=ED503835

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *EDUCAUSE Review*, *27*, 1–12.

How, M.-L., & Hung, W. L. D. (2019). Educational stakeholders' independent evaluation of an artificial intelligence-enabled adaptive learning system using Bayesian network predictive simulations. *Education Sciences*, *9*(2), 110. https://doi.org/10.3390/educsci9020110

Hsiao, C. (2014). *Analysis of panel data* (3rd ed.). Cambridge University Press.

Izzo, M. V. (2012). Universal Design for Learning: Enhancing achievement of students with disabilities. *Procedia Computer Science*, *14*, 343–350. https://doi.org/10.1016/j.procs.2012.10.039

Izzo, M. V., Murray, A., & Novak, J. (2008). The faculty perspective on Universal Design for Learning. *Journal of Postsecondary Education and Disability*, *21*(2), 60–72.

Jääskelä, P., Heilala, V., Kärkkäinen, T., & Häkkinen, P. (2021). Student agency analytics: Learning analytics as a tool for analysing student agency in higher education. *Behaviour & Information Technology*, *40*(8), 790–808. https://doi.org/10.1080/0144929X.2020.1725130

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 31–40. https://doi.org/10.1145/3170358.3170421

Jones, S. R. (1996). Toward inclusive theory: Disability as social construction. *NASPA Journal*, *33*, 347–354.

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence*, *2*, 100017. https://doi.org/10.1016/j.caeai.2021.100017

Kang, Z., Dragoo, M. R., Yeagle, L., Shehab, R. L., & Han Yuan. (2017). Adaptive learning pedagogy of Universal Design for Learning (UDL) for multimodal training. *Journal of Aviation / Aerospace Education & Research*, *27*(1), 23–48.

Kazis, R., Callahan, A., Davidson, C., McLeod, A., Bosworth, B., Choitz, V., & Hoops, J. (2007). *Adult learners in higher education: Barriers to success and strategies to improve results* (Occasional Paper No. 2007–03). U.S. Department of Labor. https://www.jff.org/resources/adult-learners-higher-education-barriers-success-and-strategies-improve-results/

Kelly, K., & Zakrajsek, T. D. (2020). *Advancing online teaching: Creating equity-based digital learning environments*. Stylus Publishing, LLC.

Khamparia, A., & Pandey, B. (2020). Association of learning styles with different e-learning problems: A systematic review and classification. *Education and Information Technologies*, *25*(2), 1303–1331. https://doi.org/10.1007/s10639-019-10028-y

Kimball, E. W., Wells, R. S., Ostiguy, B. J., Manly, C. A., & Lauterbach, A. (2016). Students with disabilities in higher education: A review of the literature and an agenda for future research. *Higher Education: Handbook of Theory and Research*, *31*, 91–156.

Kitto, K., Lupton, M., Davis, K., & Waters, Z. (2017). Designing for student-facing learning analytics. *Australasian Journal of Educational Technology*, *33*(5), Article 5. https://doi.org/10.14742/ajet.3607

Ko, S., & Rossen, S. (2017). *Teaching online: A practical guide*. Taylor & Francis.

Kohler, M., & Balduzzi, T. (2021). *The forgotten ones: Exploring the LXD model to address the barriers UDL creates for neurodivergent learners*. OLC Accelerate 2021 Conference, Washington, DC. https://onlinelearningconsortium.org/olc-accelerate-2021-session-page/?session=10316&kwds=

Kortering, L. J., McClannon, T. W., & Braziel, P. M. (2008). Universal Design for Learning: A look at what algebra and biology students with and without high incidence conditions are saying. *Remedial and Special Education*, *29*(6), 352–363. http://dx.doi.org/10.1177/0741932507314020

Kraemer, D. J. M., Hamilton, R., Messing, S. B., DeSantis, J. H., & Thompson-Schill, S. L. (2014). Cognitive style, cortical stimulation, and the conversion hypothesis. *Frontiers in Human Neuroscience*, *8*, Article 15. https://doi.org/10.3389/fnhum.2014.00015

Kraemer, D. J. M., Rosenberg, L. M., & Thompson-Schill, S. L. (2009). The neural correlates of visual and verbal cognitive styles. *Journal of Neuroscience*, *29*(12), 3792–3798. https://doi.org/10.1523/JNEUROSCI.4635-08.2009

Kraemer, D. J. M., Schinazi, V. R., Cawkwell, P. B., Tekriwal, A., Epstein, R. A., & Thompson-Schill, S. L. (2017). Verbalizing, visualizing, and navigating: The effect of strategies on encoding a large-scale virtual environment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 611–621. https://doi.org/10.1037/xlm0000314

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Kramer, R. (1988). *Maria Montessori: A biography*. Addison-Wesley.

Kuh, G. D., Kinzie, J., Ed, Buckley, J. A., Bridges, B. K., & Hayek, J. C. (Eds.). (2007). Piecing together the student success puzzle: Research, propositions, and recommendations. *ASHE Higher Education Report*, *32*(5), 1–182. https://doi.org/10.1002/aehe.3205

Laskey, M. L., & Hetzel, C. J. (2011). Investigating factors related to retention of at-risk college students. *Learning Assistance Review*, *16*(1), 31–43.

Lawrie, G., Marquis, E., Fuller, E., Newman, T., Qiu, M., Nomikoudis, M., Roelofs, F., & Dam, L. van. (2017). Moving towards inclusive learning and teaching: A synthesis of recent literature. *Teaching & Learning Inquiry*, *5*(1), 1–13. https://doi.org/10.20343/teachlearninqu.5.1.3

Lee, D. S., McCrary, J., Moreira, M. J., & Porter, J. (2020). Valid t-ratio inference for IV. *ArXiv:2010.05058 [Econ, q-Fin]*. http://arxiv.org/abs/2010.05058

Levey, J. A., Burgstahler, S. E., Montenegro-Montenegro, E., & Webb, A. (2021). COVID-19 pandemic: Universal design creates equitable access. *Journal of Nursing Measurement*, *29*(2), 185–187. https://doi.org/10.1891/JNM-D-21-00048

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Sage.

Lombardi, A. R., & Murray, C. (2011). Measuring university faculty attitudes toward disability: Willingness to accommodate and adopt universal design principles. *Journal of Vocational Rehabilitation*, *34*(1), 43–56.

Lombardi, A. R., Murray, C., & Gerdes, H. (2011). College faculty and inclusive instruction: Self-reported attitudes and actions pertaining to Universal Design. *Journal of Diversity in Higher Education*, *4*(4), 250–261. https://doi.org/10.1037/a0024961

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Sage Publications.

Lübke, K., Gehrke, M., Horst, J., & Szepannek, G. (2020). Why we should teach causal inference: Examples in linear regression with simulated data. *Journal of Statistics Education*, *28*(2), 133–139. https://doi.org/10.1080/10691898.2020.1752859

Mace, R. L. (1991). *Definitions: Accessible, adaptable, and universal design* (HDFS.4.91). Center for Universal Design, North Carolina State University. https://projects.ncsu.edu/design/cud/pubs_p/docs/Fact%20Sheet%206.pdf

Mangiatordi, A., & Serenelli, F. (2013). Universal Design for Learning: A meta-analytic review of 80 abstracts from peer reviewed journals. *Research on Education and Media*, *5*(1), 109–118.

Manly, C. A., & Wells, R. S. (2015). Reporting the use of multiple imputation for missing data in higher education research. *Research in Higher Education*, *56*(4), 397–409. https://doi.org/10.1007/s11162-014-9344-9

Manly, C. A., Wells, R. S., & Kommers, S. (2018). The influence of STEM definitions for research on women's college attainment. *International Journal of STEM Education*, *5*(1), 45. https://doi.org/10.1186/s40594-018-0144-1

Manly, C. A., Wladis, C., Hachey, A. C., & Conway, K. M. (2021). *Did COVID-19 exacerbate inequities? Course outcomes, on average and when considering students' prior online experience*. OLC Accelerate Conference, Washington, DC.

Marshak, L., Van Wieren, T., Ferrell, D. R., Swiss, L., & Dugan, C. (2010). Exploring barriers to college student use of disability services and accommodations. *Journal of Postsecondary Education and Disability*, *22*(3), 151–165.

Mattingly, K. D., Rice, M. C., & Berge, Z. L. (2012). Learning analytics as a tool for closing the assessment loop in higher education. *Knowledge Management & E-Learning: An International Journal*, *4*(3), 236–247.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, *63*(8), 760–769. https://doi.org/10.1037/0003-066X.63.8.760

Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology, 95*(4), 833–846. https://doi.org/10.1037/0022-0663.95.4.833

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, *90*(2), 312–320. https://doi.org/10.1037/0022-0663.90.2.312

Mayhew, M. J., Rockenbach, A. N., Bowman, N. A., Seifert, T. A. D., Wolniak, G. C., Pascarella, E. T., & Terenzini, P. T. (2016). *How college affects students: 21st century evidence that higher education works*. John Wiley & Sons, Incorporated.

McGuire, J. M. (2014). Universally accessible instruction: Oxymoron or opportunity? *Journal of Postsecondary Education and Disability*, *27*(4), 387–398.

McGuire, J. M., Scott, S. S., & Shaw, S. F. (2003). Universal design for instruction: The paradigm, its principles, and products for enhancing instructional access. *Journal of Postsecondary Education and Disability*, *17*(1), 10–20.

McGuire, J. M., Scott, S. S., & Shaw, S. F. (2006). Universal design and its applications in educational environments. *Remedial and Special Education*, *27*(3), 166–175.

McKeachie, W., & Svinicki, M. (2010). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers*. Cengage Learning.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. Guilford Press.

Mealor, A. D., Simner, J., Rothen, N., Carmichael, D. A., & Ward, J. (2016). Different dimensions of cognitive style in typical and atypical cognition: New evidence and a new measurement tool. *PLOS ONE, 11*(5), e0155483. https://doi.org/10.1371/journal.pone.0155483

Meyer, A., Rose, D. H., & Gordon, D. (2014). *Universal Design for Learning: Theory and practice*. CAST Professional Publishing.

Miller, C. A. (1999, October 7). *Self-directed web page creation: Web Certification*. ACUHO-IT 1999 Conference, San Jose, CA.

Montessori, M. (2014). *The Montessori method*. Transaction Publishers.

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research* (2nd ed.). Cambridge University Press.

Morris, J., Woodworth, C., Swier-Vosnos, A., Rossini, E., & Jackson, I. (2015). Relationship of sensory modality to retention of episodic memory. *Applied Neuropsychology: Adult*, *22*(4), 246–251. https://doi.org/10.1080/23279095.2014.914517

Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Murphy, M. P. (2021). Belief without evidence? A policy research note on Universal Design for Learning. *Policy Futures in Education*, *19*(1), 7–12. https://doi.org/10.1177/1478210320940206

Mustafa, Y. E. A., & Sharif, S. M. (2011). An approach to Adaptive E-learning Hypermedia System based on Learning Styles (AEHS-LS): Implementation and evaluation. *International Journal of Library and Information Science*, *3*(1), 15–28. https://doi.org/10.5897/IJLIS.9000009

Navarro, S. B., Zervas, P., Gesa, R. F., & Sampson, D. G. (2016). Developing teachers' competences for designing inclusive learning experiences. *Educational Technology & Society*, *19*(1), 17–27.

NCES. (2021). *Fast facts: Students with disabilities*. Digest of Education Statistics, 2019 (2021-009); U.S. Department of Education, National Center for Education Statistics. https://nces.ed.gov/fastfacts/display.asp?id=60

Newman, L. A., & Madaus, J. W. (2015). Reported accommodations and supports provided to secondary and postsecondary students with disabilities: National perspective. *Career Development and Transition for Exceptional Individuals*, *38*(3), 173–181. https://doi.org/10.1177/2165143413518235

Nielsen, J. (2000). *Designing web usability*. New Riders Publishing.

Office for Civil Rights. (2021). *Education in a pandemic: The disparate impacts of COVID-19 on America's students*. https://www2.ed.gov/about/offices/list/ocr/docs/20210608-impacts-of-covid19.pdf

Orr, A. C., & Hammig, S. B. (2009). Inclusive postsecondary strategies for teaching students with learning disabilities: A review of the literature. *Learning Disability Quarterly*, *32*(3), 181–196.

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, *17*(4), 49–64.

Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology*, *50*(1), 128–138. https://doi.org/10.1111/bjet.12592

Pardos, Z. A., Heffernan, N. T., Anderson, B., & Heffernan, C. L. (2006). Using fine-grained skill models to fit student performance with Bayesian networks. *Proceedings of the Workshop in Educational Data Mining*. 8th International Conference on Intelligent Tutoring Systems, Taiwan. https://doi.org/10.1201/b10274-36

Parker, D. R., Embry, P. B., Scott, S. S., & McGuire, J. M. (2003). *Disability service providers' perceptions about implementing universal design for instruction (UDI) on college campuses* (Tech. Rep. No. 03). Center on Postsecondary Education and Disability, University of Connecticut. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.510.3158&rep=rep1&type=pdf

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2009). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest*, *9*(3), 105–119. https://doi.org/10.1111/j.1539-6053.2009.01038.x

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669–710.

Pearl, J. (2009a). Causal inference in statistics: An overview. *Statistics Surveys*, *3*, 96–146. https://doi.org/10.1214/09-SS057

Pearl, J. (2009b). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.

Peterson, C. (2003). Bringing ADDIE to life: Instructional design at its best. *Journal of Educational Multimedia and Hypermedia*, *12*(3), 227–241.

Pistilli, M. D. (2017). Learner analytics and student success interventions. *New Directions for Higher Education*, *2017*(179), 43–52. https://doi.org/10.1002/he.20242

Pliner, S. M., & Johnson, J. R. (2004). Historical, theoretical, and foundational principles of Universal Instructional Design in higher education. *Equity & Excellence in Education*, *37*(2), 105–113. https://doi.org/10.1080/10665680490453913

Porcello, D., & Hsi, S. (2013). Crowdsourcing and curating online education resources. *Science*, *341*(6143), 240–241.

Porter, S. R. (2012). *Using instrumental variables properly to account for selection effects*. Department of Leadership, Policy, and Adult and Higher Education, North Carolina State University. https://eric.ed.gov/?id=ED531905

Quality Matters. (2020). *Specific review standards from the QM higher education rubric, 6th edition*. Quality Matters. https://www.qualitymatters.org/qa-resources/rubric-standards/higher-ed-rubric

Rao, K., Edelen-Smith, P., & Wailehua, C.-U. (2015). Universal design for online courses: Applying principles to pedagogy. *Open Learning*, *30*(1), 35–52. https://doi.org/10.1080/02680513.2014.991300

Rao, K., Ok, M. W., & Bryant, B. R. (2014). A review of research on universal design educational models. *Remedial and Special Education*, *35*(3), 153–166. https://doi.org/10.1177/0741932513518980

Rao, K., Ok, M. W., Smith, S. J., Evmenova, A. S., & Edyburn, D. (2020). Validation of the UDL reporting criteria with extant UDL research. *Remedial & Special Education*, *41*(4), 219–230.

Rao, K., & Tanners, A. (2011). Curb cuts in cyberspace: Universal Instructional Design for online courses. *Journal of Postsecondary Education and Disability*, *24*(3), 211–229.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications.

Riener, C., & Willingham, D. (2010). The myth of learning styles. *Change*, *42*(5), 32–35.

Roberts, K. D., Park, H. J., Brown, S., & Cook, B. (2011). Universal Design for Instruction in postsecondary education: A systematic review of empirically based articles. *Journal of Postsecondary Education and Disability*, *24*(1), 5–15.

Robinson, D. E., & Wizer, D. R. (2016). Universal Design for Learning and the Quality Matters guidelines for the design and implementation of online learning events. *International Journal of Technology in Teaching and Learning*, *12*(1), 17–32.

Rose, D. H. (2001). Universal Design for Learning: Deriving guiding principles from networks that learn. *Journal of Special Education Technology*, *16*(2), 66–67.

Rose, D. H., Harbour, W., Johnston, C. S., Daley, S., & Abarbanell, L. (2006). *Universal Design for Learning in postsecondary education: Reflections on principles and their application*. National Center on Universal Design for Learning. http://www.udlcenter.org/sites/udlcenter.org/files/UDLinPostsecondary.pdf

Rose, D. H., & Meyer, A. (2002). *Teaching every student in the digital age: Universal Design for Learning*. Association for Supervision and Curriculum Development.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Schelly, C. L., Davies, P. L., & Spooner, C. L. (2011). Student perceptions of faculty implementation of Universal Design for Learning. *Journal of Postsecondary Education and Disability*, *24*(1), 17–30.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs*. American Educational Research Association.

Schwartz, C. M., Phillippe, K., Kowalski, D., & Polec, A. (2018). Evolving from reflective to predictive: Mongomery County Community College and analytics. In J. S. Gagliardi, A. Parnetll, & J. Carpenter-Hubin (Eds.), *The Analytics Revolution in Higher Education: Big Data, Organizational Learning, and Student Success* (pp. 173–188). Stylus Publishing, LLC.

Scott, S. S., McGuire, J. M., & Shaw, S. F. (2003). Universal Design for Instruction: A new paradigm for adult instruction in postsecondary education. *Remedial & Special Education*, *24*(6), 369–379. https://doi.org/10.1177/07419325030240060801

Seale, J. K. (2013). *E-learning and disability in higher education: Accessibility research and practice*. Routledge.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Company.

Shibani, A., Knight, S., & Buckingham Shum, S. (2020). Educator perspectives on learning analytics in classroom practice. *The Internet and Higher Education*, *46*, 100730. https://doi.org/10.1016/j.iheduc.2020.100730

Silva, R. (2015). *Bayesian networks and the search for causality*. Talk given at the London Bayesian Networks Meetup. http://www.homepages.ucl.ac.uk/~ucgtrbd/talks/bayesnetmeetup_causality.pdf

Silver, P., Bourke, A., & Strehorn, K. C. (1998). Universal Instructional Design in higher education: An approach for inclusion. *Equity & Excellence in Education*, *31*(2), 47–51. https://doi.org/10.1080/1066568980310206

Sin, K., & Muthu, L. (2015). Application of big data in education data mining and learning analytics – A literature review. *ICTACT Journal on Soft Computing*, *5*(4), 1035–1049. https://doi.org/10.21917/ijsc.2015.0145

SmarterServices. (2021). *SmarterMeasure assessment details*. http://www.smartermeasure.com/research/assessment-details/

Smith, F. G. (2012). Analyzing a college course that adheres to the Universal Design for Learning (UDL) framework. *Journal of the Scholarship of Teaching and Learning*, *12*(3), 31–61.

Smith, S. J., Rao, K., Lowrey, K. A., Gardner, J. E., Moore, E., Coy, K., Marino, M., & Wojcik, B. (2019). Recommendations for a national research agenda in UDL: Outcomes from the UDL-IRN preconference on research. *Journal of Disability Policy Studies*, *30*(3), 174–185.

Snyder, T. D., de Bray, C., & Dillow, S. A. (2019). *Digest of education statistics 2017* (NCES 2018-070). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018070

Snyder, T. D., & Dillow, S. A. (2013). *Digest of education statistics 2012* (NCES 2014-105). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014015

Song, M., & Herman, R. (2010). Critical issues and common pitfalls in designing and conducting impact studies in education: Lessons learned from the What Works Clearinghouse (Phase I). *Educational Evaluation and Policy Analysis*, *32*(3), 351–371. https://doi.org/10.3102/0162373710373389

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search*. MIT Press.

StataCorp. (2019). *Stata 16 base reference manual*. Stata Press.

Street, C. D., Koff, R., Fields, H., Kuehne, L., Handlin, L., Getty, M., & Parker, D. R. (2012). Expanding access to STEM for at-risk learners: A new application of Universal Design for Instruction. *Journal of Postsecondary Education and Disability*, *25*(4), 363–375.

Swanson, H. L. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes*. Guilford Press.

Tinto, V. (2006). Research and practice of student retention: What next? *Journal of College Student Retention: Research, Theory & Practice*, *8*(1), 1–19. https://doi.org/10.2190/4YNU-4TMB-22DJ-AN4W

Tobin, T. J., & Behling, K. (2018). *Reach everyone, teach everyone: Universal Design for Learning in higher education*. West Virginia University Press.

Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners*. Association for Supervision & Curriculum Development.

Trow, M. A., & Burrage, M. (2010). *Twentieth-century higher education: Elite to mass to universal*. Johns Hopkins University Press.

Tzivinikou, S. (2014). Universal Design for Learning - Application in higher education: A Greek paradigm. *Problems of Education in the 21st Century*, *60*, 156–166.

van Buuren, S. (2012). *Flexible imputation of missing data*. CRC Press.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Wang, J., Dawson, K., Saunders, K., Ritzhaupt, A. D., Antonenko, P. "Pasha," Lombardino, L., Keil, A., Agacli-Dogan, N., Luo, W., Cheng, L., & Davis, R. O. (2018). Investigating the effects of modality and multimedia on the learning performance of college students with dyslexia. *Journal of Special Education Technology*, *33*(3), 182–193. https://doi.org/10.1177/0162643418754530

Webb, K. K., & Hoover, J. (2015). Universal Design for Learning (UDL) in the academic library: A methodology for mapping multiple means of representation in library tutorials. *College and Research Libraries*, *76*(4), 537–553. https://doi.org/10.5860/crl.76.4.537

Wellman, G. S., & Marcinkiewicz, H. (2004). Online learning and time-on-task: Impact of proctored vs. un-proctored testing. *Journal of Asynchronous Learning Networks*, *8*(4), 93–105.

Wessel, R. D., Jones, J. A., Markle, L., & Westfall, C. (2009). Retention and graduation of students with disabilities: Facilitating student success. *Journal of Postsecondary Education and Disability*, *21*(3), 116–125.

What Works Clearinghouse. (2017). *Review protocol for using technology to support postsecondary student learning practice guide, version 1.1*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Document/284

What Works Clearinghouse. (2020a). *What Works Clearinghouse procedures handbook, version 4.1*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/handbooks

What Works Clearinghouse. (2020b). *What Works Clearinghouse standards handbook, version 4.1*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/handbooks

Wladis, C., Hachey, A. C., & Conway, K. M. (2015). The representation of minority, female, and non-traditional STEM majors in the online environment at community colleges: A nationally representative study. *Community College Review*, *43*(1), 89–114. https://doi.org/10.1177/0091552114555904

Wood, D., & Wood, H. (1996). Vygotsky, tutoring and learning. *Oxford Review of Education*, *22*(1), 5.

Xenos, M. (2004). Prediction and assessment of student behaviour in open and distance education in computers using Bayesian networks. *Computers & Education*, *43*(4), 345–359. https://doi.org/10.1016/j.compedu.2003.09.005

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education*, *29*(Supplement C), 23–30. https://doi.org/10.1016/j.iheduc.2015.11.003

Zhang, H., Nurius, P., Sefidgar, Y., Morris, M., Balasubramanian, S., Brown, J., Dey, A. K., Kuehn, K., Riskin, E., Xu, X., & Mankoff, J. (2021). How does COVID-19 impact students with disabilities/health concerns? *ArXiv:2005.05438 [Cs]*. http://arxiv.org/abs/2005.05438

Zilvinskis, J., Willis, J., & Borden, V. M. H. (2017). An overview of learning analytics. *New Directions for Higher Education*, *2017*(179), 9–17. https://doi.org/10.1002/he.20239