ABSTRACT

MOORE, ROBERT LAMONT. Examining the Influence of Massive Open Online Course Pacing Condition on the Demonstration of Cognitive Presence (Under the direction of Dr. Kevin Oliver).

This research study examines the influence of the pacing condition of massive open online courses (MOOCs), be they self-paced or instructor-paced, on the demonstration of cognitive presence in such courses. The research is informed by the Community of Inquiry (CoI) model of Garrison, Anderson, and Archer (2001), a framework that identifies three domains—social, teaching, and cognitive presence—that are necessary for online learning. This study utilizes an embedded multiple-case research design to allow a thorough understanding of the area of interest, specifically, cognitive presence in MOOCs. The selected cases are three MOOCs offered by Harvard University. Each one was offered in both instructor- and self-paced pacing conditions, for a total of six courses under review in this study. The analysis of the log files and discussion forums for each of these courses presents an opportunity to explore similarities and differences in learner engagement. This study analyzed 57,650 discussion posts generated by 13,495 students across these six courses. The analysis of discussion forums within MOOCs presents many logistical challenges, resulting chiefly from the size of the datasets involved. This study explores ways in which automatic text analysis tools can be used to aid in the analysis and subsequent identification of evidence of cognitive presence in MOOCs.

Examining the Influence of Massive Open Online Course Pacing Condition on the Demonstration of Cognitive Presence

by
Robert Lamont Moore

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Curriculum and Instruction

Raleigh, North Carolina

2018

APPROVED BY:

Kevin M. Oliver Committee Chair	Sherry Booth Freeman
Shaun Kellogg	Brad Mehlenbacher

DEDICATION

This dissertation is dedicated to two people who, unfortunately, are not alive to witness this personal accomplishment: my god-aunt Angie Hill and my uncle Larry Thompson. Even though they are not here, I know that they are looking down on me and smiling.

Aunt Angie was an amazing woman. She and I shared a birthday, which may explain why we got along so well! Aunt Angie grew up with my mother and was one of her best friends. I have many pictures from when I was a baby and am being held by Aunt Angie or one of her daughters, Stacy and Kelly Douglas. I can recall visiting the Hill family on weekend trips to Frederick, Maryland, and I always enjoyed the time I spent with them. As I got older, Aunt Angie and I got closer. I remember when I was 16 and finally got to stay home alone for the weekend. The highlight of that weekend was going over to Aunt Angie's house and watching movies with her and Stacy.

When I got my first cell phone as a freshman at UNC-Chapel Hill, Aunt Angie was one of the first people I called. I had many conversations with her as I made the mile-long trek from main campus back to my South Campus dorms. Aunt Angie was diagnosed with breast cancer during my junior year, and I can remember several conversations we had as I was nearing my 21st birthday. She and I both loved to dance, so I told her that in honor of our birthday, I was going to rent out a club where we could dance the night away. I hadn't told my parents about these plans, and my mom thought something was wrong with Aunt Angie's memory because she kept mentioning this huge party that was going to be held for her birthday. Unfortunately, Aunt Angie passed away a week before our joint birthday. Nonetheless, I still had the party in her memory and danced for the both of us!

Like Aunt Angie, Uncle Larry was a childhood friend of my mother. I have many memories of visits to Uncle Larry's house. One of them involves going to watch the 1991 Super Bowl on his giant screen TV—it was the largest TV I had ever seen! I also remember Fourth of July cookouts where he would have a bushel of crabs and I would get to eat as many as I wanted. For a young kid, this was amazing! As I moved through graduate school and earned academic achievements, Uncle Larry was one of the first people with whom I would share school-related news. And the sharing went both ways. I can remember how excited he was to tell me about his own accomplishments, such as being a first ballot inductee into the Alvin G. Quinn Frederick County Sports Hall of Fame or completing his psychology undergraduate degree. It became our routine: I would call Uncle Larry as I made the drive from Chapel Hill to Raleigh for my classes at N.C. State. During these conversations, we covered everything from the Redskins to college sports (he was a Duke fan, which resulted in much trash talking), from politics to life lessons. He was always willing to share his perspective, and his support and encouraging words helped me navigate many of the challenges I have faced in my adult life. I was working on my first book chapter when he passed away. He and I had talked extensively about this particular publication; sadly, he passed before he got a chance to see it in print. That publication is dedicated to him as well.

Both Aunt Angie and Uncle Larry placed a high value on education. While I did not decide to pursue my doctorate because of them, in many ways I earned it for them. This dissertation is the culmination of a lifelong pursuit of education that has been nurtured and developed through the sacrifices of my parents. The dedication of this dissertation to the memory of my aunt and uncle serves as a memoriam of the love and dedication they gave to both to me and, more importantly, my family.

BIOGRAPHY

Rob Moore was raised in Gaithersburg, Maryland, in Montgomery County. He grew up playing soccer and spent many weekends attending the games of his younger sister, Andrea. After attending public school, Rob attended Saint Albans School (STA) in Washington, D.C., for high school. This transition to private school came at great personal and financial sacrifice to his parents. These four years at STA propelled Rob to many of his current successes. At STA, and for the first time in his life, Rob was challenged mentally, academically, and athletically. As a member of arguably 'the best class ever' at STA, Rob learned how to strive to be the best both in the classroom and on the athletic fields. Whether it was in the classroom with Malcolm Lester and Coach David Padilla, or on the track being pushed by Coaches Skip Grant and Doug Boswell, Rob learned how to succeed.

While many of his classmates went to Ivy League schools, Rob decided to attend the University of North Carolina at Chapel Hill. When asked by classmates why he chose UNC, Rob simply responded, "it is 67 percent female." At UNC, Rob put much of what he learned at STA to use in the formation of two successful student organizations. These organizations allowed Rob to have a lasting impact on the UNC community. After graduating from UNC, Rob began his full-time professional career there in 2004 as a an office assistant.

Rob's second job at UNC changed his entire career trajectory. While working as the Resource Center Manager in the Department of Romance Languages and Literatures, he started collaborating with Dr. Glynis Cowell, the director of the Spanish Language Program. Together, they transitioned a face-to-face introductory Spanish course into a hybrid course that included an online version, introducing Rob to the field of instructional technology.

In the fall of 2010, Rob accepted a position at the UNC School of Government as an instructional technology developer. He was immediately thrown into the proverbial fire by having to take the lead on the development of ethics modules for individuals covered under North Carolina's State Government Ethics Act. Each of these online modules provided the two-hour training that was mandatory for the nearly 10,000 public servants and legislative employees subject to this act. After completing this project, Rob realized that he had the skills needed to be an instructional designer; he just needed the targeted education credentials to make this a career.

Rob is an award-winning instructional designer, having received recognition for his work at the national (Association for Educational Communications and Technology (AECT)), campus (UNC-Chapel Hill), and departmental levels (Department of Romance Languages and School of Government). In his current role at UNC, Rob collaborates with faculty members on integrating innovative technology to support their instruction in face-to-face, blended, and online learning environments. His primary responsibility involves providing consultations to faculty on their instructional needs, from projects as small as creating a graphic for a PowerPoint presentation to as complex as developing a multi-module, self-paced online training program. Rob has presented at local, state, and national conferences on topics that include transitioning face-to-face courses to online courses and the UNC Romance Languages Sakai Pilot that transitioned more than 6,000 undergraduates from Blackboard to multi-section Sakai umbrella course sites. He has also developed self-paced e-learning modules and facilitated webinars on teaching technologies and best practices, including Qualtrics and PowerPoint.

ACKNOWLEDGMENTS

I would not be where I am today without the support and personal sacrifices of many people in my life. This acknowledgments section cannot even begin to articulate the appreciation I have for each of the individuals mentioned.

My family has made great sacrifices for me, and for that I am grateful. To my younger sister, Andrea, I know that I have not said this to you much, but please know that I am incredibly proud of all that you have accomplished in your life. To my parents, thank you for all that you have done to put me in the position I am now in. To this end, there are way too many things to enumerate, but the most significant sacrifice you made—and the one that had the most impact on my life—was likely the decision to send me to St. Albans (STA) for high school. Those four years shaped the person I have become today unlike any other period in my life.

But the wise parenting decisions started even earlier, for example, the no-TV rule. I hated it at the time, but this fostered a love of reading that has served me well in life, particularly in graduate school. Thank you, mom and dad, for listening to me when I struggled through life's challenges. Thank you for so much more. I love you. And mom, thank you for putting so many positive people in my life: Aunt Angie, Uncle Larry, and Aunt Barbara (both of them) to name a few. To Uncle Henry and Aunt Tracy, thank you for your support and love over the years. In the words of Uncle Henry, "One Love!" To my older sister, Tanya, my brother-in-law Roman, and my little buddies Xavier and Zuri, thanks for your love! You are all so very special to me.

My time at STA was a particularly impactful time in my life, and I am forever grateful for it. I am especially grateful to Reverend Heischmann: you did so much and made

STA such a memorable place for me. To Coach Malcolm Lester and Dean Anderoli ("Roli"), I thank you both for challenging me in the classroom in ways I had never been challenged before. Coach Grant, just know that I would not be where I am today were it not for your influence. You taught me how to be a true "STA man" and how to find my limits and push beyond them. To Coach Doug Boswell and Coach Kevin McShane, thank you for teaching me what it meant to be a good teammate, a strong competitor, and a better man. And finally, to my STA brothers, thanks to each and every one of you for pushing me to be better and always having my back. You are the brothers I never had. Our bond will always be strong, and I appreciate all of you more than you will know.

To my partner in crime, Dan, thanks for all the wild adventures. For some reason, every crazy or hilarious situation I end up in usually involves you. From the parties at Dugan's house to the races down 495 to our "international" adventure to Canada, you were always in my corner. You helped me get through this graduate school process and I cannot wait to see what trouble you and I get into in the future.

I literally would not be writing this dissertation had it not been for people I met during my studies at East Carolina University, specifically, Drs. Slagter van Tryon, Brown, and Sugar. I so appreciate each of you taking a chance on an underachieving undergraduate student who showed a bit of graduate student promise. Thank you for all you taught me in the classroom and for how you welcomed me with open arms into the AECT professional family.

To my UNC support system, first, Dr. Ervin. You helped me become the leader I am today and provided so much guidance as I worked my way through UNC. Without you, I would never have graduated, and none of my current success would have been possible

without that undergraduate foundation. But you gave me so much more, and for that I will always be grateful. To Mrs. Fuller and Dr. Cookie Newsom, I so appreciate of all our chats and conversations. You were always in my corner, and I am so grateful for that. To Dr. Glynis Cowell, you were a wonderful work colleague and have become a great friend. We accomplished a lot together, and I very much appreciate the opportunities you gave me; they allowed me to refine my skills, and I evolves as an instructor as a result. It was through my experiences collaborating with you that I was able to find my passion and my true career path. You helped set me on my current trajectory, and for that I am forever in your debt. To Georgia Allen, thanks for being the best boss and an even better friend. You took a big chance in hiring me seven years ago, and your constant support personally, academically, and professionally has allowed me to pursue this doctoral degree.

My work has benefited from the input, feedback, and support of many others. I want to acknowledge the amazing editors who have worked on so many of my articles, book chapters, and conference proposals. In particular, Lily Stephens and Viktor Agabekov, thank you for all the time you both spent editing my work. We have had some great successes, and I want you to know that they not have been possible without your assistance. I also want to acknowledge Melissa Twomey, who did extensive review of my dissertation proposal, and Abby Meister, who worked her graphical genius in designing the graphics for this dissertation. Dr. Cathy Zimmer, thank you for all the time you spent helping me work through the data analysis for this dissertation. I learned so much from you and could never have completed the analysis without your patience and expertise.

My doctoral committee has made this experience enjoyable, and I am incredibly grateful for that. Kevin, thank you for being an amazing mentor and supervisor. I have learned so much from you and appreciate all of the emails, consultations, and advice you have provided over the last four years. Working as your teaching assistant has been a great experience, and I feel I am better prepared to teach on my own having observed your instructional approach over the last two years. To Shaun, thank you for helping me work through my ideas and grow as a scholar in the MOOC field. You provided a level of expertise that was so very much appreciated and allowed me to produce such a great dissertation. To Sherry, thank you for your encouraging words and support throughout my dissertation process. Your comments and feedback went a long way in helping me complete my dissertation. To Brad, thank you for your challenging questions. You kept me honest, as they say, throughout the process, and I appreciate that immensely.

This thank-you list is in no way exhaustive. There are bound to be people whose names I neglected to mention due to the constraints of time. To everyone who falls into this category, please know that your contributions are known, valued, and appreciated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

In this study, I compare student engagement in discussion forums and the demonstration of cognitive presence between two massive open online courses (MOOCs) with different pacing conditions (instructor-paced and self-paced). I also examine potential relationships between student engagement in discussion forums and the demonstration of cognitive presence. In the following sections, I provide an overview of the theoretical framework for the study, as well as an explanation of the statement of the problem and the purpose of the study. I also highlight the significance of the study and the research questions that frame this study.

Increase in Distance Education Enrollments

In its fall 2014 enrollment report, the National Center for Education Statistics (NCES) found that higher education institutions, including community colleges, colleges, and universities, had more than more than 5.7 million students enrolled in distance education courses. In that same year, more than 2.6 million, or 15.6 percent, of undergraduates attending these institutions enrolled in at least one distance education course, and almost 2 million, or 12.1 percent, completed a fully online degree program (NCES, 2016). Distance education is now seen as an effective way for learners to gain new skills, advance their careers, and obtain the education necessary to stay competitive in our rapidly technologically-driven society. For these non-traditional students, distance education—defined as instruction where there is a separation between learners and instructors and where content can be delivered asynchronously, synchronously, or by a combination of the two in a blended, hybrid instructional model—better aligns with their needs. Instructors in these online learning environments, while blazing new trails, must nonetheless adapt their pedagogical approaches to foster the same critical thinking that occurs within face-to-face learning environments.

Online learning has also accelerated the growth of for-profit online colleges and universities, institutions that did not exist twenty years ago but now represent an important segment of the higher education community (Picciano, 2012). A report on the Carnegie Classification of Institutions of Higher Education noted that of the 483 institutions classified for the first time, 77 percent were for-profit institutions (Pence, 2012), reinforcing this spike in for-profit institutions. Traditional brick and mortar colleges and universities initially ignored these for-profit innovators—until they began drawing in students who had, up to that point, pursued exclusively residential degrees. Now, the Sloan Consortium estimates that 25 billion dollars per year is generated from tuition revenue in the online market (Gallagher & LaBrie, 2012), including online offerings from residential institutions.

Gallagher and LaBrie (2012) attribute this growing acceptance for online programs in part to the “most elite and well-branded institutions in American higher education entering into the online course development market” (p. 66). The two researchers are referring to early work done at Stanford University, Harvard University, and the Massachusetts Institute of Technology in the massive open online course (MOOC) space specifically. Thus, it seems fitting that the focus of my dissertation would be on courses offered by Harvard. When such well-respected institutions elected to create distance education programs, the higher education landscape was significantly altered. Gallagher and LaBrie (2012) observed that the involvement of these elite institutions in online education "[led to an] increase [in] credibility of online education in the mind of the public; attract[ed] great attention from peers at all levels of higher education; [and] means that these institutions are now competitors in student markets" (p. 66).

The Massive Open Online Course (MOOC)

The massive open online course, or MOOC, represents a specific type of online learning environment. The term “MOOC” entered the higher education vernacular in fall 2008, around the same time the “Connectivism and Connective Knowledge (CCK08)” course was first facilitated by George Siemens and Stephen Downes (Fini, 2009). The CCK08 course, as Fini explains, was unique in that it was not only offered for course credit through the University of Manitoba, but it was also made available to anyone in the world at no cost. Fini (2009) references research by Siemens and Downes when explaining that the “2,200 people [who] signed up, and [the] hundreds of people from around the world [who] participated in the CCK08 course . . . inspired the massive open online course (MOOC) definition” (p. 3). Thus, the MOOC has become the latest “star” in the online education constellation, continuing the field’s evolution and aligning the needs of learners with enhanced access to information. The influence of elite schools on distance learning is further demonstrated by the recent uptick in interest in and attention for MOOCs; nearly any article that mentions MOOCs and their origin references either Stanford’s or MIT’s initial MOOC courses.

Multiple variations of the acronym can be found, from “traditional MOOC” to “c-MOOC” to, simply, “MOOC”. Rodriguez (2012) classifies the CCK08 course as a “c-MOOC,” or connectivist Massive Open Online Course, “. . . that [does not] align with the course content or the instructor, but [instead aligns] to other learners and their knowledge” (p. 2). Pence (2012) would classify the CCK08 course as a “traditional MOOC” and describes MOOCs of this model as “highly social in format, [where] students create much of the course material, and the interaction resembles that in a massively multiplayer online game” (p. 28). Pence offers a second model, a mere “MOOC,” distinguishable by “formalized curricula, uniform content delivery, and examinations to verify that students have completed the learning objectives,” with a structure similar to a large on-site lecture class, despite being delivered in an online format (p. 28).

The initial MOOCs had astronomical enrollment numbers, leaving many to wonder whether they would be the new higher education revolution. Chuang and Ho (2016) report that between the years of 2012 and 2016, Harvard and MIT reached 4.5 million participants through 290 courses and issued 245,000 certificates to learners who completed course requirements. Though some of the initial interest has waned, MOOCs still provide a rich research opportunity, with multiple lines of inquiry available. If we evaluate MOOCs as we would typical online courses, the issue of attrition rate becomes significant, and it is simple to see how a MOOC could be considered unsuccessful. However, if we shift our definition of success within a MOOC to look not at the course’s completion rates but, rather, at whether or not students met their learning goals by participating in a self-selected portion of the course, we may be able to discover ways to further maximize this type of large-scale online learning environment. In this latter scenario, MOOCs may present the optimal blend of openness, connection-building, and learning opportunities for adult learners.

Pacing Conditions

Currently, there are two pacing conditions among MOOCs: instructor-paced and self-paced. An instructor-paced MOOC, also referred to as a live, session-based, or cohort MOOC, will have a specific start and end date and a defined enrollment period. These courses will have a release date (i.e., date on which the course is opened to students) and specific course deadlines, similar to what one would find in an on-campus course. While students may join at any time, many them will start on the launch date and form a cohort. This approach has been linked to higher student retention rates (Sharif & Magrill, 2015). When an instructor-paced course has passed its final deadline, it is archived. Learners can still sign up for the course, but they will not be able to earn a certificate. Enrolled learners are able to access the archived version, but new learners cannot enroll in this version of the course. Alternatively, there are self-paced, or on-demand, MOOCs. In these courses everything is released at once and there are no deadlines. Most self-paced courses do still end at some point, for practical reasons (e.g., forum support and research cycles), and they can also be archived when they end. The major providers of MOOCs, edX and Coursera, both offer MOOCs in the self-paced structure. Typically, these self-paced MOOCs feature the same content and activity requirements as their instructor-paced counterparts, with the most significant difference being the loss of the cohort moving through the course at the same time. Additionally, many of the self-paced MOOCs do not offer certificates of completion, though some do feature a teaching assistant who will have a presence in the course's discussion forums.

In this study I identified a set of HarvardX courses that were offered in both the instructor-paced and self-paced structures. This allowed me to make comparisons between these courses on important outcomes such as discussion forum activity and post characteristics and demonstrated cognitive presence to determine if one structure might be better than another. These courses were provided through a data use agreement between N.C.

State University and Harvard University’s Vice Provost for Advances in Learning (VPAL) Research Team.

Activity in MOOC Discussion Forums

The first area of interest for this study centers on activity within discussion forums. The current MOOC research tends to emphasize forum activity, given that it is the primary source of student engagement. Within online course discussion forums, students are given a space for peer connection and exchange of ideas (Akcaoglu & Lee, 2016; Kent, Laslo, & Rafaeli, 2016; Martin & Ndoye, 2016; Sharif & Magrill, 2015) through conversation and interactions with one another (Beckmann & Weber, 2016; Kent et al., 2016). Gunawardena, Lowe, and Anderson (1997) identify asynchronous discussion forums as a valuable pedagogical tool because they allow students to engage in shared knowledge creation despite the transactional distance between them. In forums, students can collaborate in knowledge construction that leads to higher levels of cognitive presence (Kent et al., 2016).

In this study, I use the individual posts as my unit of analysis and look at the substantive quality of a given post (e.g., length, type of words used); whether the post was made to a peer or to an instructor; and how much interaction, through replies, that post received. I explain my collection and analysis of this data in Chapter 3.

Community of Inquiry (CoI) Theoretical Framework

The Community of Inquiry (CoI) theoretical framework informs this research (Garrison, Anderson, & Archer, 2001). CoI has been widely researched and used to describe and understand online learning, specifically, the role of discussion forums, making it an appropriate framework for this study. Understanding that there are unique challenges in developing higher-order learning in online learning environments, Garrison, Anderson, and

Archer (2001) developed CoI. This framework can help describe the engagement in distance environments such as MOOCs through the consideration of engagement with teachers, peers, and with the content. The CoI model consists of the three domains of social presence, teaching presence, and cognitive presence, explained in more detail in Chapter 2. This framework has become the most widely cited model for understanding learning through online discussions (Breivik, 2016) and is viewed as an "essential context for higher-order learning" (Garrison et al., 2001, p. 7). The CoI model defines a community of inquiry as "a group of individuals who collaboratively engage in purposeful critical discourse and reflection to construct personal meaning and confirm mutual understanding" (Garrison, 2011, p. 2). In the context of this research, the community of inquiry is each course discussion forum with the unit of analysis, the individual posts within that specific forum. The CoI model incorporates a constructivist approach to learning and recognizes that interaction and discourse play a fundamental role in higher-order learning, along with structure (design) and leadership (facilitation and direction) (Garrison, 2007).

Teaching Presence. "Teaching presence" describes the role of instructors in course design, organization, and delivery, as well as the instructions that guide social and cognitive presences to desired learning outcomes (Anderson et al., 2001). Garrison (2007) further defines teaching presence as a significant factor for a student's satisfaction, perceptions of learning, and sense of community. Teaching presence brings the social and cognitive presences together and accounts for learners' needs and capabilities (Garrison & Anderson, 2003; Kreijns et al., 2014).

Social Presence. "Social presence" is an important aspect of online learning and is essential for high-quality asynchronous discussion forums (Akcaoglu & Lee, 2016). This presence focuses on the basic social relationships among members of the learning community and the social climate that contributes to the success of learning and the attainment of learning objectives (Rourke, Anderson, Garrison, & Archer, 1999). Social presence is the ability to project one's self and establish personal and purposeful relationships (Garrison, 2007). The three most important aspects of social presence are effective communication, open communication, and group cohesion (Garrison, 2007). Without the interaction created through social presence, the resultant collaboration and knowledge construction needed for cognitive presence cannot exist within a course (Kreijns, Van Acker, Vermeulen, & Van Buuren, 2014).

Cognitive Presence. "Cognitive presence" is defined as the exploration, construction, resolution, and confirmation of understanding through collaboration and reflection in a community of inquiry (Garrison, 2007). Cognitive presence is grounded in critical thinking literature (Garrison, Anderson, & Archer, 1999; 2001) and operationalized through the cycle of practical inquiry, where participants move deliberately from understanding the program or issue through to exploration, integration, and application (Garrison, 2007; Gibson, Ice, Mitchell, & Kupczynski, 2012). Schrire (2004) suggests the practical inquiry (PI) model as an effective way to analyze the cognitive dimension within a discussion forum. Cognitive presence is a central dimension of the PI model that describes the learning phases from the initial practical inquiry to eventual knowledge construction and problem solving (Garrison et al., 2001). Joksimovic et al. (2014) assert that "cognitive presence is recognized as a core concept in the CoI definition, and is focused on the processes of higher-order thinking" (p. 2).

Intersection of Presences. To achieve an optimal educational experience, all three presences need to be accounted for within a course. It is at the intersection of the presences that specific learning outcomes can be observed, as the presences are interconnected. The intersection of social presence and cognitive presence is important, as students are not online just for purely social reasons (Garrison, 2007). In further exploring the relationship between the three presences, studies have found that social presence is a mediator between teaching presence and cognitive presence, and teaching presence causally influences both social and cognitive presence (Garrison, Anderson, & Archer, 2010; Kreijns et al., 2014; Shea & Bidjerano, 2009). Additionally, Garrison, Cleveland-Innes, and Fung (2010) suggest that the central role of teaching presence is in establishing and maintaining social and cognitive presence. This is substantiated by the work of Gibson et al. (2012), who found that “design, facilitation, and direction laid out for the cognitive and social presences create the navigational map for a learner” (p. 2). The connection between cognitive presence and teaching presence is further clarified by Song and Yuan (2015), who explain that these two presences “go hand-in-hand to ensure the right selection, chunking and sequencing of materials” (p. 732) and explain the connection between cognitive presence and social presence as the mechanism that “creates a supportive environment for learners to interact with the learning content, as well as with each other” (p. 732).

Cognitive Presence in MOOC Discussion Forums

The second interest area within this study is cognitive presence, which will be described using cognitive processing, a measure of specific dictionary words by my analysis tool, the Linguistic Inquiry Word Count (LIWC). I explain this tool and how I used it in Chapters 2 and 3. The use of LIWC made it possible for me to compare demonstrated cognitive processing between the pacing conditions (instructor-paced versus self-paced) and to look for relationships between discussion forum activity, post characteristics, and demonstrated cognitive processing. Asynchronous forums allow time for reflection by students (Hara, Bonk, & Anjeli, 2000; McLoughlin & Mynard, 2009; Meyer, 2003; Wang, Woo, & Zhao, 2009) and for the creation of a written record, constituting an effective mechanism for students to build on each other's ideas (McLoughlin & Mynard, 2009). Through this level of engagement within the forums, students demonstrate higher levels of critical thinking and knowledge co-construction when compared to synchronous face-to-face discussions (Beckmann & Weber, 2016). The value of asynchronous discussion forums in the development of cognitive presence is evident, as higher cognitive participation in asynchronous discussions leads to the higher construction of knowledge (Webb, Jones, Barker, & van Schaik, 2004). To develop cognitive presence, a student needs to exercise critical thinking skills by communicating his or her thinking to others (Akyol & Garrison, 2011a). One should look at cognitive presence as a process (requiring sustained and continuous communications) that fosters the development of higher-order thinking skills instead of individual learning outcomes (Akyol et al., 2009; Akyol & Garrison, 2011b).

Forums by themselves will not likely support greater cognitive presence, but they can be structured to do so. Garrison (2007) explains that cognitive presence is operationalized through Dewey's (1933) practical inquiry model on reflective thinking and has four phases: triggering event, exploration, integration, and resolution (further elaborated upon in Chapter 2). The more a forum is structured to encourage these four phases, or the more a certain course structure (instructor-paced versus self-paced) allows for the presence of these phases (e.g., more likely to come to resolution in the instructor-paced condition by working synchronously with others through an issue), the more likely it may be to support greater cognitive presence. Whether the instructor-paced condition does in fact elicit more cognitive processing compared to a self-paced condition is considered in this study, with findings appearing in Chapter 4. Since student-student forum interactions are likely crucial for the demonstration of cognitive processing in a MOOC environment where instructor-student contact and instructor facilitation is limited, this dissertation examines different pacing conditions (instructor-paced and self-paced) to determine which structure best fosters cognitive processing in those forums.

Statement of the Problem

It is unclear what effects the pacing condition (self-paced versus instructor-paced) of online courses generally and MOOCs specifically has on discussion forum activity and learning in those forums as measured by cognitive processing. The structure of a course may impact how students engage in discussion forums, and an individual student's engagement within a forum may impact cognitive processing. Through an analysis of the discussion forums, I further the understanding of the levels of student participation within online and open learning environments such as MOOCs (Chiu & Hew, 2017). I examine student use of words in forums across the two MOOC pacing conditions in this study to report on how engagement and cognitive processing vary between conditions and to account for relationships between forum activity and cognitive processing within each condition. This study investigates three MOOCs offered by Harvard University: The Ancient Greek Hero in 24 Hours, Super-Earths and Life, and Visualizing Japan (1850s–1930s: Westernization, Protest, Modernity). Each MOOC was offered in two iterations: the first featured a weekly release of learning units, referred to in this study as an *instructor-paced* MOOC, and the second had a *self-paced* structure that allowed a student to move through the course at his or her own pace. In both iterations, the discussion forums played a vital role. This study compared the two iterations in terms of forum engagement and learning, as measured by cognitive processing (outcome measure from LIWC).

I use my data to test the research of Campbell, Gibbs, Najafi, and Severinski (2014), who found that students in self-paced MOOCs were still actively engaged in discussion forums despite not having a defined cohort. These findings contradict Sharif and Magrill (2015), who suggest that the opposite would occur—students would feel isolated in the self-paced MOOC and, therefore, not engage in knowledge construction.

In a more traditional online course, class size may be between twenty and thirty students and the discussion forums for the course are simple to navigate and contribute to. By contrast, the sheer volume of posts from the thousands of students enrolled in a MOOC can be overwhelming for many students. This can create problems for student learning, as it is difficult to find relevant posts and engage in the type of knowledge construction that is essential for learning. The overwhelming nature of the discussion forum can scare students away from being actively engaged in the forum, thus limiting their ability to gain knowledge and experiences from interactions with their peers. Additionally, it is not feasible for an instructor to respond and give feedback to each student's postings, and, therefore, it is essential to have an effective course structure that maximizes student-student interaction in forums. Student-student online discussions with a high level of cognitive presence might be better achieved by a particular course pacing structure that could help to compensate for the well-documented limits in MOOC teaching presence (Gašević, Adesope, Joksimović, & Kovanović, 2015). This study explored different MOOC pacing conditions (instructor-paced versus self-paced) to determine if one influences forum engagement and the development of cognitive processing more than the other.

Purpose of the Study

DeBoer, Ho, Stump, and Breslow (2014) suggest that researchers should focus on specific research variables in studying MOOCs, such as level of participation in activities, achievement, curriculum sequence, and time. These variables leverage capabilities of MOOCs to capture digital log files of participant interactions to further understand how technologies are utilized for learning and how learners interact with MOOC technologies. There is a need to study the relationships between the design aspects of MOOCs and student learning in more complex designs than single-course case studies (Veletsianos & Shepherdson, 2015; Wiebe, Thompson, & Behrend, 2015). Therefore, this exploratory case study utilizes an embedded multiple-case design to examine what influence the pacing condition of a MOOC had on the forum activity and engagement and on the demonstration of cognitive processing for learners.

The data for this study came from the discussion forums of three Harvard MOOC courses, each offered in both a self-paced and instructor-paced version. The cases are paired discussion forums from the instructor-paced and self-paced versions of each MOOC, and the embedded units of analysis are individual students' discussion forum posts within the given version of the MOOC, analyzed through automated text analysis. The courses selected are from three different subject areas—humanities (The Ancient Greek Hero in 24 Hours), art and culture (Visualizing Japan (1850s–1930s: Westernization, Protest, Modernity)), and science (Super-Earths and Life) —and are explained in more detail in Chapter 3. While each of these courses has been offered in both an instructor-paced and self-paced structure, the content and activities remained the same between both iterations.

The study uses qualitative data (in the form of discussion forum posts) and student clickstream data (such as amount of time spent in the discussion forum, number of replies and posts) to explore the similarities and differences between pacing conditions and their respective influences on student engagement. I examine the differences, in terms of both quality and quantity, in the discussion forums within each iteration of a given course (e.g., instructor-paced Ancient Greek versus self-paced Ancient Greek) to control for factors such as differences in course structure and subject area that may exist between the other courses in the study (e.g., Ancient Greek versus Visualizing Japan). My focus is on the student posts within each course discussion forum. My intent is to determine whether there is a statistically significant difference in the demonstration of cognitive processing between the instructor-paced and the self-paced course participants. From this analysis, recommendations can inform future iterations of MOOC pacing conditions. These recommendations are pertinent for other researchers, instructional designers, and MOOC instructors. Further in-depth discussion of implications and procedures occurs in Chapter 3 and Chapter 5.

One of the challenges in studying MOOCs is large dataset size. This challenge may be mitigated, however, through the leveraging of the automated text analysis tool Linguistic Inquiry Word Count (LIWC; pronounced “Luke”) (Tausczik & Pennebaker, 2010). My unit of analysis is each individual student forum posting, and with LIWC I analyze each post and calculate the percentage of words within it that fall under the “cognitive processing” category of the text analysis tool. This tool also includes categories on social words, perpetual processes, and core drives and needs (Kahn, Tobin, Massey, & Anderson, 2007; Khazaei, Lu,

& Mercer, 2017; Tausczik & Pennebaker, 2010). Simms et al. (2017) outline the process that LIWC uses to calculate the percentages for the respective categories. For example, if a given discussion post has 83 words, and 10 of these belong to the cognitive processing category, LIWC gives the post a cognitive processing score of 8.3 (10/83). This category, which is made up of 797 words, encompasses words such as “cause”, “know”, and “ought.” To offer a more refined look at the cognitive processing category, I also consider the subcategories of *insight* (e.g., “think”, “know”, “consider”); *causation* (e.g., “because”, “effect”, “hence”); *discrepancies* (e.g., “should”, “would”, “could”); *tentativeness* (e.g., “maybe”, “perhaps”, “guess”); *certainty* (e.g., “always”, “never”); and *differentiation* (e.g., “but”, “with”, “without”). For this study, LIWC provided the number of words per sentence in a given post, the word count, the number of six-letter words used in the post, and a score for “analytical thinking.” Analytical thinking is one of four global high-level text properties in LIWC (Simms et al., 2017) and is one of the summary variables output by LIWC. This summary variable measures the degree to which a post contains words that demonstrate formal, logical, and hierarchical thinking patterns—something that is a component of cognitive presence (Tausczik & Pennebaker, 2010). The summary variable of analytical thinking is one of two non-transparent dimensions, in that it is derived from algorithms (based on previous research) reflecting a particular balance of word usage across a number of categories (Faasse, Chatman, & Martin, 2016; Pennebaker, Boyd, Jordan, & Blackburn, 2015).

Cognitive presence is developed through interaction between participants. I used descriptors from each discussion post, specifically, the position of a given post in a thread, to determine what effect they had on cognitive processing and also looked at whether the fact that the post was a reply to an instructor or to a peer had any such effect. These variables are explained in Chapter 3 and findings based upon an examination of them are reported in Chapter 4.

Using the automated text analysis tool LIWC in this study allows the large dataset under review to be parsed and analyzed for these thinking skills in a more efficient and consistent way than would be achieved through the manual coding of each of the discussion forum posts by hand. By having a consistent means by which to analyze the discussion forum posts from the two course structures, similarities and differences between their respective scores can be closely analyzed to extract meaning that can inform MOOC course design decisions.

Significance of the Study

In current literature, discussion forums have been used to measure attrition, student engagement, and communication patterns, as well to predict course completion (see, e.g., Eynon, Hjorth, Yasseri, & Gillani, 2016; Gillani & Eynon, 2014; Jordan, 2015; Wang, Wen, & Rosé, 2016; Wen, Yang, & Rosé, 2014). Cognitive presence is the result of sustained communication such as that found in forums, and it requires both student-student and student-content interaction (Kanuka & Garrison, 2004). Through my use of text analysis and clickstream data, I have extended the existing research to examine the impact that course structure has on learner behaviors within a MOOC. My research not only contributes to the literature on the development of cognitive presence in MOOC courses, it also explores two distinct course structures—instructor-paced and self-paced—and their influence on this development. Further, analyzing multiple versions of the same course addressed the literature gaps on differences in learner behaviors and on levels of engagement within the same MOOC course (Gallagher & Savage, 2016).

This dissertation presents a fusion of the concepts of learner-centered instruction and structure of courses, and of the resulting impact that these concepts have on cognitive presence. This work contributes to the empirical research on delivering effective online instruction. This study explores the influence of course structure on forum activity and the demonstration of cognitive presence, which I am reporting through the measure of cognitive processing. While there has been quite a bit of research on MOOCs and discussion forums (discussed in Chapter 2), my research is unique in that it studies (1) forum activity in multiple courses with differing pacing conditions but with the same content and (2) student learning as measured by cognitive processing.

The nature of MOOCs represents both a challenge and a benefit for educational research purposes. For example, large enrollments result in extensive datasets and multiple data points, such as student log files of interactions, discussion forum posts, and activity completion scores. This vast quantity of data provides for a unique opportunity to explore the connections between student engagement, learning goals, and outcomes (Henrie, Halverson, & Graham, 2015). As Touati (2016) points out, MOOC instructors face the challenge of creating learning environments that will allow for personalized interactions in situations where there are far more students than available instructors. Thus, the personalized interactions need to occur at the student level, and understanding the influence of the MOOC structure on the creation of these types of opportunities for interactions is of critical importance for studying and improving MOOCs. In Chapter 2, I explain how the three domains of the CoI model inform the design of discussion forums and the locations within online courses where student-student interactions most often occur. In Chapter 3, I explain how these discussion forums can be efficiently analyzed using quantitative text analysis.

Overview of Approach and Research Questions

This research study employs an embedded multiple-case design in an exploratory case-study approach (Yin, 2014). This approach was selected to allow for the investigation of a specific phenomenon, namely, forum activity and learning (Gallagher & Savage, 2016; Yin, 2014). Due to the nascent nature of MOOCs and the lack of extensive empirical research on MOOC discussion forum structures, MOOCs are particularly well-suited for case-study research; they present the possibility of multiple new research areas in need of exploration and analysis (Gallagher & Savage, 2016). The research design explained in detail in Chapter 3 uses automatic text analysis to quantify the qualitative data, and from there utilizes descriptive statistics and regression models to explore the influence of the structure of a course on levels of cognitive processing within the course's discussion forum.

This study explores the similarities and differences of forum engagement and cognitive presence between instructor-paced and self-paced MOOCs. The research provides insight into learner behaviors and a deeper understanding of how course structure impacts the development of cognitive presence for course participants. The methodology explained in Chapter 3 addresses the following research questions:

1. How does cognitive processing vary between pacing conditions?
2. In what ways does pacing condition affect discussion forum activity and engagement (e.g., number of replies, number of words used)?
3. What aspects of forum activity are predictive of cognitive processing and do these relationships exist in both pacing conditions?

Definition of Terms

Cognitive Presence

The third domain of the CoI model (the others being teaching presence and social presence), cognitive presence is the central element of the *community of inquiry* (discussed below). Cognitive presence is focused on the processes of higher-order thinking, describing the learning phases from initial practical inquiry to eventual knowledge construction and problem-solving (Garrison et al., 1999; 2001; Joksimovic, Gasevic, Kovanovic, Adesope, & Hatala, 2014). Cognitive presence is the result of interaction between students, content, and instructor in a way that allows students to develop and demonstrate their critical thinking skills, which are primarily meta-cognitive in nature and require communicating one's thinking to others (Akyol & Garrison, 2011a) through a collaborative process of constructing meaningful knowledge (McLoughlin & Mynard, 2009). Typically, cognitive presence is measured through hand-coding of the discussion forums or text using the CoI framework.

Cognitive Processing

To operationalize the measurement of cognitive presence, I utilized the "cognitive processing" category within the Linguistic Inquiry Word Count (LIWC) tool. One limitation of LIWC is that it looks at individual words and not at combinations of words or at context. However, due to the sheer volume of posts in this study, I believe that this tool is the most useful one for extracting meaning from the discussion posts. LIWC's cognitive processing dictionary has 797 words that are used to calculate the level of cognitive process in each discussion post. Examples of words in this dictionary include "cause", "know", and "ought."

Community of Inquiry

The theoretical framework for this study, the community of inquiry features three interconnected presences: teaching, social, and cognitive. Together, these presences form the educational experience for the learner.

Critical Thinking

Beckmann and Weber (2016) define critical thinking as “a state of thinking that is beyond one’s own, or even groups’, interests and is dependent on the quality of standards and depth of experience the thinker has in respect to a particular problem or question” (p. 57). Lai (2011) relates critical thinking skills to student learning outcomes such as metacognition, motivation, collaboration, and creativity. In the context of this study, critical thinking will mean the way cognitive presence is demonstrated.

Dictionary

Tausczik and Pennebaker (2010) explain that LIWC uses *dictionaries* to categorize words and that the term “dictionary” here refers to the “collection of words that define a particular category” (p. 27).

Instructor-Paced MOOC

In an instructor-paced, or live, MOOC, there is a specific start and end date and students are often able to earn a certificate of completion. In the literature, this type of MOOC is also referred to as *session-based* (primarily by courses offered by Coursera); as having a *weekly release structure*; or as a *cohort model*, in that learners will be moving through the course together. These courses follow an instructor-led model, much like traditional on-campus courses. For simplicity, I use the term “instructor-paced” to encapsulate all of these variations.

Linguistic Inquiry Word Count (LIWC)

This study utilized the Linguistic Inquiry Word Count (LIWC) (pronounced “Luke”) to analyze the unit of analysis, the MOOC discussion forum posts. This validated tool uses dictionaries to categorize and quantify language used in text and provides a calculation of the percentage of words within defined categories (Kahn et al., 2007; Khazaei et al., 2017; Simms et al., 2017; Tausczik & Pennebaker, 2010).

Massive Open Online Courses (MOOCs)

Massive open online courses (MOOCs) provide the context for this study. MOOCs feature enrollments ranging from 200 to tens of thousands of learners. The two most common types of MOOCs are cMOOCs and xMOOCs. A cMOOC uses connectivism as its guiding learning theory and focuses on networked learning between and among individuals; an xMOOC follows the more traditional course delivery structure that features modules, lessons, and pre-determined activities. This study focuses specifically on xMOOCs.

Pacing Condition

In this study, the “pacing condition” refers to whether the course was offered in a self-paced or instructor-paced format. Pacing condition can also be referred to as “course structure.”

Self-Paced MOOC

A self-paced MOOC, also referred to as an on-demand or, simply, a MOOC, does not have a specific start period but does, for logistical reasons, have a defined end period. In these MOOCs, learners can move through the course at their own pace and, while they often have access to the same course materials as students in an instructor-paced MOOC, they may not have the ability to earn a certificate of completion.

Organization of the Study

This chapter presented a background and a rationale for the study. Chapter two provides a review of literature examining the connection between cognitive presence and online discussion forums. Details about the study's theoretical framework are interspersed throughout the chapter, along with connections to the literature as it relates to the research on MOOCs. Chapter three discusses study methodology in detail, specifically, how quantitative content analysis (QCA) is used in this study. It also scrutinizes the limitations and constraints that may affect findings and implications. Chapter four presents the study's findings. Chapter five provides a discussion of those findings and how they relate to the Community of Inquiry (CoI) framework and addresses the research question. It also contains the conclusion and a discussion of future research directions.

CHAPTER TWO: LITERATURE REVIEW

In this literature review, I provide an overview of the research on massive online courses (MOOCs), including discussions on MOOC dimensions, structure, research directions, and challenges. I then address extant research on MOOC discussion forums. I build off of the discussion in Chapter one of the theoretical framework of CoI and its relationship to online learning environments and conclude with how cognitive presence and discussion forums are linked.

Organizational Structure

For my literature review, I use an organizational structure (Figure 1) to discuss and analyze the literature around the three main components of my study: the pacing condition, the discussion forum activity and the characteristics of the forum posts, and then cognitive processing and its development in MOOC discussion forums. Starting with the pacing condition, I offer an overview of research on MOOCs, including the literature on the elements and dimensions of MOOCs, as well as on MOOC discussion forums. I then move into a discussion of the different MOOC course structures I examine in this study. For the discussion forum activity and post characteristics sections, I explain the role of the CoI model as it relates to engagement in distance education environments. This discussion leads into one on the connection between cognitive presence and discussion forums.

Figure 1. Organizational Structure of literature review

Overview of MOOCs

While many researchers are accustomed to class sizes ranging from 25 to 100 students, massive open online courses (MOOCs) have class enrollments in the thousands and even tens of thousands and thus present a tremendous opportunity to gain a greater understanding of the type of learning occurring within online environments. The principles of critical thinking development that function in traditional face-to-face classes do not always transfer into online learning environments (Kanuka & Garrison, 2004). However, an online learning environment offers new opportunities for students to reach deeper levels of learning, much of which will occur in discussions forums (Shea & Bidjerano, 2009). While there is consensus among researchers that evidence of cognitive presence can be found and nurtured within online discussion forums (see, e.g., McLoughlin & Mynard, 2009; Webb et al., 2004), little is known about the specific structures that can create meaningful discourse (Gilbert & Dabbagh, 2005) or the impact that these forums have on learning (Bergner, Kerr, & Pritchard, 2015). These online discussion forums potentially contain a wealth of information about the types of learning that occur in MOOCs, but analysis of the data can present a challenge, owing to the sheer quantity of it. Thus, much remains unknown about the effectiveness of MOOC discussion forums. This study seeks to close this gap through an analysis of discussion forums in both instructor-paced and self-paced MOOCs to determine if these pacing structures influence the development of cognitive presence.

Elements of a MOOC

Nawrot and Doucet (2014) offer the following as goals of MOOCs: (1) provide open access to high-quality education to learners on a global scale, (2) leverage social interactions to support new opportunities for knowledge creation, and (3) advance research on learning. Stevens (2013) summarizes Cormier's essential elements of a MOOC as follows: (1) it is a course with a critical mass of participants, allowing knowledge to originate from participants instead of from just an instructor; (2) it is free of charge, has an open and accessible syllabus, and is flexible, permitting students to guide their own learning; (3) it is online; and (4) it has structure (p. 2). The three main MOOC pedagogies are cognitive-behaviorist, social constructivist, and connectivist (Anderson & Dron, 2011). Cormier’s four MOOC components align well with the “traditional” MOOC definition offered by Pence (2012). When compared to an introductory science class held on a university campus, a MOOC presents more differences than similarities. The onsite science class could have multiple sections, each exceeding 500 undergraduate students. While such a class could be considered “massive,” it lacks the other necessary MOOC criteria, e.g., offered fully online, flexible enough to let participants practice self-guided learning. Today, with Internet access, anyone can become a student. This diversity of learners—each with varying needs, previous educational backgrounds, and the desire to dictate the terms of their learning—is another element that sets MOOCs apart from traditional face-to-face courses (and even from other online courses).

Typology and Dimensions

There is no all-encompassing definition of a MOOC in the literature. Instead, the term is used to refer to a variety of offerings (Major & Blackmon, 2016). The two most common types are cMOOC and xMOOC. Major and Blackmon (2016) differentiate between the two by focusing on the role of the learner, which, in turn, results in significantly different learner experiences. In a cMOOC, or connectivist MOOC, the expectation is for the learner to not only participate in discussions, but to take an active role in the production, discovery, and debate about course content. A distinguishing characteristic of a cMOOC is the use of various external tools, such as Twitter or blogs, produced by students outside of the course site but brought into the course space by the students. This aligns with connectivist pedagogy that views learning as learner-centered and influenced by openness, autonomy, and connections across a network of distributed knowledge (Major & Blackmon, 2016; Siemens, 2005). In contrast, xMOOCs are more teacher-driven and use a behaviorist view of learning. These tend to be offered by institutions and follow a pedagogical approach found in more traditional online courses, using a linear course structure with defined sequential units. Recently, however, there has been an evolution in xMOOCs, whereby many now incorporate characteristics found in cMOOCs (Major & Blackmon, 2016).

Major and Blackmon (2016) offer a useful typology for MOOCs that focuses on the dimensions of "(1) affiliation, (2) size, (3) accessibility, (4) duration, (5) timing, (6) relation to knowledge, (7) content, (8) structure, (9) authority and control, (10) participant and (11) pedagogy" (p. 21). The courses examined in this study, discussed in more detail in Chapter 3, were offered both in an instructor-paced and a self-paced structure. These courses were specifically selected because they permit a comparison of discussion forum activity in different structures while allowing for a control of course content and instructor influence. Application of the dimensions from Major and Blackmon (2016) results in the two structures having slightly different dimension aspects. While both offerings would be considered massive, the structure of the instructor-paced, or time-board, course would be regarded as linear, while the self-paced course would be regarded as adaptive (since learners can move around the entire course without specific time structures). Also, authority and control differ in the instructor-paced versus the self-paced offering, in that the latter would be more learner-centered.

Research Directions and Challenges

While the availability and flexibility of MOOCs make them attractive learning options, these same factors also pose obstacles. One notable challenge is that MOOCs have less-than-full acceptance within the existing landscape of higher education. There are several reasons for this, some of the most pressing being the lack of college credit offered and the inability to verify who completed the coursework. Despite the fact that several third-party providers, such as Coursera and edX, offer certificates of completion, they are not currently accepted by higher education institutions as credit-hour replacements or supplements (Gwynne, 2013). This deficit of credit provides little incentive for learners to stay engaged in courses and contributes to the incredibly high attrition rate of MOOCs. Also, since cMOOCs are nonlinear, meaning that students determine their learning paths rather than following one set path, MOOCs can be overwhelming for individuals who are more accustomed to structured learning environments.

Liyanagunawardena, Adams, and Williams (2013), and Veletsianos and Shepherdson (2016) worked on comprehensive systematic reviews of MOOC research conducted between

2008 and 2012, and 2013 and 2015, respectively. Forty-five works published between 2008 and 2012 were studied; a majority of them were journal articles and conference publications, and more than half (twenty-six) were published in 2012 (Liyanagunawardena et al., 2013). The majority of the papers were case studies; papers focusing on educational theory were second in prevalence. Most publications examined the learner perspective, which leaves a gap in the literature on the perspective of the facilitator (Liyanagunawardena et al., 2013). The review by Veletsianos and Shepherdson (2016) investigated 183 articles published between 2013 and 2015, with the majority (129) being published in 2014, substantiating the significant upward trend in publishing MOOC research that Liyanagunawardena et al. also observed. Five research strands were identified—student-focused, design-focused, context and impact, instructor-focused, and "other"—and more than 80 percent of the articles were student-focused, while less than 10 percent were instructor-focused (Veletsianos & Shepherdson, 2016).

Other challenges inherent in MOOC research include the large number of students involved, the new ways of learning due to the freedom of choice by students, and the copious amounts of data produced (Raffaghelli, Cucchiara, & Persico, 2015). However, these challenges create new research opportunities that do not exist and are not possible in traditional online learning environments. With data that potentially spans millions of users, researchers can better identify connections between engagement, goals, and outcomes within MOOCs and other online courses (Henrie et al., 2015). Along these lines, Breslow et al. (2013) encourage researchers to take advantage of the huge quantity of data MOOCs generate to identify in greater detail what contributes to and constrains students' learning. Baer (2003) pinpoints the factors of accounting for personal differences in learning goals and ambitions, learner backgrounds, and the pace of learning as being critical for success within MOOCs. And while retention is a concern, and many studies have focused on ways to predict attrition, Littlejohn, Hood, Milligan, and Mustain (2016) suggest that research should consider the goals of learners and their motivations.

Finally, MOOC instructors' inability to establish one-to-one interactions and to provide quality feedback to all participants is another challenge (Touati, 2016). The primary spaces used for student interactions in MOOCs are discussion forums, as they provide for the exchange and sharing of ideas among learners (Akcaoglu & Lee, 2016; Sharif & Magrill, 2015; Wong, Pursel, Divinsky, & Jansen, 2016). Thus, it becomes critical for instructors to understand how to use these forums to spawn the types of social interactions that form the connections between learners, connections that encourage them to persist in the course.

MOOC Discussion Forums

MOOC research tends to emphasize forum activity, given that this is the primary source of student engagement. Sharif and Magrill (2015) suggest that the experiential learning that can occur in a MOOC discussion forum aligns well with the connectivist approach of cMOOCs. Most of the research on forums is done from a macro perspective (Sun, Li, & Lin, 2016), and findings suggest that the structure of the discussion significantly impacts its quality (Hewitt, 2003; Swan, Shea, Fredericksen, Pickett, & Pelz, 2000; Vonderwell, 2003; Vrasidas & McIsaac, 1999). But despite this potential, the precise impact of discussion forums on learning is not yet fully understood (Bergner et al., 2015), making further exploration of discussion forum impact critically important. Since it is within the forums that the main interaction between learners occurs, it makes sense that forums have been one of the most studied areas in MOOC research. A variety of methods have been used to look at discussion forum activity, from collecting descriptive statistics of the number of posts to leveraging social network analysis to explore the subcommunities of students that have developed and their interaction patterns. From this analysis of the activity in the discussion forums, researchers have been able to describe learner behavior in MOOCs by focusing on the areas of attrition (see, e.g., Jordan, 2015; Onah et al., 2014), engagement (see, e.g., Wang, Yang, Wen, Koedinger, & Rosé, 2015), and communication patterns (see, e.g., Eynon et al., 2016; Gillani & Eynon, 2014).

MOOC Pacing Conditions

There are two primary ways for a MOOC to be delivered: in an instructor-paced or in a self-paced pacing condition. Research suggests that these conditions may impact forum engagement (see, e.g., Campbell et al., 2014; Sharif & Magrill, 2015). Campbell et al. compared learner intent in instructor-paced and self-paced MOOCs and found similar interaction patterns when comparing the two groups. Their most significant finding for purposes of my study is the observation of extensive discussion forum use in the self-paced course. Sharif and Magrill offer a counter perspective, suggesting that self-paced courses would not be socially welcome enough to encourage active participation and linking learner engagement to whether the learner persists or resigns from the course. To explore this in greater depth, this study compares discussion forum posts in instructor-paced versus self-paced MOOCs. In the next section, I describe how interaction and engagement occurs, and is different, within the two course pacing conditions of my study.

Instructor-Paced Courses

Many consider the instructor-paced structure, where there is a linear and sequential course structure dictated by an instructor, to be the traditional structure for instruction.

Courses with this structure have a defined start and end period, are of a specified duration, and are based on the cohort model, which Sharif and Magrill (2015) suggest significantly influences students' decisions to either resign from or persist in a MOOC. The instructor's role is to determine how learners will move through the course. Joksimovic et al.(2014) maintain that the interaction occurring among the students is influenced by the teaching presence (in the form of the instructional design of the course). In instructor-paced courses, it is the teaching presence that brings the social and cognitive presences together and takes learners' needs and capabilities into account (Garrison & Anderson, 2003; Kreijns et al., 2014). Therefore, it is important to consider how the structure of a course is impacting and fostering engagement and interaction between students (Garrison, Cleveland-Innes, & Fung, 2010).

Given the fact that student engagement has been shown to be linked to all three CoI presences, most notably cognitive presence, this is a topic of interest for my study. Jordan (2015), citing the research study conducted by Perna et al. (2014), explains that her study "corroborates the findings of [Perna et al., 2014] in that the initial weeks of courses are key for student engagement" (p. 353). This is relevant for my study because students in an instructor-paced course will begin at the same time, so each will have a similar week-to-week schedule. This is not the case in self-paced courses, which could impact student engagement in such courses. Wise, Hausknecht, and Zhao (2014) further examined the concept of student engagement in their analysis of online discussions of an instructor-paced MOOC. They proposed a conceptual model of what they term online "listening," designed to show that engaging in a discussion forum is more active than the terms "reading" or "lurking," which are typically used to describe learner behavior in online discussion forums.

The connection between student activity and engagement levels in MOOC discussion forums has been studied primarily in instructor-paced courses. One such study was conducted by Hecking, Chounta, and Hoppe (2015), who established a correlation between the amount of posts and number of threads created by a given student and the total number of votes, which appear and function similar to a Facebook “like”, received by that student. Students who were active within the forums were viewed as influencers within the course, and they received more votes and replies than their less-active peers. I examined this same relationship by looking at the number of replies to, and the position of posts by, a given student within a thread to track engagement. I examined my dataset to determine if there was a correlation in the instructor-paced courses between higher levels of cognitive presence in threads that had more replies and the role of originator of the post being replied to, as this would be indicative of ongoing discussions and conversations between students and test the findings of Hecking et al. The study by Wang et al. (2016) honed in on engagement and found that higher learning gains were demonstrated by students who were not only active in the forums, but who were making on-topic posts. I expected to see such higher learning gains in instructor-paced courses after analyzing the position of posts within a thread. If there were high levels of cognitive presence in posts that appeared later in a discussion thread, that would suggest that the post was on-topic. As students spend more time engaging with content through the discussion forums, I would expect to see a higher cognitive presence score. In my study, I expected to determine whether student engagement, as measured by number of replies and posts in forums and number of words and sentences in posts, was greater in the instructor-paced condition, a result that might be expected based on the aforementioned research into instructor-paced courses. Through this analysis, I was interested in exploring the impact of the instructor-paced course pacing condition on the demonstration of cognitive presence.

My area of interest is cognitive presence, a phenomenon that requires sustained communications. To measure this within my courses, I looked not only at what was said within the forum posts, but also at where a given post fell within a given thread. Creating an environment where students are interacting with each other is important (and can be a challenge, particularly with MOOCs). But interactions within forums is not enough; students must also engage in meaningful discussions over a period of time to demonstrate cognitive presence.

Yang et al. (2014) compiled discussion forum data from three Coursera MOOCs offered in 2013 to address the concepts of engagement and interaction. Their analysis identified a two-way influence found in discussion forums. As users converse in a specific thread, they influence the conversation, both among each other and within the subcommunity as a whole. Students continuing to engage and influence each other's posts influences the direction and formation of the associated subcommunities. It is for these reasons that one would expect to see higher levels of cognitive presence in instructor-paced course discussion forums, as the students there are engaging with each other and the content at the same time. It is also expected that since students move through the content at the same time in an instructor-paced course, replies and interactions will occur at similar times. Since the replies will be sent close in time to when the original post was made, there are more opportunities for conversations. Thus, I would expect to see longer threads in an instructor-paced course. This presents an advantage for the development of cognitive presence in instructor-paced courses, since there is more engagement, both between students and also between students and the content, through these longer threads. As noted by Yang et al., "subcommunity structure arises both from the pattern of connections embedded within the network constructed from the threaded reply structure and the behaviors reflected through the text contributed within that structure" (p. 219).

There are two interconnected challenges to keeping students engaged in MOOC courses and fostering cognitive presence. The first relates to unimpressive participation rates within discussion forums. Many studies have found that participation rates in MOOC discussion forums can be on the low side (see, e.g., Breslow et al., 2013; Onah et al., 2014; Sharif & Magrill, 2015). The second challenge recognizes that if students are not active within discussion forums, their opportunities for engaging with each other are limited, and this, in turn, results in fewer possibilities for developing cognitive presence. Further complicating this research is the fact that even those students who are actively engaged in a forum at the outset tend to gradually decrease their activity levels as the value of the forum is perceived to fade over time (Yang, 2014). One reason for this is that as more students drop out of the course, there are fewer opportunities for interaction within the posts and, therefore, the posts have less perceived value. In an instructor-paced course, this drop-out factor and the declining participation rate in the discussion forum could negatively impact a student’s ability to engage with others as the course continues, resulting in the student seeing little value in continuing to post to fewer and fewer people. In the analysis of cognitive presence, this would show as lower levels of cognitive presence as the course continues, as well as fewer replies and shorter threads.

Self-Paced Courses

The self-paced, or on-demand, structure has a more adaptive learning path. In this structure, there are no definitive start and end periods for the course and the orientation to learning is student-centered. The learner can decide in what order he or she will view content. As such, student engagement and interactions will differ from those in instructor-paced courses.

In a self-paced course, students are not part of a cohort. The lack of a cohort has been identified as a factor that influences a student's decision about whether or not to continue with a course (Sharif & Magrill, 2015). While students in self-paced courses may not be as engaged with each other as their instructor-paced course counterparts, the self-paced MOOC structure can actually be more advantageous for self-directed learners if the content is of a high quality (Campbell et al., 2014). Campbell et al. suggest that self-directed learners have less reliance on instructors and, therefore, are likely to prefer the self-paced MOOC structure.

The type of student interactions will also differ between self-paced and instructor-paced MOOCs. In the self-paced course, there is likely to be less peer interactions than in the instructor-paced course, but this is not necessarily a bad thing. Campbell et al. (2014) discovered that even without cohort presence in self-paced MOOCs, extensive discussion forum usage nonetheless resulted.

The way students use a self-paced MOOC discussion forum impacts the way peers within the course influence each other. This influence will differ from that found in an instructor-paced course in that there will likely be more time between posts and replies. A student in a self-paced course may be reading a discussion post made three or four weeks prior and may choose to offer his or her own thoughts or insights. But the likelihood of the original/early people who posted to the forum returning to a given conversation and replying to new/later posts is low, and this minimal response to later posters' contributions might be demotivating for them (amounting, basically, to posting to dead air with no hope of reply). At the same time, by not focusing on specific deadlines, students in self-paced courses may benefit from reading more peer posts before making posts of their own. It is for this reason that I examined the amount of time students spend in forums to determine if there is any difference in how the forums are being used within the two course conditions.

Sharif and Magrill (2015) argue that self-paced MOOCs are too socially isolating and that the resulting disconnection damages the potential for "knowledge scaffolding." Scaffolding refers to the probing and follow-up questions that are posted in response to initial postings and which further the discussion (Whipp, 2003). This is a legitimate concern, as even in instructor-paced MOOCs, many students struggle to develop the social connections necessary to persist in the course. This can be exacerbated in a self-paced structure. Alternatively, Campbell et al. (2014) found that there was still extensive discussion forum usage in self-paced MOOCs despite the lack of a defined cohort presence. My study explores these contradictory findings to determine what influence, if any, exists between the course structure and the engagement and demonstration of cognitive presence in discussion forums.

The work by Yang et al. (2014) provides useful insight for my study. Specifically, the researchers' analysis of the way students need connections and need to develop subcommittees provides a context for me to explore the differences between the instructor-paced and self-paced courses. In the cohort model, it is assumed that students will be going through the course with a group and will be able to get the peer feedback and connections they need to feel engaged in discussion forums. While the instructor-paced students' level of cognitive presence could be expected to decline over time due to lack of discussion forum posts, it could actually be an advantage in the self-paced courses, as learners there will encounter forums already chock-full of posts, making them more likely to see value in the fourms and to engage. So, while the self-paced student may not post as frequently as his or her instructor-paced peer, the posts could still have high levels of cognitive presence, as the student was able to read, and benefitted from, the other posts in the thread. Also, following the argument asserted earlier that longer posts are likely to be more on-topic, a self-paced student who focused on longer threads would gain more opportunities to engage, and to engage more deeply, with the content, something that Hecking et al. (2015) suggested would result in higher learning gains. My study explored the influence of that assumption by looking at interactions in discussion forums in the self-paced versions of each course. By exploring the discussion forum activity in instructor-paced and self-paced courses, I was able to study relationships between engagement variables and cognitive presence.

Engagement in Distance Education Environments

Understanding that there are unique challenges in developing higher-order learning in online learning environments, Garrison et al. (2001) developed the “Community of Inquiry” model, or CoI. It consists of the three domains of social presence, teaching presence, and cognitive presence (Figure 2). This framework posits that to achieve high-quality online learning, learners need to have social interaction and collaboration with peers, need to be able to connect new knowledge to past experience, to must be able to immediately apply content to their current lives and experience a learning environment that is supportive of self-reflection and self-regulated learning (Cercone, 2008; Garrison, 2007; Ke, 2010; Shea & Bidjerano, 2009).

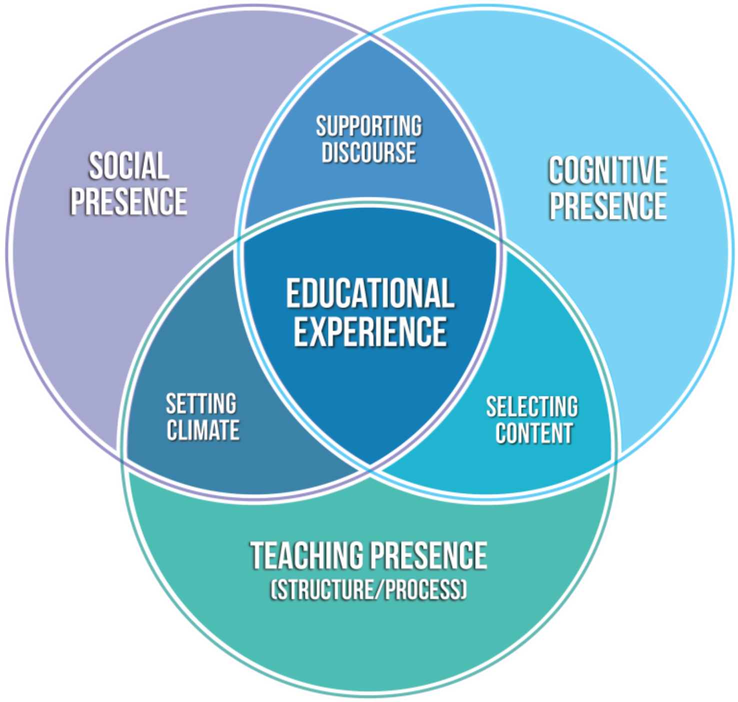

Figure 2. Community of Inquiry Framework. Adapted from Garrison et al. (2001)

The CoI model defines a community of inquiry as "a group of individuals who collaboratively engage in purposeful critical discourse and reflection to construct personal meaning and confirm mutual understanding" (Garrison, 2011, p. 2). In the context of this research, the community of inquiry was each course discussion forum. The CoI model uses a constructivist approach to learning and recognizes that interaction and discourse play key roles in higher-order learning, along with structure (design) and leadership (facilitation and direction) (Garrison, 2007). Further, the CoI model recognizes that collaboration, interaction, and community are vital parts of the learning experience (Lee, 2014).

Social Presence

The purpose of social presence is to create the conditions that will lead to inquiry and quality interaction. Social presence is focused on the human-human interaction aspect of online learning (Kim, Kwon, & Cho, 2011). The three core aspects of social presence are *effective communication*, *open communication*, and *group cohesion* (Garrison, 2007). In an online course, much of the human-human interaction occurs within discussion forums, and thus high-quality discussion forums require social presence (Akcaoglu & Lee, 2016). Moreover, while social presence is a requirement for high-quality forums, it can also explain lower feelings of satisfaction among students. As students respond less frequently to their peers' posts, there is less interaction, and thus a lower sense of social presence (Akcaoglu & Lee, 2016; Kreijns et al., 2014; Moore, 2016).

Relationships are therefore an essential part of this domain, and they can be difficult to cultivate in online learning environments. The transactional distance between online learners can make it a challenge for students to feel connected to each other, and this is only amplified in a MOOC, where you have a larger number of students, more diversity in previous experiences, and more variations in schedules. The development of a social relationship has been linked to the achievement of a greater sense of community and to higher levels of student satisfaction (Moore, 2014), as well as to higher levels of learning and attainment of learning objectives (Rourke et al., 1999), thus making relationship-building a crucial aspect within a community of inquiry. Further, social presence is the process through which students develop closeness to each other, which allows for the formation of a learning community (Kim, Kwon, & Cho, 2011, p. 1513). However, in forming relationships, it is imperative that learners foster connections that are more than personal and shift focus to academic purposes and activities within the online learning space (Garrison, 2007). As Garrison (2007) suggests, social presence is about quality over quantity. Supporting this, Kim et al. (2011) found that media integration, quality instruction, and interactivity are good predictors of social presence, while only media integration and quality instruction predict learning satisfaction.

Teaching Presence

The second CoI domain examines teaching presence and is focused on the role of instructors in course design, organization, and delivery and on instructions that guide social and cognitive presences to desired learning outcomes (Anderson, Rourke, Garrison, & Archer, 2001). The overall design of a course has a significant impact on the development of a community of inquiry. In fact, it has been found to have a greater impact than course duration (Lambert & Fisher, 2013). Teaching presence combines the social and cognitive presences and takes learners' needs and capabilities into account (Garrison & Anderson, 2003; Kreijns et al., 2014). It is a significant predictor of student satisfaction, perceived learning, and sense of community (Garrison, 2007). In a traditional face-to-face class, the teacher's presence is far different than in an online course, particularly in a MOOC. In online courses, students must play more active roles in their learning, specifically, within discussion forums. Joksimovic et al. (2014) explain that simply establishing interaction between students is not enough. they argue that, instead, interaction should be guided through a careful instructional design, i.e., teaching presence. In online environments, this is especially apparent in that the structure of the course and the opportunities for peer feedback and guidance become "intrinsic components of metacognition in a community of inquiry" (Kovanovic et al., 2015, p. 75).

Cognitive Presence

The third and final CoI domain, cognitive presence, is the central element of the community of inquiry and focuses on the process of higher-order thinking, describing learning phases from the initial practical inquiry to the eventual knowledge construction and problem-solving (Garrison et al., 1999, 2001; Joksimovic et al., 2014; Lee, 2014). To develop cognitive presence, students must exercise critical thinking skills, primarily meta-cognitive in nature and requiring the communication of one's thinking to others (Akyol & Garrison, 2011a) through a collaborative process of constructing essential knowledge (McLoughlin & Mynard, 2009; Shea & Bidjerano, 2009). Thus, one should view cognitive presence as a process (requiring sustained and continuous communications) that describes the development of higher-order thinking skills instead of individual learning outcomes (Akyol et al., 2009; Akyol & Garrison, 2011b). Kovanovic et al. (2015) identified cognitive presence as a measure of critical thinking because oral and text-based communications, such as those found in discussion forums, have been shown to stimulate the development of critical thinking skills. Shea and Bidjerano (2009) suggest that by focusing on pedagogical decisions that develop cognitive presence, the CoI can be a useful framework for online instructional design. Cognitive presence is defined in terms of a cycle operationalized through Dewey's (1933) practical inquiry model on reflective thinking, which has four phases (Figure 3): a triggering event, exploration, integration, and resolution (Garrison, 2007; Gibson, Ice, Mitchell, & Kupczynski, 2012). As Figure 3 shows, the four phases progress in a step-like way, with the highest phase being the resolution phase.

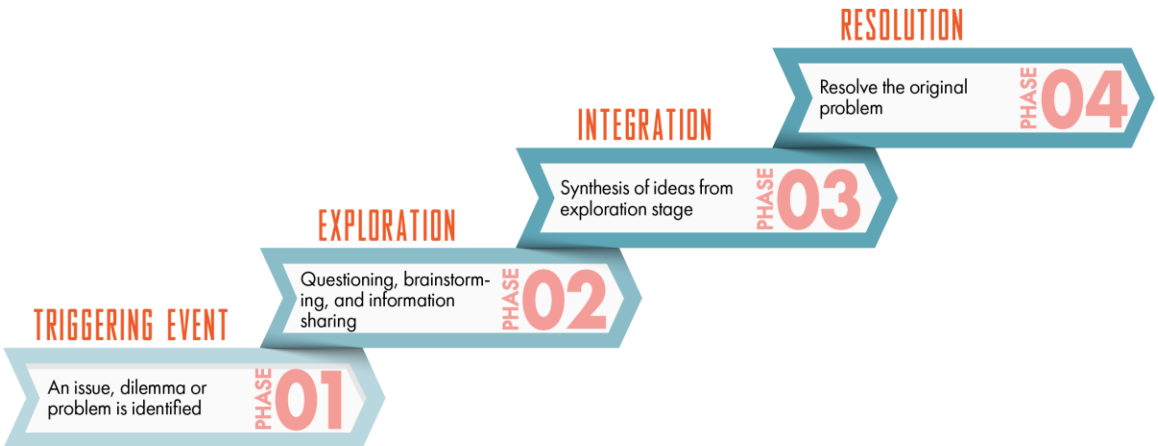

Figure 3. Four Phases of Cognitive Presence. Adapted from Garrison et al. (2001).

Triggering Event. In this initial phase, the learning cycle is initiated by a problem or dilemma, which, in the course context, is typically introduced by the instructor. In a discussion forum, this would be the initial prompt the instructor has posed to learners.

Exploration. In the second phase, students move to exploration, brainstorming, and other activities in which they gather information relevant to the problem or task at hand. For many discussion forums, this is the phase in which students spend the most time.

Integration. In this phase, after gathering an appropriate body of information, students synthesize and integrate different components of it, while being selective and filtering out all irrelevant information. It is at this stage where higher levels of cognitive presence are demonstrated.

Resolution. This final phase of cognitive presence is typically the most difficult to reach, in part due to the educational context. In this phase, students reach a settlement of the original problem. But if this is a new subject domain for the learners, they may not be able to attain this level within the relatively short duration of the discussion forum. It is also common to see the resolution of the original problem launch a new learning cycle, with an accompanying new triggering event (Kovanovic et al., 2015).

Interaction of Presences

The three CoI presences are interconnected and must all work together to create a high-quality learning environment, although the exact inter-relationships have not been clearly defined (Lee, 2014). Garrison and Anderson (2003) state that once social presence has been established in a course, cognitive presence can be enhanced and sustained. Additionally, Garrison, Cleveland-Innes, and Fung (2010) suggest that the central role of teaching presence is in the establishment and sustainment of social and cognitive presence. It is the specific course design and facilitation approaches (or teaching presence), such as chunking and sequencing of materials, used to develop and sustain cognitive and social presence for learners that creates the pathway to learning (Gibson et al., 2012; Kozan, 2016; Song & Yuan, 2015). Studies by Garrison and Cleveland-Innes (2005) and Akyol and Garrison (2011b) looked at learning approaches within communities of inquiry and found that specific forms of teaching presence—such as instructional leadership in facilitation, direct instruction, and appropriate course structure—have a positive effect on the promotion of learning, thus establishing and sustaining high levels of cognitive presence (Kovanovic et al., 2015).

Cognitive Presence and Discussion Forums

In this study, I examined the relationship between engagement in forums and the demonstration of cognitive processing. One reason online courses feature discussion forums is that these forums give students a place, and a way, to engage in conversations and interact with each other (Beckmann & Weber, 2016; Kent et al., 2016; Martin & Ndoye, 2016). It is through this interaction that social presence is created, and the ultimate result is cognitive presence. In this sense, online discussion forums operate in support of the constructivist theory of learning, in that they encourage learners to work collaboratively with peers (McLoughlin and Mynard, 2009). And while forums have the potential to allow for the co-construction of knowledge and for collaborative learning (Akyol & Garrison, 2011b; Lee, 2014; Shea & Bidjerano, 2009), many instructors shy away from using them for fear of lack of engagement and participation by students (Moore, 2016). In the MOOC context, there is minimal participation in the discussion forums (Onah et al., 2014), with some studies estimating the participation rate to be as low as 5 to 20 percent (Breslow et al., 2013; Hecking, Hoppe, & Harrer, 2015; Kizilcec, Schneider, Cohen, & McFarland, 2014; Sharif & Magrill, 2015). This low rate can be attributed to a number of factors, including students who enrol in a MOOC with no intention of completing the course, the lack of student interest in the topics in discussion forums, or students becoming overwhelmed by the number of forum posts and not feeling as if they can make meaningful contributions. This low forum participation level is concerning on several fronts, not the least of which is attrition, as research has demonstrated a positive connection between successful MOOC course completion and activity within forums (Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014; Sharif & Magrill, 2015). Therefore, it is imperative that instructors learn to effectively incorporate discussion forums into their courses.

Learning Is Collaborative

A successful online learning environment will provide opportunities for both collaboration and reflection for learners (Garrison et al., 2001; McLoughlin & Mynard, 2009). Kovanovic et al. (2015), citing the work by Lust et al. (2012) and Clarebout et al.

(2013), identify four keys for successful learning in current online learning environments: learners must be able to (1) recognize the tools or study tactics available in the learning environment; (2) make the connection between the available tools and specific assignments (e.g., using discussion forums); (3) obtain the meta-cognitive skills necessary to leverage said tools effectively; and (4) feel motivated enough to invest time and effort into using a tool and to meta-cognitively monitor and control the use of the tool in relation to their learning tasks. In addition to providing a space for collaboration, discussion forums can also provide a space for the co-construction of knowledge, as higher cognitive participation in asynchronous discussions has been shown to lead to higher construction of knowledge (Webb et al., 2004). Therefore, teaching presence, or instructional design, is critical and is connected to cognitive presence. This study investigated how students are co-constructing and collaborating in MOOC discussion forums by looking both at what they are saying (through the text analysis), and how they are engaging, in the forum (through the number of replies and position of specific posts within a thread). The analysis examined how these independent variables influenced the demonstration of cognitive presence.

Development of Higher-Order Thinking Skills

An analysis of discussion forums can provide new insight into the learning that occurs—or does not occur—within MOOCs. These forums may also reveal clues about the ways adult learners develop communities of practice and engage with each other in online learning environments. McLoughlin and Mynard (2009) found that several elements of discussion forums supported the development of higher-order thinking in students, such as time to reflect on one's writings and the creation of a written record that allows students to more effectively building upon each other's ideas. These researchers further suggest that the type of assigned task and the wording of the initial discussion forum prompt can affect the type of higher-order thinking processes that will emerge in online discussion. Breivik (2016) considers the use of online discussions to be a learning activity and explains that "the development of critical thinking is an important rationale for higher education and plays a central role, both as a goal for and as a prerequisite of successful online discussions" (p. 5).

Several researchers have tried to operationalize critical thinking in online contexts. Breivik cites research done by Weltzer-Ward (2011) noting that between 2002 and 2009, there were fifty-two different research frameworks and coding schemes examining online educational discussions. And while there were quite a few frameworks, DeWever, Schellens, Valcke, and Van Keer (2006) found that there was a lack of "coherence between the theoretical bases and operationalization for a number of the frameworks . . . examined" (as cited by Breivik, 2016, p. 2). This both highlights the importance of understanding critical thinking in online learning contexts and suggests that it is difficult to effectively and efficiently measure such thinking. Given the challenges in measuring critical thinking, it understandably becomes difficult to identify best practices for fostering and encouraging this type of deeper learning, particularly in online spaces.

Development of Cognitive Presence

In my study, I explore the factors and variables which may influence the development of cognitive presence. Others have studied cognitive presence in MOOC discussion forums (Beckmann & Weber, 2016; Dowell & Graesser, 2014; Kovanović et al., 2016; Kovanović, Gasevic, Hatala, & Siemens, 2017). Using automatic text analysis in such examinations can make the process more efficient (Dowell, Graesser, & Cai, 2015; Dowell & Graesser, 2014; Kovanović et al., 2016, 2017; Tausczik & Pennebaker, 2010).

The use of automatic text analysis tools is one method for examining discussion forums for evidence of cognitive presence (Dowell & Graesser, 2014; Kovanović et al., 2016). Other studies have coded for cognitive presence using the CoI framework; these have typically been hand-coded studies. The instrument employed in the CoI framework is tedious and difficult to code, resulting in hours of manual processing (Kovanovic et al., 2016). This method of analysis is far more feasible when the dataset in a study is smaller or when there is a team of researchers working on the study. However, it is not feasible for looking at MOOC discussion forums due to the sheer volume of posts that must be analyzed. One solution to this problem is to use a text analysis tool such as Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004; Graesser, McNamara, & Kulikowich, 2011; McNamara et al., 2014). The tool I selected for my study is the Linguistic Inquiry Word Count tool, or LIWC (Tausczik & Pennebaker, 2010).

Research Questions (RQs) and Hypotheses (Hs)

I developed a conceptual model (Figure 4) (1) to see what influence the course pacing condition (instructor-paced versus self-paced) has on MOOC discussion forum activity and post characteristics (e.g., number of replies, length of post, type of post) (2) to understand how this discussion forum activity impacts the demonstration of cognitive processing and (3) to explore what role the course pacing condition has on the demonstration of cognitive processing in MOOC discussion forums. The model explores the relationships between (1) my independent variable (IV), the pacing condition, (2) the mediators, which are the discussion forum activity and post characteristics measures, and (3) the dependent variables (DVs) generated by the LIWC analysis, which are cognitive processing, insight, differentiation, certainty, tentative, discrepancy, and causation. The discussion forum activity and post characteristic measures include type of post, whether the post was made in reply to an instructor or a peer post, whether it was endorsed by the instructor, where it appeared in the discussion thread, and how many comments and replies the post generated. To explore the substantive nature of the post, I look at variables such as words per sentence and word count, along with the dependent variables, all calculated by LIWC. LIWC produces a numerical value for each of the variables, e.g. word count, number of six letter words, which is appended to the row for each discussion post being analyzed. The theoretical framework uses cognitive presence, and my conceptual model explained this through cognitive processing. . The other six dependent variables are sub-scores for the cognitive processing score. In Figure 4, the solid lines represent direct relationships and dotted lines represent indirect relationships. I explain the research hypotheses and how I used this model in more detail in Chapter 3.

Figure 4. Conceptual Model

My research answers the following questions and addresses the associated hypotheses (Hs):

1. How does cognitive processing vary between pacing conditions?
    - H1: Pacing condition indirectly affects cognitive processing, such that students in the self-paced courses will have higher average cognitive processing scores than students in the instructor-paced courses.

2. In what ways does pacing condition affect discussion forum activity and engagement (e.g., number of replies, number of words used)?
    - H2: Pacing condition affects discussion forum activity and engagement, such that students in the instructor-paced courses will post more frequently and have more replies than students in the self-paced courses and students in the self-paced courses will make longer posts (word count and thread length) than students in the instructor-paced courses.
3. What aspects of forum activity are predictive of cognitive processing and do these relationships exist in both pacing conditions?
    - H3: Discussion forum activity and post characteristics affect cognitive processing, such that students in self-paced courses will have posts that have higher levels of insight, certainty, and differentiation than students in instructor-paced courses. Additionally, longer (in terms of word count) posts made as discussion posts will have higher levels of cognitive processing.

Hypothesis 1 (H1): Pacing Condition Indirectly Affects Cognitive Processing

To test this hypothesis, I used an independent samples t-test to compare the cognitive processing (and sub-parts) within each pacing condition. I expected to find that students in the self-paced courses would have higher average cognitive processing scores than students in the instructor-paced courses. This expectation is based on the work by Kanuka and Garrison (2004), who posit that cognitive presence requires sustained communication and student-student and student-content interaction. Because I expected students in the self-paced courses to have more time to read and consider their peer discussion posts, I expected to find students in the self-paced courses would show higher levels of cognitive presence than students in the instructor-paced courses I also expected to find that students in the self-paced courses would make longer posts and demonstrate higher levels of cognitive processing. Following this same logic, I also expected to find that the additional time available to students in self-paced courses to consider their peers' posts would result in longer threads than are found in instructor-paced courses. Despite the lower total number of students, I expected that there would still be activity within the self-paced discussion forums (Campbell et al., 2014) and that the additional time and consideration of the posts would lead to more opportunities for students to engage in knowledge construction with each other, which, in turn, would be reflected in higher levels of cognitive presence (Kent et al., 2016).

Hypothesis 2 (H2): Pacing Condition Affects Discussion Forum Activity and Engagement

For this hypothesis, I looked at the relationship that the course structure—instructor-paced versus self-paced—has on the discussion forum activity measures. I considered this it to be an important factor because, by understanding what students are doing within the discussion forums, we can start to understand how they are participating in online environments such as MOOCs (Chiu & Hew, 2017). I expected to find that students in the instructor-paced courses would post more frequently and have more replies than students in the self-paced courses, owing to the fact that students in the former category were moving through the course with a cohort, and thus there were more opportunities for interaction between peers. I also wanted to provide some information and analysis to more thoroughly describe discussion forum activity in these two pacing conditions, since much is still unknown about the effects of MOOC discussion forums upon student learning (Chiu & Hew, 2017). I expected that while the self-paced students would post fewer posts, those posts would be longer in terms of word count and thread length.

Hypothesis 3 (H3): Discussion Forum Activity and Engagement Affect Cognitive Processing

To test this hypothesis, I created several multi-level regression models (findings are set out in in Chapter 4). I created a regression model for each of the dependent variables (cognitive processing, insight, causation, discrepancies, tentativeness, certainty, differentiation). For this hypothesis, I looked at the relationship that the discussion forum activity had on the demonstration of cognitive processing and the associated sub-scores (insight, causation, discrepancies, tentativeness, certainty, and differentiation) that make up the overall value of cognitive processing. Chiu and Hew (2018) hold the view that making a post (or a comment) utilizes higher levels of cognitive processing because the student who is posting must not only read, but also must understand, the content to which he or she is responding as well as what has been previously posted on that content. Additionally, I expected to find that longer (in terms of word count) posts would have higher levels of cognitive processing. Due to the clustering effect of students making multiple posts within a course and, in some cases, posting in both course versions (self- and instructor-paced), I used a random-intercepts model, also known as a multi-level mixed-effects linear regression. I ran a separate model for each dependent variable and course and report the findings below. I also used post characteristics (e.g., response post, comment versus comment thread) as independent variables to explore their impact on cognitive processing. Using the cognitive processing sub-scores as dependent variables allows for a more refined look at how the discussion forum engagement variables predict the associated elements of cognitive processing. I expected to find that posts by students in the self-paced discussion forums would demonstrate higher levels of insight, certainty, and differentiation than posts by students in the instructor-paced discussion forums.

To prepare my models, I dummy-coded all categorical variables. Since I am studying individual student posts, I clustered observations by student. Because some variables—specifically, original poster, responded to, endorsed, and thread descriptor—were not applicable for every post, I ran three separate models. To provide more interpretable results, I have reported word count by multiples of 100, and words per sentence and six-letter words by multiples of 10.

Elements Affecting Cognitive Processing

My study explored how the pacing condition and specific post-related variables affect the demonstration of cognitive processing by students. To explore these relationships, I identified five elements that affect cognitive processing: (1) pacing condition, (2) post characteristics, (3) post relationship, (4) comment popularity, and (5) substantive quality (Figure 5).

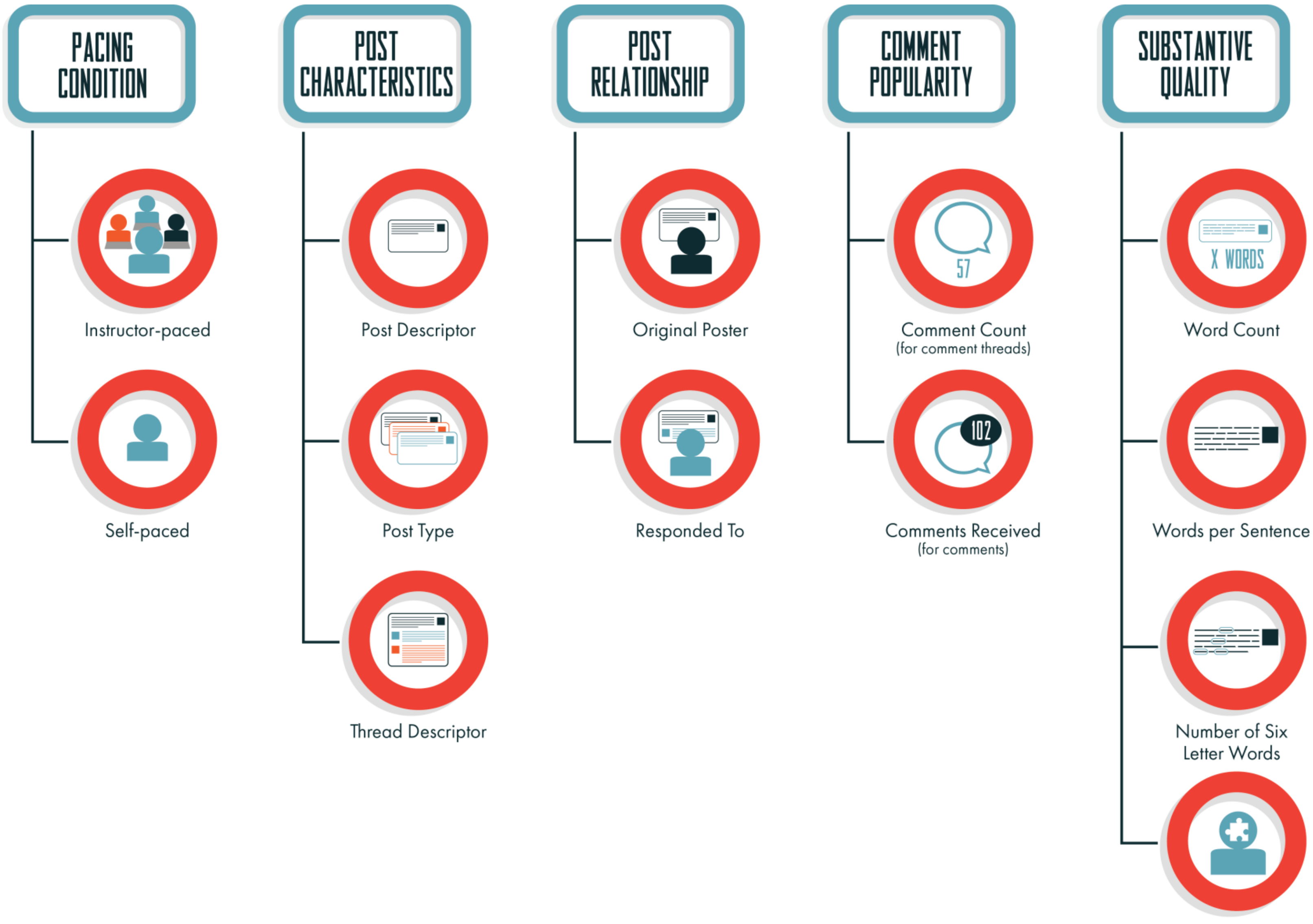

Figure 5. Elements That Affect Cognitive Processing

Pacing condition. For the institution and course provider, a self-paced course may use less resources and be easier to manage, but it is not clear what, if any, impact this pacing condition has on the discussion forum participation by students. Additionally, there is contradictory literature on the question of whether students will feel too isolated to actively participate in self-paced course discussion forums versus the forums in an instructor-paced course.

Post characteristics. It is useful to be able to describe the types of posts that students are making. This study uses variables for post descriptor, post type, and thread descriptor to characterize each post. Students are given the option of (1) making a response or a comment and (2) deciding whether a post will take the form of a new discussion or a question in an existing discussion. Exploring how these decisions and their associated substantive qualities impact cognitive processing can also inform pedagogical decisions. For instance, if it is determined that students are more likely to engage with discussions started by other students, this can be encouraged in the course design.

Post relationship. The original poster and the responded to variables are used to look at the post relationship. Exploring whether a comment came to an instructor or to a student post helps to determine what types of posts are generating student interest and interaction.

Comment popularity. For this component, I looked at the number of comments in comment threads and the number of comments posted about other comments. This information helped narrow the focus to posts that triggered the most forum engagement. Since my unit of analysis was at the individual post level, I was able to look at the associated substantive qualities of the posts that were generating the most number of comments to explore what type of characteristics they had.

Substantive quality. For this variable, I used measures calculated by LIWC, specifically, word count, words per sentence, number of six-letter words, and analytical tone. Each of these can be indicators of substantive quality. By pairing these specific variables with the others that describe the post, I was able to observe which types of words triggered which types of responses. With LIWC providing the sub-parts of cognitive processing (e.g., insight, certainty, discrepancy, differentiation, causation), I analyzed the discussion posts at a more granular level. By determining what types of words make up the most substantive posts, course designers can craft their prompts and forum guidelines to encourage these types of characteristics.

Conclusion

MOOCs are a viable learning environment and have great potential to support the growing call for on-demand instruction from adult learners. Higher education in general needs to gain an enhanced understanding of MOOCs, including how they fit into the current education landscape. Such knowledge can be used to create learning environments that can meet the needs of all kinds of learners. It is also crucial for instructors and course developers to understand the purpose of a MOOC before it is released to students. Is it to provide remediation before a traditional course? To develop communities of practice for learners? Or is it about reaching a certain percentage of completed learners? Presently, there are numerous MOOCs, with a myriad of uses.

Prior research into MOOC discussion forums will provide useful context. Fixating on the goal of retaining users, however, will not be useful—it is a losing battle. Instead of being guided predominantly by thoughts of retention, researchers should also be looking at how to improve the learning environment for the students (Wang et al., 2015). I utilized clickstream data, or data that is collected by a learning management system and tracks student activity within courses, on forum engagement (e.g., number of posts, how long spent in forums) to see which variables are predictive of cognitive presence. Understanding these variables can then inform pedagogical decisions. However, I am mindful that clickstream data by itself can be misleading. While such data can show the start and end points of an action, it does not provide information about exactly what occurs in the time between those two points (e.g., a student could just open a post, move on to something else, and return to it at a later date). However, by pairing clickstream data with a contextual analysis of discussion posts, as I do in my study, one can achieve a more thorough understanding of student activity between start and end points and evaluate the influence the activity has on engagement and the subsequent development of cognitive presence.

A tremendous amount of data is generated from MOOC discussion forums, and determining how best to study the text of MOOC student forums posts to find evidence of cognitive presence can be challenging. As the studies cited above have demonstrated, an automatic text analysis can handle and mitigate that challenge. The impact of this is significant. Being able to efficiently analyze all discussion forum posts allows for more (quantitatively speaking) analysis on learner behavior. This makes it possible for me to take information from the other studies mentioned above—such as how to identify communication patterns, how to use clickstream data—and pair it with what the students are saying within the forums I am studying to see how these variables are influencing the development of cognitive presence. Through an understanding of which factors influence cognitive presence and the extent of that influence, specific pedagogical structures can be identified that, ultimately, could help create more collaborative and engaged learners within MOOC discussion forums.

This chapter provided a description of the community of inquiry and the three associated presences of social, teaching, and cognitive presence. The chapter focused on cognitive presence and explained how it can be linked to critical thinking, a core goal of education. As MOOC discussion forums provide the context for my study, I provided definitions and dimensions of MOOCs and explored how discussion forums have been studied in prior research and how these prior studies inform my own research. I concluded with a look at different ways MOOC discussion forums have been studied and connected each of these approaches to my own work. In the next chapter, I focus on the methodology for my study.

CHAPTER THREE: METHODS

This study explored the influence of MOOC pacing structure (instructor-paced versus self-paced) on forum activity and cognitive processing. The methodology explained in this chapter addresses the following research questions:

1. How does cognitive processing vary between pacing conditions?
2. In what ways does pacing condition affect discussion forum activity and engagement (e.g., number of replies, number of words used)?
3. What aspects of forum activity are predictive of cognitive processing, and do these relationships exist in both pacing conditions?

Design of the Study

This research study employed an embedded multiple-case design in an exploratory case-study approach (Figure 6; Yin, 2014). In this study, the unit of analysis was each individual discussion forum post. I used data taken from instructor-paced and self-paced MOOCs to perform within-course comparisons and explorations (e.g., instructor-paced Visualizing Japan course/forums versus self-paced Visualizing Japan course/forums). A case-study approach was selected to allow for the investigation of a specific phenomenon, particularly, learner engagement in MOOC discussion forums and cognitive presence, in a real-life context (Gallagher & Savage, 2016; Yin, 2014). The multiple-case approach was chosen to avoid the influence of confounding factors such as trying to compare two different courses with different instructional approaches or materials. Instead, each case focused on a specific course and examined the influence that the course structure—i.e., whether it was instructor-paced or self-paced—had on discussion forum engagement, holding constant other variables, such as course requirements, resources provided to students, etc.

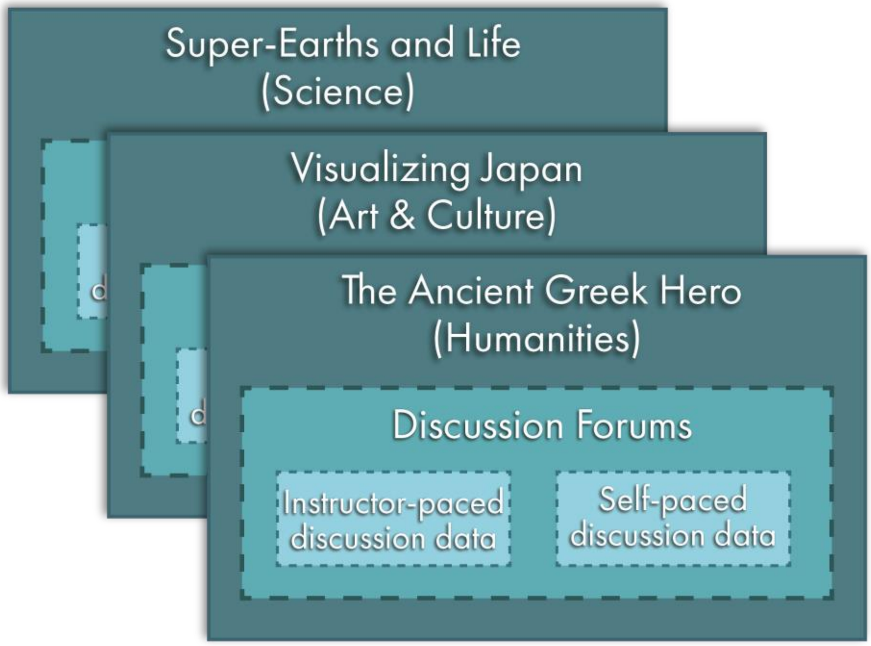

Figure 6. Embedded Case-Study design. Adapted from Yin (2014)

MOOCs are particularly well-suited for case-study research, as they are a fairly new phenomenon and thus present many new research areas that merit further exploration and analysis (Gallagher & Savage, 2016). MOOC research can present major challenges, however, particularly when one is trying to analyze effectively and efficiently the many data points generated. Learning analytics, specifically, text mining, can be used in both unsupervised and supervised machine learning approaches to make this analysis more manageable. The different approaches and frameworks that have been developed to measure critical thinking have been met with various levels of support or criticism, with the Community of Inquiry (CoI) framework rising to the top in terms of popularity. While the CoI model is a well-researched and validated framework, it can be tedious to hand-code responses under this approach (Kovanović et al., 2016).

Generation of Variables

The variables for this study were generated via the Linguistic Inquiry Word Count (LIWC) tool or through an analysis of the discussion forum log files.

LIWC variables. The LIWC variables were generated by analyzing specific words used in discussion forum posts and are explained in Table 1. LIWC uses proprietary dictionaries and scans each post to find words that match specific terms within a given dictionary. LIWC then reports a value, expressed as a percentage, for those words. For example, a score of 8.3 for a post's cognitive processing means that 8.3 percent of the words used in the post were in the cognitive processing dictionary (Simms et al., 2017). To create these values, I uploaded a .csv file, with each discussion forum post occupying its own row, and LIWC then appended the different values.

Discussion forum log files. The discussion forum log files were used to generate the measures for post description and classifications, as well as for response relationships. These variables are explained in full detail in the case context section.

Table 1. LIWC-Generated Variables

LIWC Variable	Description
Word count	The number of words in a post
Words per sentence	The number of words in each sentence
Six-letter words	The number of six-letter words used in a post
Analytical thinking	This is a summary variable which encompasses algorithms made from various LIWC variables based on previous language research
Cognitive processing	This variable is derived from a dictionary of 797 words
Insight, cause, discrepancy, tentativeness, certainty, differentiation	Each of these are sub-parts of cognitive processing, and each has an associated dictionary

Research Context

In previously cited studies of MOOC discussion forums, leveraging learning analytics and, more specifically, examining discussion forum text, allowed for more in-depth text analysis. Some researchers examined MOOC discussion forums to look for evidence of cognitive presence using automatic text analysis tools (e.g., Dowell & Graesser, 2014; Kovanović et al., 2016). Other studies have coded for cognitive presence using the CoI framework, though these have typically been hand-coded studies. The instrument used for the CoI framework is tedious and difficult to code, resulting in hours of manual processing (Kovanovic et al., 2016). This is a feasible approach when the dataset for a study is smaller or when there is a team of researchers, but it is far less feasible for studies looking at MOOC discussion forums, given the volume of posts to be analyzed. A limitation of my analysis tool is that it did not measure interaction or presence between people, or context within the post itself, but instead did a unigram analysis at the individual level.

Case Context

My study focused on the discussion forums of three courses offered by Harvard University on the edX platform: (1) The Ancient Greek Hero in 24 Hours; (2) Visualizing Japan (1850s–1930s: Westernization, Protest, Modernity); and (3) Super-Earths and Life. These courses had a combined student enrollment of 153,768 (Table 2). These courses have evolved and been revised incrementally over time. To track which these changes, Harvard uses a version number which is shared below for each of the observed courses. The version numbers increase with each iteration.

The Ancient Greek Hero. The Ancient Greek Hero course, in the humanities subject area, was originally offered as a seventeen-week course with an estimated time commitment of five to eight hours each week. The instructor-paced course in this study ran from 03/13/2013 to 08/26/2013 and was version 1 for the course. The self-paced version ran from 08/13/2015 to 12/21/2015 and was also version 1. The self-paced version was the first available self-paced course offered after the initial instructor-paced course. This course seeks to use classical texts such as the *Iliad* and *Odyssey* to teach learners about Ancient Greece. No prior knowledge of Greek history or literature is required for enrollment. The course has several learning objectives, including the exploration of the connections and relationships between ritual and myth and between the concepts of lyric and epic.

Visualizing Japan. The Visualizing Japan course, in the art and culture subject area, is a collaboration between MITx and Harvardx and is organized into four modules that focus on the ways images can be used to explain and explore the time period between the 1850s and 1930s. These modules cover Westernization, the expedition to Japan in 1853–1854 by Commodore Perry, the 1905 Hibiya Riots in Tokyo, and the impact of modernity. The instructor-paced version is designed to take six weeks and requires between three and five hours of effort by each learner weekly. The instructor-paced version used in this study ran from 09/15/15 to 11/03/15 and was version 2. The self-paced version ran from 09/01/16 to 08/31/17 and was version 3. The learning objectives for the course include learning how to visualize Japanese history between the 1850s and 1930s; examining the relationship between social protest, modernity, and Westernization using digital images; and understanding how to learn history through visual sources.

Super-Earths and Life. The Super-Earths and Life course falls in the science category, and the instructor-paced version is a six-week course which has a suggested effort of three to five hours per week. The instructor-paced version used in this study ran from 10/13/2015 to 11/29/2015 and was version 2. The self-paced version ran from 10/18/2016 to 02/26/2017 and was version 3. This course focuses on the intersection of astronomy and biology and considers the presence of alien life. The course has as its learning objectives understanding the origin of life on Earth, exploring the discovery of planets, examining the factors that make a planet inhabitable, and discussing how we search the universe for signs of life.

Table 2. Course Name, Dates, and Student Enrollment

Course Name	Course Dates	Student Enrollment
The Ancient Greek Hero (Instructor-Paced) Abbreviation: HeroIN	03/13/13–08/26/13	43,539
The Ancient Greek Hero (Self-Paced) Abbreviation: HeroSE	08/13/15–12/21/15	8,094
Super-Earths and Life (Instructor-Paced) Abbreviation: SuEarIN	10/13/15–11/29/15	81,121
Super-Earths and Life (Self-Paced) Abbreviation: SuEarSE	10/18/16–02/26/17	9,798
Visualizing Japan (Instructor-Paced) Abbreviation: VjXIN	09/15/15–11/03/15	3,834
Visualizing Japan (Self-Paced) Abbreviation: VjXSE	09/01/16–08/31/17	7,382
	Total	**153,768**

Discussion Forum Structure

Each course uses the edX learning management platform and has similarly structured discussion forums. The discussion forum is organized into discussion topics determined by the course team. There are two types of discussion topics: course-wide and content-specific. Course-wide topics are general and affect the entire course, e.g., “General” and “Teaching Staff Office Hours” (see Figure 7). These course-wide topics are accessible through the discussion page within the course site, and students can post questions or new discussions or replies to existing posts. Content-specific discussion topics are added as part of a course unit and relate to specific video lectures, reading assignments, homework problems, or other course content, e.g. “Day 1: Course Goals” and “Day 1: Introduction to Shiseido Module” (see Figure 7). Content-specific discussion topics can be accessed through the discussion link or within the course unit on the course page.

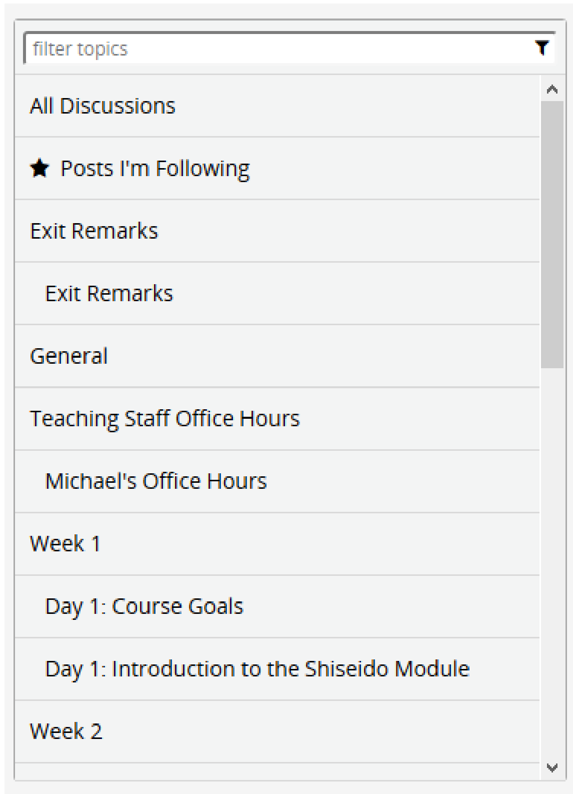

Figure 7. Screenshot of Course-Wide and Content-Specific Discussion Topics.

Adding a new post. When preparing to post, a student has the choice of framing his or her post as a question or as a discussion. If the student opts for the latter, he or she must then designate a topic area (Figure 8).

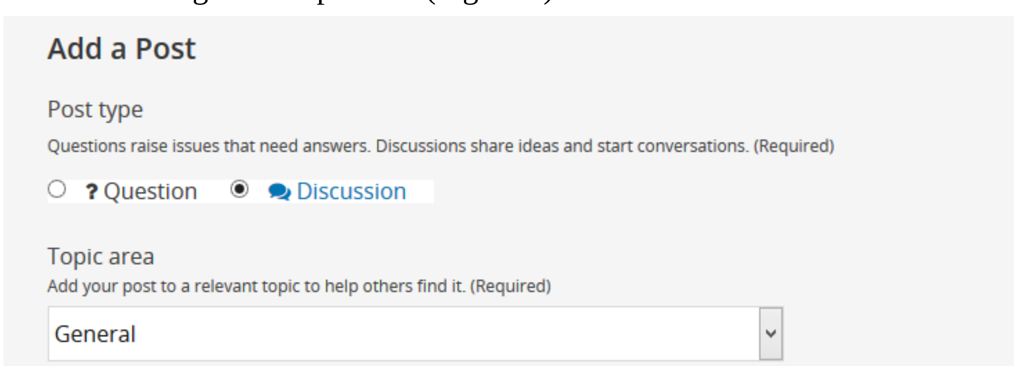

Figure 8. Screenshot of Add a Post Options

Posting to content-specific topics. For content-specific topics, students can initiate a conversation (an initial post), make a reply to a post (a response), or expand on a response (comment) (Figure 9). Students must assign their posts to one of the topics created by the course team; they cannot make up their own topics.

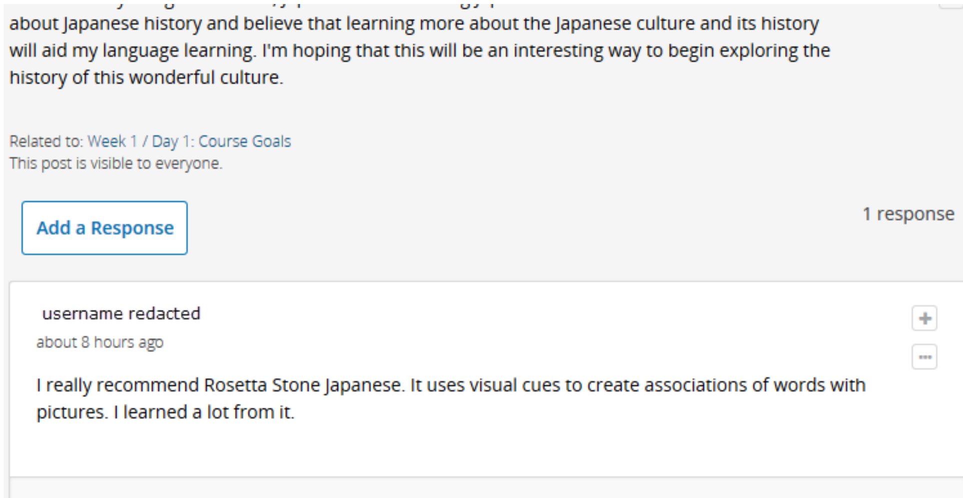

Figure 9. Screenshot of Response Screen.

Additional functionality. The discussion forum provides additional functionality for students, such as the ability to pin and follow posts, to post anonymously, and to vote specific posts "up" or "down" (akin to the "Like" feature on Facebook). These options, while available to students within the courses studied, were used only minimally, and so they were excluded from the analysis.

Case Selection and Participants

The courses in my study were selected using several criteria. The first requirement was that the course needed to be offered by Harvard as both an instructor-paced and a self-paced MOOC. It was essential that the course materials, assignments, and assessments remained the same between the two course structures. Another requirement was that there needed to be activity within the discussion forum in the self-paced MOOC, as that is the area of interest for my study. Harvard provided me with data from multiple iterations of the courses in the both the self-paced and instructor-paced formats. In my review of the associated files, I selected cases that had sufficient discussion forum activity. The percentage of students posting in a given course's discussion forum was fairly consistent between the pacing conditions. The overall content remained consistent between the courses. When a course became a self-paced course, the course designers started with the instructor-paced course, copied it, and then made changes to it, such as removing the deadlines and making other minor modifications to the course's structure. I purposefully selected courses representing several different subject areas—humanities, science, and art and culture. When a student signs up for a MOOC with Harvard, he or she agrees to be part of research efforts (see http://harvardx.harvard.edu/research-statement).

As previously mentioned, 153,768 students registered for the six courses in this study. Some students registered for both the instructor-paced and self-paced versions of the course. Approximately 52 percent of all registered students were male (Table 3) and approximate 44 percent had either a bachelor's or a master's degree (Table 4).

Table 3. Registered Student's Indicated Gender, by Course

Gender	HeroIN	HeroSE	SuEarIN	SuEarSe	VjXIN	VjXSE	**Total**
Male	21,043	3,305	45,093	5,134	1,761	3,146	**79,482**
Female	18,408	3,577	31,644	3,684	1,572	3,079	**61,964**
Other	113	50	395	104	37	130	**829**
None provided	3,975	1,162	3,989	876	464	1,027	**11,493**
Total	43,539	8,094	81,121	9,798	3,834	7,382	153,768

Table 4. Registered Student's Education Level, by Course

Education Level	HeroIN	HeroSE	SuEarIN	SuEarSe	VjXIN	VjXSE	**Total**
Bachelor's degree	13,735	2,202	19,494	2,363	1,163	2,175	**41,132**
Master's or professional degree	10,693	1,988	9,213	1,395	978	1,586	**25,853**
Doctorate	2,403	405	1,241	240	188	279	**4,756**
Associate’s degree	504	276	8,144	738	166	313	**10,141**
Secondary/high school	9,110	1,513	31,444	3,000	652	1,482	**47,201**
Junior secondary	1,121	226	3,113	606	83	214	**5,363**
Elementary	222	37	540	131	11	25	**966**
No formal education	256	22	243	30	9	20	**580**
Other education	1,320	161	2,742	247	72	114	**4,656**
Did not specify	4,175	1,264	4,947	1,048	512	1,174	**13,120**
Total	43,539	8,094	81,121	9,798	3,834	7,382	153,768

My study focused on student participation within discussion forums. For purposes of this study, I excluded analysis of instructor posts. There were 13,495 unique students who posted in the course discussion forums across the six course variations studied. As previously mentioned, some students were enrolled in multiple courses and posted in more than one course discussion forum, which explains why some of the totals for posters will exceed 13,495. Discussion forum participation tended to be higher in the instructor-paced courses than in the self-paced courses (Table 5), with the overall participation rate coming in at a bit over 9 percent of registered students. One explanation for this is that there was a considerably higher number of student responses to the initial introduction posts (made in the instructor-paced courses) than to the posts in the self-paced courses.

Table 5. Number of Students Who Posted in Forum, by Course

Course	# of Posters	% of Enrolled Students (Total Number)
The Ancient Greek Hero (Instructor-Paced)	3,871	8.9 (43,539)
The Ancient Greek Hero (Self-Paced)	534	6.6 (8,094)
Super-Earths (Instructor-Paced)	8,003	10.0 (81,121)
Super-Earths (Self-Paced)	881	9.0 (9,798)
Visualizing Japan (Instructor-Paced)	481	12.5 (3,834)
Visualizing Japan (Self-Paced)	461	6.2 (7,382)
Total	14,231 posters	9.3% 100.0%

If one filters the student population by those who visited a given course at least once (Table 6), the forum participation rate becomes 20 percent. If one filters by those who viewed at least half of the chapters of the course (Table 7), the forum participation exceeds 100 percent.

Table 6. Students Who Viewed the Course at Least Once and Posted at Least One Discussion Forum Post

Course	Did Not View/Post	**Did View/Post**	Total
The Ancient Greek Hero (Instructor-Paced)	305	**3,566**	3,871
The Ancient Greek Hero (Self-Paced)	58	**476**	534
Super-Earths (Instructor-Paced)	209	**7,794**	8,003
Super-Earths (Self-Paced)	104	**777**	881
Visualizing Japan (Instructor-Paced)	14	**467**	481
Visualizing Japan (Self-Paced)	16	**445**	461
Total	706	**13,525**	14,231

Table 7. Students Who Viewed at Least Half of the Chapters and Made at Least One Discussion Forum Post

Course	Did Not View/Post	**Did View/Post**	Total
The Ancient Greek Hero (Instructor-Paced)	2,855	**1,016**	3,871
The Ancient Greek Hero (Self-Paced)	401	**133**	534
Super-Earths (Instructor-Paced)	4,432	**3,571**	8,003
Super-Earths (Self-Paced)	486	**395**	881
Visualizing Japan (Instructor-Paced)	223	**258**	481
Visualizing Japan (Self-Paced)	232	**229**	461
Total	8,629	**5,602**	14,231

Data Collection and Management

This study relied on existing data provided by Harvard University's Vice Provost for Advances in Learning (VPAL) Research Team. The study received approval from the institutional review boards of North Carolina State University and Harvard University. After a comprehensive data use verification process, which included multiple security protocols, a data use agreement was signed by the relevant parties at both schools. The data was digitally transferred as an encrypted package of .csv files and stored on an encrypted hard drive. The full data transfer included data from 59 courses, clickstream data, and survey responses, as well as full discussion forum texts from 867,597 student and instructor records.

Data Source

The courses chosen for this study are running on the edX platform, which captures clickstream data, discussion forum engagement variables, and written posts. My primary research question asks, “How does cognitive processing vary between pacing conditions?” This was answered by using LIWC to calculate the number of words and sentences in each discussion forum post. I used the clickstream data provided by Harvard to determine the number of replies to a given post, the position of the post within the thread, and whether the post was a reply to an instructor or to a student. I also used the classification types, e.g., comment or thread, initial or response post, to further describe and analyze each discussion post. These served as my “engagement variables.” The expectation was that more words and sentences per post, more replies to posts, and deeper posts/threads would signal more engagement. These variables allowed me to answer my second and third research questions, which concern the activity and engagement within courses and determining which of these variables are predictors for cognitive processing. Cognitive processing has several sub-parts, and I used LIWC to find the values for the dependent variables of cognitive processing, insight, certainty, tentative, discrepancy, differentiation, and causation. The latter scores are the sub-scores that make up the cognitive processing score. I included these sub-scores to allow for a more refined analysis of the types of words being used within posts.

Data Cleaning and Structuring Process

This study used a post-hoc analysis in that no interaction existed between the researcher and the learners and the data was analyzed after the conclusion of each course studied. Harvard’s course-delivery structure (the edX platform) was consistent for each of the studied courses. The courses had defined start and end dates. In addition to selecting courses that had instructor-paced and self-paced versions, I selected courses that had activity in the discussion forums. In order to clean and structure the data, I used a combination of Excel functions and formulas, Stata commands, and hand-coding and review (Figure 10).

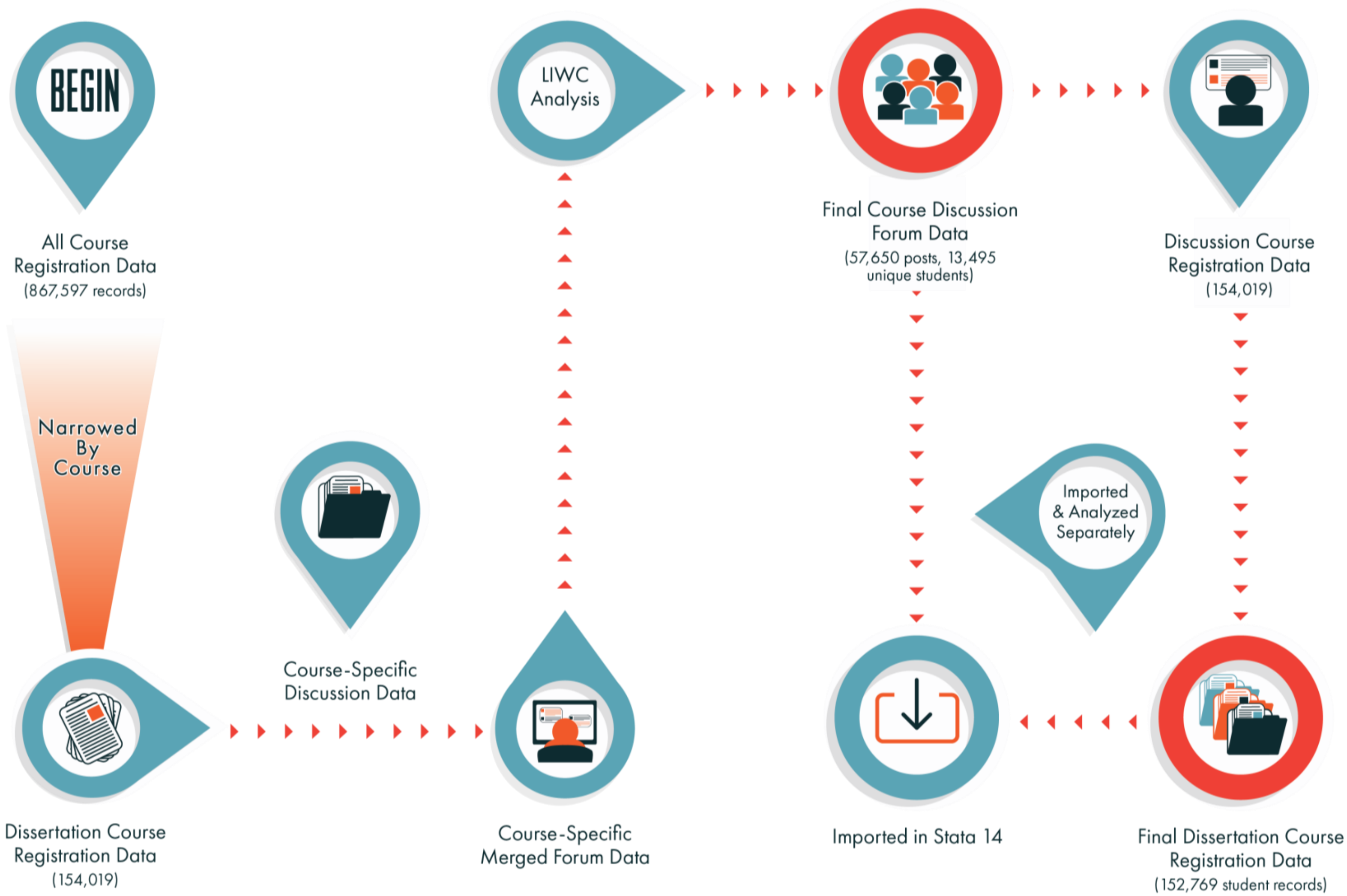

Figure 10. Data Cleaning and Structuring Process

The steps in my data cleaning and structuring process are outlined below.

1. Create master roster. The first step of the data cleanup was to create the dissertation course file. Harvard sent me a file that included demographic information, including roles, pre-course survey responses, and log file information per user. This log file data included actions such as how many course content chapters, activity within the forum (forum events), and, where applicable, quiz and test scores (assessment performance). It also included the data about how many students viewed, explored, completed, and certified in each of the courses. This file had 867,597 records and included instructor, staff, and student data. For the purposes of my study, I changed any users with the role of staff to be "instructor," creating two roles in the dataset – either student or instructor. This file included a column featuring the course identification (ID) number for each offering, and I used that column to reduce the file to reflect only the students and instructors participating in the six courses selected for my study. This reduced the file to 154,019 instructor and student records.

2. Create discussion forum file. Each course had separate "forum", "forum_posts", "forum_person", and "course_axis .csv" files. The data was slightly different between the different forum .csv files, which had an identifier between the files that allowed me to merge the relevant fields into one file for each course. For instance, the forum_posts file contained information about the original poster and the responded to element, but the forum file did not. The forum_posts file only had the first 100 characters of discussion forum text, while the forum file contained full text. Using vlookup functions, I created a new column in my merged forum file that listed the role of the user derived from the master roster file. From there, I deleted any posts in my merged forum files that were made by instructors. The columns reporting "original poster" and "responded to" listed usernames. I used another vlookup function to match these usernames to my roster file and reported whether each was an instructor or a student. Once I had identified whether the original poster or responded to entry related to an instructor or to a student, I removed the original columns that listed the usernames. For the purposes of my study, I did not track which specific student or instructor the post was made in response to. I simply examined whether it was a response to either an instructor or a student or whether it was an initial post.

3. Discussion forum post cleanup. Next, I did data cleaning of the actual discussion forum posts. I had originally planned to use a spellcheck application for the posts but ultimately decided not to for a few reasons. First and foremost, I felt that spellchecking a post would potentially influence the cognitive processing score in a way that would not accurately reflect the content of the actual post. Also, any spelling mistakes in a post may have adversely impacted peer engagement with it (e.g., reading, replying, commenting), and correcting the spelling could have skewed the data analysis. The second reason I decided against spellchecking has to do with the sheer volume of my data. With more than 57,000 posts, it simply was not feasible to spellcheck every post. Finally, the fact that I did not know content-specific terminology used in the courses would have made it nearly impossible for me to even identify, much less correct, perceived spelling errors. I deleted any discussion posts that obviously were spam messages, were gibberish, or were made in a different language or made using indecipherable characters. Additionally, if the only text a student used in a post was an emoticon or a grouping of emoticons, e.g., ☺, I removed the post. There were some data inconsistencies between the master student roster and the forum data. In these cases, if a given student's ID number was not in the master student roster file provided by Harvard, I removed the student from the dataset. Twenty-seven students in all were so removed.

4. Creation of post position variable. I created a variable for comments and replies made to an existing comment or post. To do this, I sorted first by the initial post ID (which tracked whether there was a reply to the initial post) and then by the post creation time stamp. I then numbered the posts sequentially in the order they were made. Posts receiving higher numbers tended to be student replies to the instructor's general initial posts, such as those on introductions or course goals.

5. Merging course-specific forum data. At this stage, I now had six separate course-specific forum data files, which I then merged into one .csv file.

6. LIWC analysis. I used LIWC and ran the merged forum data file and LIWC appended variables for word count, words per sentence, number of six-letter words, and the cognitive processing and associated sub-part scores. After the data cleanup and merging of the discussion forum files, I ended up with 57,650 discussion posts from 13,495 students across all of the courses. Each self-paced course had at least 1,600 discussion forum posts (Table 8).

Table 8. Number of Posts, by Course

Course	**Number of Posts**	Percent of Total Posts
The Ancient Greek Hero (Instructor-Paced)	**30,206**	52.4
The Ancient Greek Hero (Self-Paced)	**4,118**	7.14
Super-Earths (Instructor-Paced)	**18,202**	31.57
Super-Earths (Self-Paced)	**1,866**	3.24
Visualizing Japan (Instructor-Paced)	**1,604**	2.78
Visualizing Japan (Self-Paced)	**1,654**	2.87
Total	**57,650**	100

7. Update to discussion course registration data. One of the things I wanted to report on was the question of whether students who viewed or explored a given course also participated in the course's discussion forum. To do this, I did another vlookup, which added a new column to the master roster file that indicated whether a student had made at least one discussion forum post. Once I had made that change, I removed the instructors from the course registration data file, and this reduced that file to 153,769 student records.

8. Preparing for Stata upload. Next, to facilitate the upload to Stata and subsequent data analysis, I converted any categorical variables into numbers. I used the following categorical variables in my analysis: course name, post descriptor, post type, thread descriptor, endorsement, original poster, and responded to. Each of these categorical variables was generated through edX and provided in the forum and forum_posts files. I converted the text to the numbers as shown in my codebook (Table 9). Once the values were entered into Stata, I created labels for each variable, allowing me to see the text in my outputs.

Table 9. Codebook for Categorical Variables

Categorical Variable(s)	Values
Course	1 = The Ancient Greek Hero (Instructor-Paced) 2 = The Ancient Greek Hero (Self-Paced) 3 = Super-Earths (Instructor-Paced) 4 = Super-Earths (Self-Paced) 5 = Visualizing Japan (Instructor-Paced) 6 = Visualizing Japan (Self-Paced)
Post descriptor	1 = Initial post 2 = Response post 3 = Comment
Post type	1 = Comment 0 = Comment thread
Thread descriptor	1 = Discussion 0 = Question
Endorsement	1 = Yes 0 = No
Original Poster, Responded to	1 = Student 0 = Instructor

9. De-identification of data. I received identifiable data, but specific names were not important to my study. As explained in the streamlining process, I broke one of the identifiable connections by converting the “original poster” and “responded to” username references to “instructor”, “student”, or “null”. After I finished my data cleanup and uploaded my respective .csv files to Stata, I utilized Stata functions to anonymize the user IDs. I also deleted the post identifiers, such as the Mongoid and initial post IDs, since they were no longer needed. The resulting Stata dataset is not identifiable to any specific user in any of the courses.

Data Analysis

Online discussion forums, particularly in MOOCs, are the spaces where a significant amount of student-student interaction takes place (Wong, Pursel, Divinsky, & Jansen, 2015). For an online course, the instructor may have weekly assignments requiring students to post original thoughts and replies to the course's discussion forum. Analyzing discussion forums can be challenging, given the amount of data points that can exist, and this is particularly true when one is dealing with a MOOC, which can have between 200 and 10,000 learners in each course. One way to better understand and analyze discussion forums is through learning analytics, an approach that You (2016) says enables "data-driven decision making while improving institutional productivity" (p. 23). Ezen-Can, Boyer, Kellogg, and Booth (2015) point out that there has been a marked increase in interest in using learning analytics to better understand student activities within MOOCs. Specifically, the researchers state that "one very important source of data in MOOCS is the textual dialogue among students . . . on discussion forums" (p. 146). Again, when looking at datasets as large as those generated from MOOCs, it is imperative to find and use automated ways to analyze such data (Donnelly & Gardner, 2011).

Quantitative Content Analysis (QCA)

Joksimovic et al. (2014) found that different phases of cognitive presence contain distinct levels of word use. These researchers point out that the analysis of automatically-extracted linguistic features of online discussion transcripts can be beneficial in identifying psychological factors of learning in communities of inquiry (CoIs). Tausczik and Pennebaker

(2010) elaborate by stating that there is great potential in exploring the connection between the three dimensions of the CoI model and the psychological meaning of words. The process of automatically analyzing these discussion forum posts is termed “quantitative content analysis (QCA),” which Corich, Kinshuk, and Hunt (2006) identify as one of the most popular research approaches for measuring the evidence of cognitive presence in online discussion forums. They define QCA as “involving breaking transcripts into units, assigning the units to a category and counting the number of units in each category” (p. 2). This approach is sometimes referred to as the quantifying of qualitative data. One of the challenges in analyzing discussion forum posts is that a vast amount of text needs to be coded and read through. With QCA, there are techniques to automate this process through the use of tools such as Coh-Metrix (McNamara et al., 2014) or Linguistic Inquiry and Word Count (LIWC) (Tausczik & Pennebaker, 2010).

QCA tools present new opportunities for measuring and evaluating evidence of critical thinking in online discussion forums. These tools can drive future pedagogical approaches in online learning environments. My research design used automatic text analysis to analyze the qualitative data. I paired this analysis with selected clickstream data points, creating a regression model that explored the influence of course structure on levels of cognitive processing within a given discussion forum.

The CoI model provided the theoretical framework for this study. I selected this model because it is the most studied of the models dealing with online discussion forums (Breivik, 2016), as I outlined in Chapter 2. But while the CoI model provides a useful lens for evaluating the development of cognitive presence in discussion forums, it is difficult to apply and analyze in the context of the large datasets found with MOOC discussion forums.

To help mitigate this, I deployed quantitative content analysis (QCA) in a unique way that allowed for the analysis of discussion forums to detect cognitive processing. The utility of QCA, that is, its ability to efficiently analyze a large corpus using automated text analysis, creates new opportunities to make pedagogical changes and to see how those changes can influence the demonstration of cognitive processing within MOOCs.

Linguistic Inquiry Word Count (LIWC). I selected the 2015 version of the LIWC tool to analyze discussion forum text. As Kahn et al. (2007) explain, LIWC compares text to pre-determined dictionaries with specified categories and outputs a numerical frequency value of post content along these categories, ranging from simple articles and prepositions to nuanced emotion, affective, and cognitive words. Pennebaker, Boyd, Jordan, and Blackburn (2015) use the term “target words”, which they define as the words analyzed by LIWC. In this study, the target words are those found in the discussion forum posts. The words found in LIWC dictionaries are called “dictionary words”, and the term “word categories” refers to groups of words that describe a specific domain, such as “insight” or “discrepancy” (Pennebaker et al., 2015). For this study, I am particularly interested in the concept of cognitive presence, which is measured using the “cognitive processes” word category in LIWC. This category is made up of a total of 797 words that can be further divided into sub-categories of “insight”, “cause”, “discrepancies”, “tentativeness”, “certainty”, and “differentiation” (Table 10). The words that appear in these subcategories (e.g., “insight”, “causation”) are also in the main word category (“cognitive processes”). Reporting the overall word category value as well as the scores for the sub-categories allowed for a deeper analysis of which word categories had the most influence on cognitive processes within the discussion forums. It also allowed for a deeper understanding of the levels and type of student engagement within the forums. LIWC's usefulness has been validated through various studies, including those looking at cognitive presence (Kovanović et al., 2016; Oztok, Zingaro, Brett, & Hewitt, 2013; Pennebaker et al., 2015), and its cognitive processing score has been found to have high levels of predictive validity (Slotter & Ward, 2015; Tausczik & Pennebaker, 2010).

Table 10. LIWC Word Categories

Category	Examples	# of Words in Category
Cognitive processes	Since, seem, sort, sense, wish, somewhat	797
Insight	Think, realize, question, accept, notice, perspective	259
Causation	Because, effect, hence, affect, based, create	135
Discrepancy	Should, would, could, hope, lack, odd	83
Tentative	Maybe, perhaps, guess, unknown, wonder	178
Certainty	Always, never, clearly, confident, commit, apparent	113
Differentiation	Hasn't, but, else, despite, although, except, exclude	81

Procedures

In order to answer my research questions, I used independent sample t-tests, as well as multi-level mixed-effects regression models, also referred to as random-intercept models. As outlined, I had a large dataset, and the size of the data can make just about any statistical test show a significant difference (Sullivan & Feinn, 2012). For each of my t-tests that produced statistically significant results, I did a post-hoc analysis to report the effect size as measured by Cohen's *d*, which is a way to report effect size (Khazaei et al., 2017). This study also uses multi-level mixed-effects regression models to explore the influence of an independent variable (course structure) and mediators (e.g., post type, number of words in a post) on the dependent variables (cognitive processing score and the associated sub-scores).

Independent variable. For this study, I used the pacing condition—instructor-paced and self-paced— as my independent variable.

Mediators. The mediators for this study were the following measures of discussion forum activity and post characteristics: (1) number of words in the post, (2) number of replies to the post, (3) words per sentence within the post, (4) number of six-letter words in the post, (5) analytical thinking, (6) number of comments, (7) the post's position within the discussion thread, and (8) the post's characteristics (e.g., type, descriptor). The mediators were used to explore the direct relationship between dependent variables, as well as the indirect relationship that these mediators have with the independent variable on the dependent variable.

Dependent variables. In this study I used several dependent variables, including cognitive presence, which is measured in this study using the cognitive processing variable in LIWC. In addition, I used the following sub-scores of the cognitive processing variable: (1) insight, (2) certainty, (3) tentative, (4) discrepancy, (5) differentiation, and (6) causation. These specific variables are all components of cognitive processing. Each post was analyzed by LIWC for each of the dependent variables, and separate regression models were run for each dependent variable.

For the data analysis I used Stata 14.2. I created a representative sample version of one of the course files to do test analyses and to work out issues with Stata commands or data structuring. One of the benefits of using Stata are the do-files. These files allow users to save commands and then apply them to other datasets. Since all of the commands and models had to be run on three sets of data (from the three courses studied), the do-files were a way to manage that process and ensure that each dataset was getting identical treatment and analysis.

In a regression, there is an assumption that all observations are independent of each other. However, in this study, there was a clustering effect, as some people may have a certain style to their posts that may make them more like other people's posts. Given this clustering effect, it was essential to use a multi-level model. This grouping or clustering could have affected the standard errors, and thus would have created Type 1 errors that would have caused predictors to appear to have significant effects when in fact they did not (Keith, 2015; Paterson & Goldstein, 2010; Steenbergen & Jones, 2010). Using a multi-level model also provided for a more thorough analysis of precisely what influence group-level variables had on individual-level variables (Keith, 2015). For each dataset, I used the student's ID number as the clustering variable. In my study, the unit of analysis was the discussion forum, and group characteristics such as number of replies and sustained communication were measured through an individual-level variable, the dependent variable of cognitive processing. I created regression models for each course and identified which of the independent variables had influence on the demonstration of cognitive processing. I present my findings for each case; this presentation allows for some cross-case comparisons, which are shared in Chapters 4 and 5.

Reliability and Validity

For the content analysis of this study, I used peer reviews and external audits of my findings to check for inconsistencies. I also used triangulation of the data collected (Creswell, 2013; Merriam & Tisdell, 2016). I worked with experts in quantitative analysis to inform the structuring of my data and the subsequent completion of my analysis. These experts had the dual roles of being peer reviewers and external auditors. I also shared my preliminary analysis and findings with peers studying MOOC discussion forums, particularly those using automatic text analysis, to get their feedback on my findings and interpretations. Finally, as part of the dissertation process, I worked closely with my committee members on the analyses done within my study. External validity was established by using multiple subject domains (science, humanities, and art and culture) (Creswell & Clark, 2011). This allowed for generalizations to be made about the interpretations in the study. Additionally, the use of multi-level analysis further provided a test of the generalizability of the findings (Steenbergen & Jones, 2010). For the automatic text analysis, I used a validated tool, LIWC, which allowed me to establish criterion-related validity (Creswell & Clark, 2011; Pennebaker et al., 2015; Tausczik & Pennebaker, 2010).

Limitations

The most significant limitation in this study was that I was not able to hand-code discussion forum post content and align it to a specific model. Instead, I relied on the validity of the LIWC tool to identify the cognitive processes within each of the discussion posts examined. But this may not necessarily be a bad thing, as LIWC is not subject to human bias (Slotter & Ward, 2015; Tausczik & Pennebaker, 2010). Another limitation was that I was only able to measure cognitive processing for the students in the three courses based on what they shared in the discussion forums; I cannot account for other ways they may have demonstrated their learning. As a significant number of the enrolled students will not make any posts, let alone engage in sustained communication, a majority of the enrolled students were effectively excluded from this study. This limits the impact and true understanding of cognitive processing in these MOOCs because only the most active students are being studied. This study did not attempt to understand the motivations behind a student's decision not to participate in a discussion forum; I sought only to make recommendations for future course pacing structure decisions. Ultimately, this study provided empirical findings regarding the level of engagement and activity of students in self-paced and instructor-paced MOOCs and also further supported the previous works that have used learning analytics to better understand cognitive presence within MOOC discussion forums. A final limitation is that LIWC only counted the number of times specific words appeared in a given post; it was not able to account for context and positioning of one post in relation to another post.

Conclusion

In this chapter, I outlined the methodological approach for my study, which made use of quantitative content analysis to measure the levels of cognitive processing in three selected MOOCs. I detailed my process for data cleanup and structure and provided initial descriptive statistics, including demographic information and forum participation. In Chapter 4, I provide my findings from the analysis described in this chapter.

CHAPTER FOUR: FINDINGS

This study explores the influence of massive open online course (MOOC) pacing structure (instructor-paced versus self-paced) on course discussion forum activity and cognitive processing. A multiple-case study approach was used to focus on three MOOCs offered by HarvardX that each featured both an instructor- and a self-paced pacing version. This chapter provides findings on the following research questions:

1. How does cognitive processing vary between pacing conditions?
2. In what ways does pacing condition affect discussion forum activity and engagement (e.g., number of replies, number of words used)?
3. What aspects of forum activity are predictive of cognitive processing, and do these relationships exist in both pacing conditions?

The chapter commences with a thorough explanation of the way discussion forum posts were classified in this study, including the descriptive statistics for each of the post classifications. Next, each of the three research hypotheses introduced in Chapters 2 and 3 are revisited and statistical findings for each are presented. The results are shared for each paired course.

Course Profiles

This study analyzed 57,650 discussion forum posts generated by 13,495 students across three paired courses. The way each post was classified is shown in Figure 11.

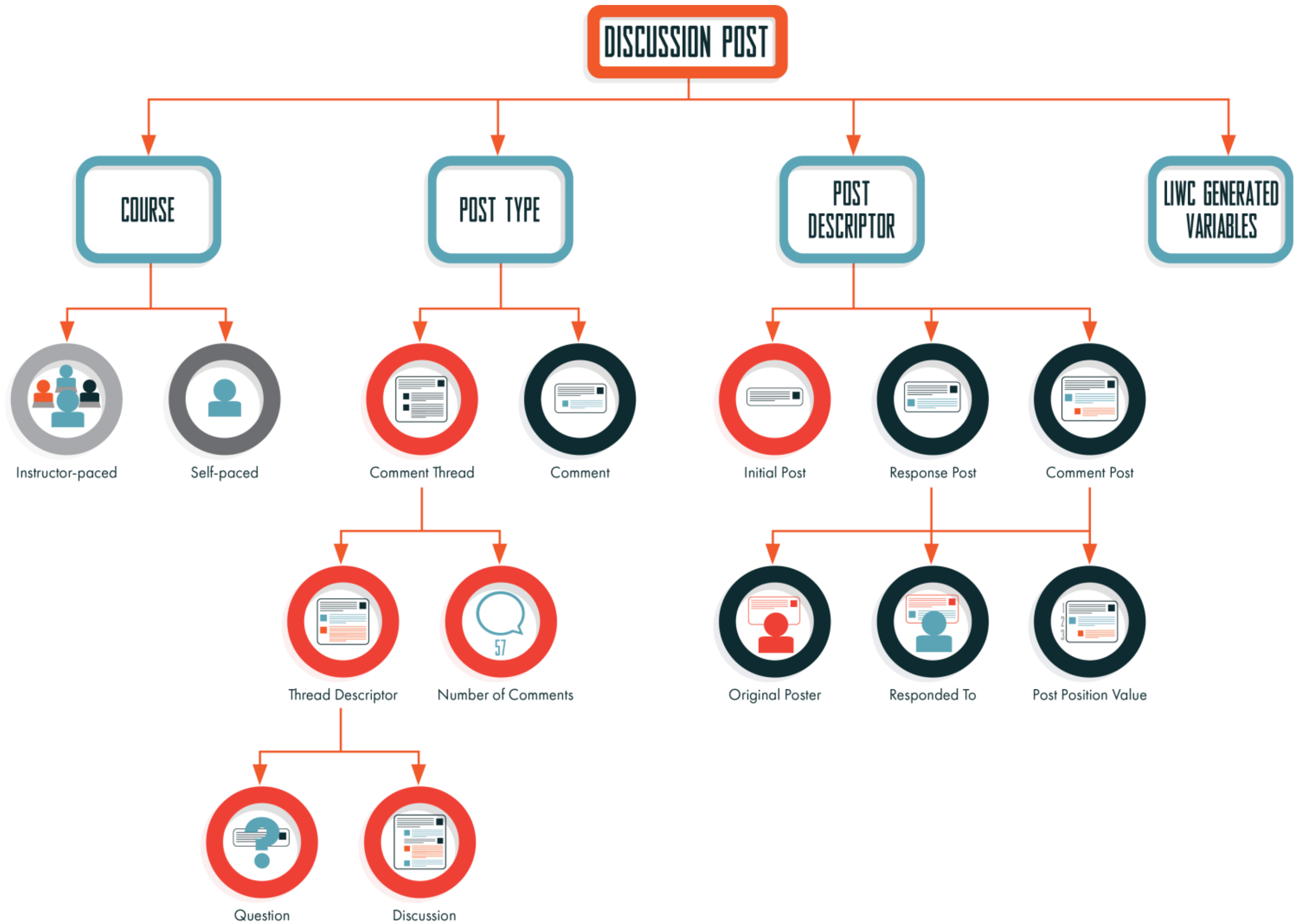

Figure 11. Post Classification

For each post, there was a pacing condition, post type, and post descriptor, all of which came from the discussion forum log files. These classifications are universal across all edX courses. Every post has a post type: either a comment or a comment thread. The post type can be further classified by the post descriptor variable. A comment post type can have a post descriptor of either a response post or a comment post; every comment threads has a post descriptor of an initial post. Just as the post descriptor refines the comment post type, the thread descriptor variable refines the comment thread post type, characterizing a comment thread as either a question or a discussion. For each response or comment post, the original poster variable shows whether that post was made in reply to an instructor or a student and excludes initial posts made by students. The responded to variable is also for either response or comment posts and tracks whether the post was made in response to a student or an instructor's previous post and whether it was generated by the log files. This variable refines comment tracking and can show whether an instructor-originated thread is generating student-to-student interaction. The log files gave the username for the original poster and the responded to variables. I used a series of Excel formulas to recode these elements as either "instructor" or "student." Posts that were particularly helpful or useful were tagged by instructors as "endorsed". With such a large volume of posts, this proved to be an effective way to help students identify the most important posts.

I appended additional variables for each post using the Linguistic Inquiry Word Count (LIWC) text analysis tool. These variables included word count in a given post, number of words per sentence in the post, and number of six-letter words used. an analytical thinking score (a summary variable calculated by LIWC) and scores for cognitive processing and each of the sub-parts of cognitive processing (insight, certainty, discrepancy, differentiation, tentative, and causation). For some of these variables, there were empty or null values (e.g., thread descriptor, original poster) due to the fact that the variable was only applicable to a specific type of post. For some posts, no variables were applicable, and these posts were ignored in the Stata analysis.

Ancient Greek Hero

In the Ancient Greek Hero in 24 Hours course, the students were examining literary texts and, more specifically, exploring the definition of “hero” in the context of classical Greek texts. The course had several learning objectives, including the exploration of the connections and relationships between ritual and myth and between the concepts of lyric and epic. While there was a large difference in the number of student posts in the instructor-paced versus the self-paced version of this course (30,206 to 4,118), the breakdown by post classification type was consistent between the two course versions (Table 11). Both pacing conditions appeared to have robust student-to-student interaction. The higher percentage of discussion-type versus question-type initial posts in both course variations indicates that students understood the course content and structure and were focusing on exploring the content of the course.

Instructor-paced. There was a total of 30,206 student posts in this version of the course. Most of these posts (73.2 percent, or 22,099) were comments. Of these comments, slightly more than 54 percent were comment posts. The other 8,107 posts in this course were initial posts, and of them were discussions. The students in this instructor-paced course were more likely (almost 80 percent) to respond to an initial post made by a student and were more likely (nearly 86 percent) to respond to a student post generally. Each of these indicators suggests that students were interacting with each other’s posts within the discussion forum.

Self-paced. The self-paced version of the Ancient Greek Hero course had 4,118 student posts. Similar to the instructor-paced version, the majority (nearly 87 percent) of the posts were made as comments instead of comment threads. Also like the instructor-paced version, most of these posts (about 51 percent) were made as comment posts. When students made initial posts, they marked most of them (80 percent) as discussions. Students were more likely (54 percent) to respond to an initial post made by a student and more likely (86 percent) to respond to a student post generally.

Table 11. Classification of Ancient Greek Hero Student Posts, by Pacing Condition

Pacing Condition	Post Type (% of Total Posts)	Post Descriptor (% of Total Posts)	Thread Descriptor (% of Total Posts)	Original Poster (% of Total Posts)	Responded to (% of Total Posts)
Instructor-Paced	Comments (73.2%)	Comment Post (54.1%)	Discussion (100%)	Students (79.6%)	Students (85.95%)
Self-Paced	Comments (86.9%)	Comment Post (50.9%)	Discussion (80.4%)	Students (53.8%)	Students (85.8%)

Note: Percentages reflect those expressed within each classification.

Super-Earths and Life

In the Super-Earths and Life course, the students learned about the intersection between astronomy and biology and explored the existence of alien life. The course had as its learning objectives understanding the origin of life on Earth, exploring the discovery of planets, examining the factors that make a planet inhabitable, and discussing how we search the universe for signs of life. Just as with the Ancient Greek Hero course, the classification percentages observed in the Super-Earths course were consistent between the two pacing conditions (Table 12). A noticeable difference between the Ancient Greek Hero and the Super-Earths courses was the increased instructor presence in the Super-Earths offering. Students in the Super-Earths course responded to more specific instructor-initiated prompts than their Ancient Greek Hero counterparts, and the course had more of a structured and sequential approach than the Ancient Greek Hero course. The Super-Earths course instructors were decidedly more active within discussion forums than in the other two courses in this study.

Instructor-paced. There was a total of 18,202 student posts in this course. Most of these posts (81 percent, or 14,791) were comments. Of these, slightly more than 59 percent were response posts. The other 3,411 posts in this course were initial posts. Eighty-four percent of these were discussion posts. Students were more likely (64 percent) to respond to a post that was initially posted by an instructor and more likely (almost 80 percent) to respond to a student post generally. These indicators suggest that the instructor in this course version was guiding and facilitating the discussion. However, while students were more likely to respond to an instructor-created thread, they were responding to each other at a high rate.

Self-paced. The self-paced version of the Super-Earths course had 1,866 student posts. Similar to the instructor-paced version, the majority (nearly 82 percent) of the posts were made as comments instead of comment threads. Also like the instructor-paced version, most of these posts (about 42 percent) were made as response posts. When students made initial posts, they marked most (81 percent) as discussions. Students were more likely (70 percent) to respond to an original post made by the instructor and more likely (79 percent) to respond to another student's post generally. These indicators suggest that despite the self-paced pacing condition, there was still a strong instructor presence in the guidance and facilitation of the self-paced course discussions.

Table 12. Classification of Super-Earths and Life Student Posts, by Pacing Condition

Pacing Condition	Post Type (% of Total Posts)	Post Descriptor	Thread Descriptor	Original Poster (% of Total Posts)	Responded to (% of Total Posts)

		(% of Total Posts)	(% of Total Posts)		
Instructor-Paced	Comments (81.3%)	Response Post (59.2%)	Discussion (83.6%)	Instructor (64.3%)	Students (79.1%)
Self-Paced	Comments (81.6%)	Response Post (41.6%)	Discussion (80.8%)	Instructor (69.6%)	Students (78.7%)

Note: Percentages reflect those expressed within each classification.

Visualizing Japan

In the Visualizing Japan (1850s–1930s: Westernization, Protest, Modernity) course, historians from both the Massachusetts Institute of Technology and Harvard provided students with an examination of the westernization of Japanese history, looking at the time period between the 1850s and 1930s. The course was organized into four modules that focused primarily on images and the ways they can be used to explain and explore this time period. These modules covered Westernization, the expedition to Japan in 1853–1854 by Commodore Perry, the 1905 Hibiya Riots in Tokyo, and the impact of modernity. Of the three courses studied, this course had the most inconsistent classification breakdowns between the pacing conditions (Table 13). The self-paced version of this course is the only one that had more comment threads than comments and more student posts than were logged in the instructor-paced version.

Instructor-paced. There was a total of 1,604 student posts in this course. Most of these posts (85 percent) were comments. Of these, nearly 62 percent were response posts. The other 235 posts in this version of the course were initial posts. Ninety-four percent of those posts were discussions. Students were more likely (90 percent) to respond to an initial post made by an instructor and more likely (97 percent) to respond to a student post generally. Similar to the Super-Earths course, these indicators show that discussions in this version of the course were guided by instructors and that students engaged with each other in response to these instructor posts.

Self-paced. The self-paced version of the Visualizing Japan course had 1,654 student posts. While there were more comments made in the instructor-paced version of this offering, a higher percentage (about 69 percent) of student posts in the self-paced course were comment threads; most of these were classified as initial posts. These comment threads were overwhelming classified (96 percent) as discussions versus questions. Not surprisingly, an equally overwhelming percentage of the responses (88 percent) were made to student-originated posts and were classified (63 percent) as student replies in this version of the course. These indicators show that instructor presence was lower in the self-paced version of Visualizing Japan and that students had more influence over the direction of the forum discussions.

Table 13. Classification of Visualizing Japan Student Posts, by Pacing Condition

Pacing Condition	Post Type (% of Total Posts)	Post Descriptor (% of Total Posts)	Thread Descriptor (% of Total Posts)	Original Poster (% of Total Posts)	Responded to (% of Total Posts)
Instructor-Paced	Comments (85.4%)	Response Post (61.5%)	Discussion (94%)	Instructor (90.4%)	Students (97.1%)
Self-Paced	Comment Threads (68.7%)	Initial Post (68.7%)	Discussion (96.2%)	Students (87.6%)	Students (63.4%)

Note: Percentages reflect those expressed within each classification.

Summary

For the Ancient Greek Hero and Super-Earths courses, the classifications were consistent between the two pacing conditions. In these courses, the instructor-paced version appeared to have more instructor-guided discussions, while the self-paced course appeared to have more student-guided discussions. The outlier was the Visualizing Japan course, where the classifications were not as consistent. Interestingly, the self-paced Visualizing Japan course is the only self-paced course studied that had more student posts than the course's instructor-paced version, although the difference is small (1,654 as compared to 1,604). There were two classification variables that were consistent across all six course studied: most comment threads were identified as discussions, as opposed to questions, and students in all six offerings replied more to student posts than to instructor posts (a possible indicator of student-to-student interaction). Overall, few posts (just 1.23 percent of the total discussion posts) were endorsed, with endorsements being observed in the instructor-paced Ancient Greek Hero course (407 endorsed posts, or 1.84 percent of comment posts) and the instructor-paced Super-Earths course (112 endorsed posts, or 0.76 percent). No posts in the Visualizing Japan courses were endorsed. This outcome can likely be attributed to the lower teacher presence in the self-paced versions of the Ancient Greek Hero and Super-Earths courses.

Research Questions (RQs) and Hypotheses (Hs)

In this section, I revisit each of my research questions and their associated hypotheses, reporting the experimental findings for each of the six courses studied.

Research Question 1 (RQ1) and Hypothesis 1 (H1)

My first research question (RQ1) is “How does cognitive processing vary between pacing conditions?”, and the associated hypothesis (H1) is that pacing condition directly affects cognitive processing. I expected to find that students in the self-paced courses would have higher average cognitive processing scores than students in the instructor-paced courses. This expectation is based on the work by Kanuka and Garrison (2004), who posit that cognitive presence requires sustained communication and student-student and student-content interaction. Because I expected that students in self-paced courses would have more time to read and consider peer discussion posts than students in instructor-paced courses, I expected to find that self-paced students would show higher levels of cognitive presence than students in instructor-paced courses. I also expected to find that students in the self-paced courses would make longer posts and demonstrate higher levels of cognitive processing. Following this same logic, I also expected to find that the additional time spent on considering peers’ posts would result in longer threads within the self-paced courses. Despite a lower total number of students in the self-paced courses studied, I expected that there would still be activity within the discussion forums of these courses (Campbell et al., 2014) and that the additional time available to consider peer posts would lead to more opportunities for students to engage in knowledge construction with each other, which, in turn, would be reflected in higher levels of cognitive presence (Kent et al., 2016).

I utilized Levene's robust test statistic for equality of variances, with most of the group means for the Ancient Greek Hero and Super-Earths courses having an unequal variance and most of the Visualizing Japan courses having an equal variance. To test this hypothesis, I conducted an independent samples t-test to compare the cognitive processing (and sub-parts) within each pacing condition. For any comparisons resulting in a significant difference, I ran a post-hoc analysis to standardize the effect size, Cohen's *d*, and reported the practical significance or effect size. Cohen, in his seminal work (1988), provides the interpretations of *d* as follows: for a value of less than 0.2, there is little to no difference between the two groups compared; for a value of between 0.2 and 0.5, the difference between the groups is small; for a value of between 0.5 and 0.8, the difference is moderate; for a value of greater than 0.8, the two groups have considerable difference between them. These interpretations were revised by Sawilowsky (2009) to provide more detail and are shown in Table 14. I used Sawilowsky's interpretations in my reporting of the effect sizes.

Table 14. Levels of Practical Significance. Adapted from Sawilowsky (2009)

Value of *d*	Effect Size Description
0.01	Very small
0.20	Small
0.50	Medium
0.80	Large
1.2	Very large
2.0	Huge

Even with a very small to small practical significance, there is a difference between groups compared; it is just not something that can be observed by the naked eye (Cohen, 1988, 1992; Sawilowsky, 2009). I reported each of the results based on the cognitive processing and associated sub-parts and reported the summary of effect sizes in Table 15; the full table is set out in Appendix A.

Cognitive processing. In the Ancient Greek Hero courses, students in the instructor-paced version (M = 11.72, SD = 8.56) demonstrated slightly lower levels of cognitive processing than students in the self-paced version (M = 11.88, SD = 7.42). The difference between the two groups (-.16) was not significant$t$(5721.14) = -1.3032, p > .05. In the Super-Earths courses, students in the instructor-paced version (M = 13.42, SD = 14.16) demonstrated higher levels of cognitive processing than students in the self-paced version (M = 11.54, SD = 8.42). The difference between the two groups (1.88) was significant, t(3076.15) = 8.4693, p < 0.001, though it had a very small to small practical significance (d = .14). In the Visualizing Japan courses, while students in the instructor-paced version (M = 12.84, SD = 6.87) demonstrated slightly higher levels of cognitive processing than students in the self-paced version (*M* = 12.65, *SD* = 7.21), the difference between the two groups (.19) was not significant, $t(3256) = 0.7684$, $p > .05$.

Insight. In the Ancient Greek Hero courses, students in the instructor-paced version (*M* = 3.34, *SD* = 4.17) demonstrated slightly lower levels of insight than students in the self-paced version (*M* = 3.55, *SD* = 3.96). The difference between the two groups (-.21) was significant, $t(34322) = -3.0471$, $p < .01$, though it had a very small to small practical significance ($d = .05$). In the Super-Earths courses, students in the instructor-paced version (*M* = 5.07, *SD* = 13.16) demonstrated higher levels of insight than students in the self-paced version (*M* = 3.14, *SD* = 4.01). The difference between the two groups (1.92) was significant, $t(7334.33) = 14.2825$, $p < .001$, but it had a very small to small practical significance ($d = .15$). In the Visualizing Japan courses, students in the instructor-paced version (*M* = 3.98, *SD* = 4.05) demonstrated slightly higher levels of insight than the students in the self-paced version (*M* = 3.92, *SD* = 3.66), but the difference between the two groups (.06) was not significant, $t(3256) = 0.4778$, $p > .05$.

Causation. In the Ancient Greek Hero courses, students in the instructor-paced version (*M* = 1.31, *SD* = 2.25) demonstrated slightly lower levels of causation than students in the self-paced course (*M* = 1.43, *SD* = 2.14). The difference between the two groups (-.11) was significant, $t(34322) = -3.0399$, $p < .01$, Though it had a very small to small practical significance ($d = .05$). In the Super-Earths courses, students in the instructor-paced version (*M* = 1.35, *SD* = 2.66) demonstrated slightly lower levels of causation than students in the self-paced version (*M* = 1.58, *SD* = 2.73). The difference between the two groups (-.24) was significant, $t(2244.83) = -3.6180$, $p < .01$, though it had a very small to small practical significance ($d = .1$). In the Visualizing Japan courses, students in the instructor-paced version (M = 2.07, SD = 2.87) demonstrated slightly higher levels of causation than the students in the self-paced version (M = 1.95, SD = 2.56), but the difference between the two groups (.11) was not significant, $t(3256) = 1.2049$, $p > .05$.

Discrepancy. In the Ancient Greek Hero courses, students in the instructor-paced version (M = 1.41, SD = 2.59) demonstrated slightly higher levels of discrepancy than students in the self-paced version (M = 1.35, SD = 2.08). The difference between the two groups (.06) was not significant, $t(6006.47) = 1.6888$, $p > .05$. In the Super-Earths courses, students in the instructor-paced version (M = 1.33, SD = 2.67) demonstrated slightly lower levels of discrepancy than students in the self-paced version (M = 1.67, SD = 2.50). The difference between the two groups (-.34) was significant, $t(2321.35) = -5.6330$, $p < .001$. However, while statistically significant, this difference had only a very small to small practical significance (d = .13). In the Visualizing Japan courses, students in the instructor-paced version (M = 1.49, SD = 2.21) demonstrated slightly higher levels of discrepancy than the students in the self-paced version (M = 1.29, SD = 1.99). The difference between the two groups (.20) was significant, $t(3256) = 2.7342$, $p < .01$, though it had a very small to small practical significance (d = .10).

Tentativeness. In the Ancient Greek Hero courses, students in the instructor-paced version (M = 2.68, SD = 3.62) demonstrated slightly lower levels of tentativeness than students in the self-paced version (M = 2.73, SD = 3.48). The difference between the two groups (-.04) was not significant, $t(5398.2) = -0.7464$, $p > .05$. In the Super-Earths courses, students in the instructor-paced version (M = 2.49, SD = 3.81) demonstrated slightly lower levels of tentativeness than students in the self-paced version (M = 3.01, SD = 4.28). The difference between the two groups (-.51) was significant, $t(20066) = -5.4661$, $p < .001$, though it had a very small to small practical significance (d = .13). In the Visualizing Japan courses, students in the instructor-paced version (M = 2.85, SD = 3.18) demonstrated slightly higher levels of tentativeness than the students in the self-paced version (M = 2.80, SD = 3.28). The difference between the two groups (.04) was not significant, $t(3256) = 0.3803$, $p > .05$.

Certainty. In the Ancient Greek Hero courses, students in the instructor-paced version (M = 1.75, SD = 4.03) demonstrated slightly higher levels of certainty than students in the self-paced version (M = 1.45, SD = 2.69). The difference between the two groups (.30) was significant, $t(6923.31) = 6.3145$, $p < .001$, though it had very small to small practical significance (d = .08). In the Super-Earths courses, students in the instructor-paced version (M = 2.20, SD = 3.92) demonstrated slightly higher levels of certainty than students in the self-paced version (M = 1.48, SD = 2.69). The difference between the two groups (.72) was significant, $t(2747.54) = 10.4070$, $p < .001$, though it had a small practical significance (d = .20). In the Visualizing Japan courses, students in the instructor-paced version (M = 1.29, SD = 2.28) demonstrated slightly lower levels of certainty than the students in the self-paced version (M = 1.36, SD = 3.58). The difference between the two groups (-.06) was not significant, $t(3256) = -0.5610$, $p > .05$.

Differentiation. In the Ancient Greek Hero courses, students in the instructor-paced version (M = 2.82, SD = 3.54) demonstrated slightly lower levels of differentiation than students in the self-paced version (M = 3.02, SD = 3.25). The difference between the two groups (-.20) was significant, $t(5539.49) = -3.7577$, $p < .001$, though it had a very small to small practical significance (d = .06). In the Super-Earths courses, students in the instructor-paced version (M = 2.35, SD = 3.39) demonstrated slightly higher levels of differentiation than students in the self-paced version (*M* = 2.43, *SD* = 3.24). The difference between the two groups (-.08) was not significant, $t(20066) = -0.97514$, $p > .05$. In the Visualizing Japan courses, students in the instructor-paced version (*M* = 2.65, *SD* = 2.96) demonstrated slightly lower levels of differentiation than students in the self-paced version (*M* = 2.77, *SD* = 3.15). The difference between the two groups (-.11) was not significant, $t(3256) = -1.0677$, $p > .05$.

Table 15. Effect Sizes of Dependent Variables (DVs), by Course

DV	Ancient Greek Hero	Super-Earths	Visualizing Japan
Cognitive Processing	Self-paced (+0.16)	Instructor-paced (+1.88)	Instructor-paced (+0.19)
Effect Size	-	**Very small to small**	-
Insight	Self-paced (+0.21)	Instructor-paced (+1.93)	Instructor-paced (+0.06)
Effect Size	**Very small to small**	**Very small to small**	-
Causation	Self-paced (+0.05)	Self-paced (+0.23)	Instructor-paced (+0.12)
Effect Size	**Very small to small**	**Very small to small**	-
Discrepancy	Instructor-paced (+0.06)	Self-paced (+0.34)	Instructor-paced (+0.20)
Effect Size	-	**Very small to small**	**Very small to small**
Tentativeness	Self-paced (+0.05)	Self-paced (+0.52)	Instructor-paced (+0.05)
Effect Size	-	**Very small to small**	-
Certainty	Instructor-paced (+0.30)	Instructor-paced (+0.72)	Self-paced (+0.07)
Effect Size	**Very small to small**	**Very small to small**	-
Differentiation	Self-paced (+0.20)	Self-paced (+0.08)	Self-paced (+0.12)
Effect Size	**Very small to small**	-	-

Note: Pacing condition with higher mean shown in table, with difference in means shown in parentheses.

Summary. In the Ancient Greek Hero courses, three of the variables (cognitive processing, discrepancy, and tentativeness) showed no statistical difference between the pacing conditions. The Super-Earths courses had statistically significant differences for each dependent variable except for differentiation, but there was not consistency between which pacing condition was higher for a specific variable. Once again, the Visualizing Japan courses were outliers, resulting in only one statistically significant variable (discrepancy).

Overall, the statistical significance is influenced by the large sample size, and the practical significance was reported for any variables that were statistically significant. In looking through the practical significance levels, the levels ranged from very small to small, suggesting that the pacing condition, by itself, had a small impact on cognitive processing (or its associated sub-parts) in the courses, though it had an impact nonetheless.

Research Question 2 (RQ2) and Hypothesis 2 (H2)

My second research question is "In what ways does pacing condition affect discussion forum activity and engagement (e.g., number of replies, number of words in each post)?", and the associated hypothesis (H2) is that pacing condition affects discussion forum activity and engagement. I expected to find that students in the instructor-paced courses would post more frequently and have more replies than students in the self-paced courses. I based this expectation on a review of the relevant literature (e.g. Jordan, 2015; Sharif & Magrill, 2015), which stresses the influence that the cohort nature of a MOOC can have on discussion forum activity. Research suggests that because students in instructor-paced courses have a cohort, there will be timely replies to posts by peers, leading to more engagement and activity in discussion forums. The presence of a cohort can also lead to the development of the sub-communities that Yang et al.(2014) believe demonstrate evidence of the two-way influence found in discussion forums. I also expected to find that students in the self-paced courses would make longer posts (word count and thread length) than students in the instructor-paced courses. I expected to find this because students in the self-paced versions would have more time to consider and interact with peer posts as well as the course content.

I again used Levene's test statistic for equal variances and found that most of the group means had unequal variances. To test this hypothesis, I used a two-tailed independent samples t-test to compare the post characteristics within each pacing condition. For any t-test that showed a significant difference, I ran a post-hoc analysis to standardize the effect size, Cohen's *d*, and reported the practical significance. I reported each of the results based on the post characteristics and summarized them in Table 16; the full table is set out in Appendix A.

Post position. In the Ancient Greek Hero courses, students in the instructor-paced version (M = 27.43, SD = 61.97) had longer post threads than students in the self-paced version (M = 20.07, SD = 29.66). The difference between the two groups (7.35) was significant, $t(9648.14) = 11.3439$, $p < .001$, though it had a very small to small practical significance (d = .13). In the Super-Earths courses, students in the instructor-paced version (M = 934.34, SD = 1397.69) had much longer post threads than students in the self-paced version (M = 197.49, SD = 229.02). The difference between the two groups (736.85) was significant, $t(14155.6) = 57.0996$, $p < .001$. A unique feature of the Super-Earths discussion forums was that students were asked to solve specific problems that were presented to them in the form of instructor-initiated posts. The large thread length seen in these courses resulted not only from large enrollment numbers, but also from the fact that students provided answers to the same initial (problem-solution-seeking) post from the instructor. Also contributing to the larger thread length in the Super-Earths courses was the fact that students were asked to post introductions in the discussion forums, and these introductions required them to share a fair amount of information, including their reasons for taking the course and their goals for the course. The difference in thread lengths between the two versions of the Super-Earths course had a medium practical significance (d = .60). In the Visualizing Japan courses, students in the instructor-paced version (M = 40.94, SD = 39.16) had much longer post threads than the students in the self-paced version (M = 3.40, SD = 5.68). The difference between the two groups (37.54) was significant, $t(1512.22) = 34.5137$, $p < .001$, and it had a large to very large practical significance (d = 1.12).

Number of comments. In the Ancient Greek Hero courses, students in the instructor-paced version (M = 2.14, SD = 7.33) received fewer comments on their posts than students in the self-paced version (M = 3.56, SD = 5.61). The difference between the two groups (-1.42) was significant, $t(8646) = -4.4094$, $p < .001$, though the practical significance was very small to small (d = .10). In the Super-Earths courses, students in the instructor-paced version (M = 1.52, SD = 4.58) received slightly more comments on their posts than students in the self-paced version (M = 1.34, SD = 1.79). The difference between the two groups (.20) was not significant, $t(3752) = 0.8074$, $p > 0.05$. In the Visualizing Japan courses, students in the instructor-paced version (M = 0.56, SD = 0.98) received slightly more comments on their posts than students in the self-paced version (M = 0.39, SD = 1.18). The difference (.16) was significant, $t(1370) = 1.9845$, $p < 0.05$, though it had a very small to small practical significance (d = .11).

Word count. In the Ancient Greek Hero courses, students in the instructor-paced version (M = 78.39, SD = 125.75) made posts using fewer words than students in the self-paced version (*M* = 84.04, *SD* = 99.92). The difference between the two groups (-5.65) was significant, *t*(6048.45) = -3.2902, *p* < .001. While a difference in word count of five more words, as was the case here, is statistically significant, it represents a very small to small practical significance (*d* = .05). In the Super-Earths courses, students in the instructor-paced version (*M* = 47.44, *SD* = 69.67) made posts using fewer words than students in the self-paced version (*M* = 54.77, *SD* = 68.22). The difference between the two groups (-7.33) was significant, *t*(20066) = -4.3393, *p* < .001. The practical significance of this difference, however, was very small to small (*d* = .11). In the Visualizing Japan courses, students in the instructor-paced version (*M* = 60.00, *SD* = 50.01) made posts using more words than the students in the self-paced version (*M* = 53.00, *SD* = 42.07). The difference between the two groups (7) was significant, *t*(3129.12) = 4.3158, *p* < .001, though it had a small practical significance (*d* = .20).

Words per sentence. In the Ancient Greek Hero courses, students in the instructor-paced version (*M* = 15.64, *SD* = 11.67) made posts using fewer words per sentence than students in the self-paced version (*M* = 17.12, *SD* = 9.85). The difference between the two groups (-1.48) was significant, *t*(5812.43) = 8.8223, *p* < .001, though it represented a very small to small practical significance (*d* = .12). In the Super-Earths courses, students in the instructor-paced version (*M* = 12.57, *SD* = 9.13) made posts using fewer words per sentence than students in the self-paced version (*M* = 14.38, *SD* = 9.92). The difference between the two groups (-1.81) was significant, *t*(2201.47) = -7.5818, *p* < .001, though it represented a small practical significance (*d* = .20). In the Visualizing Japan courses, students in the instructor-paced version (*M* = 19.98, *SD* = 10.50) made posts using slightly more words per sentence than the students in the self-paced version (*M* = 19.91, *SD* = 9.66). The difference between the two groups (.07) was not significant, *t*(3256) = 0.1961, *p* > 0.05.

Number of six-letter words. In the Ancient Greek Hero courses, students in the instructor-paced version (*M* = 18.74, *SD* = 10.34) made posts using slightly fewer six-letter words than students in the self-paced version (*M* = 19.54, *SD* = 8.90). The difference between the two groups (-.79) was significant, *t*(5745.67) = -5.2862, *p* <.001, though it had a very small to small practical significance (*d* = .10). In the Super-Earths courses, students in the instructor-paced version (*M* = 20.29, *SD* = 14.84) made posts using slightly more six-letter words than students in the self-paced version (*M* = 20.25, *SD* = 10.11). The difference between the two groups (.049) was not significant, *t*(2766.78) = 0.1830, *p* > .05. In the Visualizing Japan courses, students in the instructor-paced version (*M* = 24.25, *SD* = 9.69) made posts using slightly fewer six-letter words than the students in the self-paced version (*M* = 24.85, *SD* = 9.71). The difference between the two groups (-.60) was not significant, *t*(3256) = -1.7629, *p* > 0.05.

Analytical thinking. In the Ancient Greek Hero courses, students in the instructor-paced version (*M* = 62.55, *SD* = 30.97) made posts that were lower in the analytical thinking score than students in the self-paced version (*M* = 66.19, *SD* = 28.95). The difference between the two groups (-3.65) was significant, *t*(5483.4) = -7.5139, *p* < .001, though it had a very small to small practical significance (*d* = .12). In the Super-Earths courses, students in the instructor-paced version (*M* = 54.18, *SD* = 32.07) made posts that were less analytical than students in the self-paced version (*M* = 59.33, *SD* = 30.44). The difference between the two groups (-5.15) was significant, *t*(2310.69) = -6.9285, *p* < .001, though the practical significance of this difference was small (*d* = .20). In the Visualizing Japan courses, students in the instructor-paced version (M = 72.86, SD = 26.46) made posts that were slightly less analytical than students in the self-paced version (M = 74.59, SD = 26.14). The difference between the two groups (-1.74) was not significant, $t(3256) = -1.8841$, $p > 0.05$.

Table 16. Effect Size of Post Characteristics, by Course

Characteristic	Ancient Greek Hero	Super-Earths	Visualizing Japan
Post Position	Instructor-paced (+7.36)	Instructor-paced (+736.85)	Instructor-paced (+37.54)
Effect Size	**Very small to small**	**Medium**	**Large to very large**
# of Comments	Self-paced (+1.42)	Instructor-paced (+0.18)	Instructor-paced (+0.17)
Effect Size	**Very small to small**	**-**	**Very small to small**
Word Count	Self-paced (+5.65)	Self-paced (+7.33)	Instructor-paced (+7.00)
Effect Size	**Very small to small**	**Very small to small**	**Small**
Words per Sentence	Self-paced (+1.48)	Self-paced (+1.81)	Instructor-paced (+0.07)
Effect Size	**Very small to small**	**Small**	-
# of Six-Letter Words	Self-paced (+0.80)	Instructor-paced (+0.04)	Self-paced (+0.60)
Effect Size	**Very small to small**	-	-
Analytical Thinking	Instructor-paced (+3.64)	Self-paced (+5.15)	Self-paced (+1.73)
Effect Size	**Very small to small**	**Small**	-

Note: Pacing condition with higher mean shown in table, with difference in means shown in parentheses.

Summary of findings. The findings detailed above suggest that pacing condition affects post characteristics, though it is not clear which pacing condition is more effective because some characteristics are supported more in the instructor-paced condition (e.g., post position and number of comments), while other characteristics are supported more in the self-paced condition (e.g., word count, words per sentence, and number of six-letter words). The instructor-paced versions of all three of the courses studied had considerably longer posts in terms of thread length than their self-paced counterparts. This can likely be attributed to larger enrollments and longer introductory posts. For the Ancient Greek Hero course, excluding the post position variable, the self-paced version reported higher on all measures when compared to the instructor-paced version. But while there were statistically significant differences between the two groups, the practical significance of these differences was relatively small. Conversely, in the Visualizing Japan course, results for all variables except the number of six-letter words and analytical thinking scores were statistically significant, with the instructor-paced version of the course having the higher respective scores in comparison to the self-paced version. The Super-Earths course, while not displaying a consistent pattern across results, did report more characteristics (word count, words per sentence, and analytical thinking score) with higher scores in the self-paced versus the instructor-paced version. The words per sentence and analytical thinking scores were considerably higher for the self-paced course. Interestingly, all three courses had higher analytical thinking scores for the self-paced versions (although the difference was not significant in the Visualizing Japan course), which suggests that students in the self-paced versions were making high-quality posts.

Research Question 3 (RQ3) and Hypothesis 3 (H3)

My third research question is "What aspects of forum activity are predictive of cognitive processing, and do these relationships exist in both pacing conditions?", and the associated hypothesis (H3) is that discussion forum activity and post characteristics affect cognitive processing. I expected to find that students in self-paced courses would have posts that have higher levels of insight, certainty, and differentiation than students in instructor-paced courses. Chiu and Hew (2018) hold the view that the act of making a post utilizes higher levels of cognitive processing because the poster must not only read, but also must understand, previous posts and/or the content to which he or she is responding. I also expected to find that longer (in terms of word count) posts would have higher levels of cognitive processing. Due to the clustering effect of students making multiple posts within a course and, in some cases, students posting to discussion forums in both course versions, I used a random-intercepts model, also known as a multi-level mixed-effects linear regression. I ran a separate model for each dependent variable and course, and I report the findings, based on dependent variables, below. To prepare my models, I dummy-coded all categorical variables. Since I am studying individual student posts, I clustered observations by student. Because some variables, specifically, "original poster", "responded to", "endorsed", and thread descriptor, were not applicable for every post, I ran three separate models. To provide more interpretable results, I have reported word count by multiples of 100, and words per sentence and six-letter words by multiples of 10.

Cognitive processing. The "cognitive processing" word category in the LIWC text analysis tool is made up of 797 words (e.g., since, seem, sort, sense, wish, somewhat) with sub-parts (insight, causation, differentiation, certainty, and tentativeness, and discrepancies). Results of the regressions for each of the three courses appears in Table 17.

Table 17. Unstandardized Regression Coefficients Estimating Cognitive Processing

	Hero Courses Coefficient	Super-Earths Courses Coefficient	Visualizing Japan Courses Coefficient
Intercept	13.68***	3.69***	16.06***
(SE)	0.58	0.72	0.64
Self-Paced	0.33	-1.92***	0.18
(SE)	0.22	0.55	0.32
Post Descriptor	0.62***	2.80***	-0.25
(SE)	0.17	0.23	0.33
Post Type	-0.83***	-[c]	-[c]
(SE)	0.19	-	-
Word Count	0.27***	0.62***	0.27
(SE)	0.05	0.10	0.20
Analytical Thinking	-0.08***	-0.06***	-0.09***
(SE)	0.00	0.00	0.01
Words per Sentence	0.70***	0.14	0.54***
(SE)	0.11	0.14	0.14
Six-Letter Words	1.15*	5.67***	0.69***
(SE)	0.34	0.40	0.20

Note: **p* < .05; ***p* < .01; ****p* < .001; [c] = variable dropped because of collinearity.

Pacing condition. The Ancient Greek courses had 4,156 clusters of student observations in the initial model. The Super-Earths courses had 8,486 clusters of student observations in the initial model. The pacing condition in the Super-Earths courses was negatively associated with cognitive processing, where students in the self-paced version had posts approximately 1.92 percentage points lower for cognitive processing than those in the instructor-paced version, while holding other variables constant. The Visualizing Japan courses had 879 clusters of student observations in the initial model.

Post descriptor. In the Ancient Greek Hero courses, I found that, when compared to either an initial post or a response post, a comment as the post descriptor was positively associated with cognitive processing. I expected that this would account for cognitive processing scores that were about 0.62 percentage points higher, while holding other variables constant. In the Super-Earths courses, I found that, when compared to either a response post or a comment post, an initial post as the post descriptor was positively associated with cognitive processing and accounted for an expected 2.80 percentage points increase, while holding other variables constant.

Post type. In the Ancient Greek Hero courses, the post type was negatively associated with cognitive processing score, as post types marked as comments was expected to be approximately 0.83 percentage points lower for the cognitive processing value than those marked as comment threads, while holding other variables constant.

Word count. In both Ancient Greek Hero course versions, for every additional 100 words used in a post, I expected to see an approximately 0.27 percentage point increase in cognitive processing, while holding other variables constant. For the Super-Earths courses, for that same additional 100 words measure, I expected to see an approximately 0.62 percentage point increase, while holding other variables constant. Between the three courses studied, word count had the greatest expected influence on cognitive processing in the Super-Earths courses.

Analytical thinking. In all three courses studied, the analytical thinking scores were negatively associated with cognitive processing, while holding other variables constant. I expected that, for every one-unit increase in the analytical thinking score, the cognitive processing score would be lowered, albeit by a very small amount (between 0.06 and 0.09 units).

Words per sentence. In the Ancient Greek Hero courses, for every 10 words added to a sentence in a post, I expected to see an increase of approximately 0.70 percentage points. That expected association was slightly lower (but still positive) in the Visualizing Japan courses, where the increase was approximately 0.54 percentage points in cognitive processing for every additional 10 words added to a post sentence, while holding other variables constant.

Six-letter words. For all three courses studied, the number of six-letter words used in a post increased the cognitive processing score, with every ten additional six-letter words adding an expected 1.1 percentage points (Ancient Greek Hero), 5.7 percentage points (Super-Earths), and 0.50 percentage points (Visualizing Japan) to the cognitive processing scores, while holding other variables constant. I would expect that adding additional six-letter words to a post would have the most influence in the Super-Earths courses.

Additional characteristics. To analyze the influence of other variables, such as original poster, responded to, and thread descriptor (question or discussion), I ran additional models for each of the three courses studied. In the Ancient Greek Hero courses, endorsed posts were positively associated with cognitive processing, and I expected that an endorsed post would have a cognitive processing score that was approximately 0.85 percentage points higher than a post that was not endorsed, while holding other variables constant. Interestingly, the opposite was true for the Super-Earths courses, where endorsed posts were negatively associated with cognitive processing by an expected identical 0.85 percentage points, while holding other variables constant. This difference was even more significant with the Visualizing Japan courses, where an endorsed post added an expected approximately 4.14 percentage points to the cognitive processing score, while holding other variables constant. In the Ancient Greek Hero courses, thread descriptor was negatively associated with cognitive processing and posts marked as discussions had cognitive processing scores an expected approximately 1.53 percentage points lower than discussions marked as questions, while holding other variables constant. That same negative association was found in the Super-Earths courses, where the discussions were expected to score approximately 2.24 percentage points lower for cognitive processing, while holding other variables constant. In the Super-Earths courses, for posts made by students who responded to a post that was originally posted by a student, I expected such posts to be approximately 1.39 percentage points higher for cognitive processing. In this course, the posts of students who responded to peer reply posts were approximately 0.41 percentage points higher for cognitive processing, while holding other variables constant. No other predictors were significant.

Insight. The “insight” sub-category of the “cognitive processing” category in the LIWC text analysis tool contains 259 words (e.g., think, realize, question, accept, notice, and perspective). A summary of results for each course appears in Table 18.

Table 18. Unstandardized Regression Coefficients Estimating Insight

	Hero Courses Coefficient	Super-Earths Courses Coefficient	Visualizing Japan Courses Coefficient
Intercept	3.09***	-3.82***	4.07***
(SE)	0.29	0.58	0.39
Self-Paced	0.21*	-1.57***	-0.18
(SE)	0.10	0.00	0.20
Post Descriptor	0.20*	0.21	-0.04
(SE)	0.07	0.16	0.22
Post Type	-0.12	-[c]	-[c]
(SE)	0.09	-	-
Word Count	-0.02	-0.88***	-0.64***
(SE)	0.02	0.11	0.12
Analytical Thinking	-0.01***	-0.01***	-0.02***
(SE)	0.00	0.00	0.00
Words Per Sentence	0.20***	-1.20***	0.10
(SE)	0.04	0.20	0.10
Six Letter Words	0.50*	5.62***	0.60***
(SE)	0.20	0.40	0.20

Note: *$p < .05$; **$p < .01$; ***$p < .001$; [c] = variable dropped because of collinearity.

Pacing condition. In the Ancient Greek Hero courses, pacing condition was positively associated with insight. I expected student posts in the self-paced version to score approximately 0.21 percentage points higher for insight than posts in the instructor-paced version, while holding other variables constant. In the Super-Earths courses, pacing condition was negatively associated with insight. I expected posts by students in the self-paced version to score approximately 1.57 percentage points lower for insight than those in the instructor-paced version, while holding other variables constant.

Post descriptor. In the Ancient Greek Hero courses, I found that, when compared to either an initial post or a response post, a comment as the post descriptor was positively associated with insight and accounted for an expected increase of about 0.62 percentage points, while holding other variables constant.

Word count. In the Super-Earths courses, I expected that, for every 100 words added to a post, there would be an expected approximately 0.88 percentage point decrease, while holding other variables constant. In the Visualizing Japan courses, I expected that, for every 100 words added to a post, there would be approximately 0.64 percentage point decrease, while holding other variables constant. As the posts in these courses get longer, students were using fewer words indicative of insight.

Analytical thinking. In all three courses studied, analytical thinking was negatively associated with insight, with a one-unit increase in the analytical thinking score slightly lowering the insight score, while holding other variables constant.

Words per sentence. In the Ancient Greek Hero courses, I found that words per sentence was positively associated with insight and that, for every ten additional words added to a post, there would be an expected approximately 0.20 percentage point increase in the insight score, while holding other variables constant. The expected influence was approximately six times greater in the Super-Earths courses, albeit in the negative direction. In those courses, for every additional ten words added to a post, I expected that there would be an approximately 1.2 percentage point decrease in the insight score, while holding other variables constant.

Six-letter words. In all three courses studied, I found that the number of six-letter words used in a post was positively associated with insight, while holding other variables constant. The expected influence of six-letter words was highest in the Super-Earths courses. In those courses, for every ten additional six-letter words added to a post, I expected that the insight score would be approximately 5.62 percentage points higher, while holding other variables constant.

Additional characteristics. As I did with cognitive processing, I ran other models to explore additional characteristics that may be predictors for insight. In the Ancient Greek Hero courses, I found that the “responded to” role was negatively associated with insight, as the posts of students who responded to peer posts scored approximately 0.30 percentage points lower for insight than the posts of students who responded to instructor posts, while holding other variables constant. In the Super-Earths courses, I found that the thread descriptor was negatively associated with insight and that posts marked as discussions had insight scores approximately 0.65 percentage points lower than posts marked as questions, while holding other variables constant. No other predictors were found to be significant.

Causation. The “causation” sub-category of the “cognitive processing” category in the LIWC text analysis tool contains 135 (e.g., because, effect, hence, affect, based, create). A summary of findings appears in Table 19.

Table 19. Unstandardized Regression Coefficients Estimating Causation

	Hero Courses Coefficient	Super-Earths Courses Coefficient	Visualizing Japan Courses Coefficient
Intercept	1.30***	0.41***	2.02***
(SE)	0.07	0.07	0.26
Self-Paced	0.13*	0.06	-0.12
(SE)	0.05	0.13	0.15
Post Descriptor	0.13***	0.46***	-0.00
(SE)	0.03	0.06	0.15
Post Type	-0.32***	-[c]	-[c]
(SE)	0.04	-	-
Word Count	0.07***	0.68***	-0.08
(SE)	0.01	0.05	0.08
Analytical Thinking	-0.01***	-0.00	-0.01***
(SE)	0.00	0.00	0.00
Words Per Sentence	0.20***	0.35***	0.20***
(SE)	0.02	0.04	0.05
Six Letter Words	0.11***	0.02	0.20*
(SE)	0.02	0.01	0.08

Note: *p < .05; **p < .01; ***p < .001; [c] = variable dropped because of collinearity.

Pacing condition. In the Ancient Greek Hero courses, pacing condition was positively associated with causation. I expected the posts of students in the self-paced course version to score approximately 0.13 percentage points higher for causation than student posts in the instructor-paced version, while holding other variables constant.

Post descriptor. In the Ancient Greek Hero courses, I found that, when compared to either an initial post or a response post, a comment as the post descriptor was positively associated with causation and accounted for an expected approximately 0.13 percentage-point increase, while holding other variables constant. In the Super-Earths courses, posts that were made initial posts, when compared to response posts or comment posts, were positively associated with causation and were expected to score approximately 0.46 percentage points higher, while holding other variables constant.

Post type. In the Ancient Greek Hero courses, the post type was negatively associated with the causation score. I expected that a post type marked as a comment would score approximately 0.32 percentage points lower than one classified as a comment thread, while holding other variables constant.

Word count. In the Ancient Greek Hero courses, I expected that, for every additional 100 words added to a post, the insight scores for the post would be approximately 0.07 percentage points higher, while holding other variables constant. This expected influence was again considerably higher with the Super-Earths courses. In those courses, the same additional 100 words would increase the insight score by an expected approximately 0.68 percentage-point increase, while holding other variables constant.

Analytical thinking. For the Ancient Hero and Visualizing Japan courses, the analytical thinking score was negatively associated with causation.

Words per sentence. In all three courses studied, words per sentence was positively associated with causation. This expected association was about two times higher in the Super-Earths courses, where, for every ten words added to a sentence, I expected to see an approximately 0.35 percentage-point increase, while holding other variables constant.

Six-letter words. In the Ancient Greek Hero and Visualizing Japan courses, I found that the number of six-letter words in a post was positively associated with causation.

Additional characteristics. I created separate models to explore the influence of other predictors on causation. In the Super-Earths courses, for posts by students who responded to a post originally posted by a student, I expected to see approximately 0.67 percentage points added to the causation scores for these posts, $b = 0.67$, $z = 7.17$, $p < .001$, while holding other variables constant. Additionally, in these same courses, posts that were endorsed were negatively associated with causation, $b = .70$, $z = -3.88$, $p < .001$, while holding other variables constant. I also found that the thread descriptor was negatively associated with causation and that posts marked as discussions had insight scores which I expected to be approximately 0.52 percentage points lower than posts marked as questions, $b = -0.52$, $z = -4.30$, $p < .001$, while holding other variables constant. No other predictors were found to be significant.

Discrepancy. The “discrepancy” sub-category of the “cognitive processing” category of the LIWC text analysis tool contains 83 words (e.g., should, would, could, hope, lack, odd). A summary of results appears in Table 20.

Table 20. Unstandardized Regression Coefficients Estimating Discrepancy

	Hero Courses Coefficient	Super-Earths Courses Coefficient	Visualizing Japan Courses Coefficient
Intercept	2.35***	1.27***	2.16***
(SE)	0.07	0.13	0.19
Self-Paced	-0.01	0.23	-0.17
(SE)	0.06	0.13	0.10
Post Descriptor	0.01	0.44***	0.01
(SE)	0.04	0.07	0.10
Post Type	-0.08*	-[c]	-[c]
(SE)	0.04	-	-
Word Count	0.02***	0.18***	0.06
(SE)	0.04	0.03	0.06
Analytical Thinking	-0.01***	-0.01***	-0.01***
(SE)	0.00	0.00	0.00
Words Per Sentence	0.10***	0.30***	0.10*
(SE)	0.02	0.04	0.03
Six Letter Words	-0.10*	-0.08***	-0.12*
(SE)	0.02	0.02	0.04

Note: $*p < .05$; $**p < .01$; $***p < .001$; [c] = variable dropped because of collinearity.

Post descriptor. In the Super-Earths courses, student posts made as initial posts, when compared to response posts or comment posts, were positively associated with discrepancy and were expected to be approximately 0.44 percentage points higher, while holding other variables constant.

Post type. In the Ancient Greek Hero courses, the post type was negatively associated with the discrepancy score. A post type marked as a comment was expected to be approximately 0.08 percentage points lower than one marked as a comment thread, while holding other variables constant.

Word count. In the Ancient Greek Hero and Super-Earths courses, word count was positively associated with discrepancy, while holding other variables constant

Analytical thinking. In all three courses studied, the analytical thinking scores were negatively associated with discrepancy scores, while holding other variables constant.

Words per sentence. In all three courses studied, the number of words per sentence was positively associated with discrepancy, while holding other variables constant.

Six-letter words. In all three courses studied, the number of six-letter words used in a post was negatively associated with discrepancy, while holding other variables constant.

Additional characteristics. I explored the influence of other predictors through separate models. In the Super-Earths courses, I expected the posts of students who responded to posts originally posted by a student to be approximately 0.58 percentage points higher. $b = 0.58$, $z = 3.97$, $p < .001$, while holding other variables constant. I also found that the thread descriptor variable was negatively associated with causation and that posts marked as discussions were expected to score approximately 0.49 percentage points lower for causation than posts marked as questions, $b = -0.49$, $z(2346) = -3.86$, $p < .001$, while holding other variables constant. In the Visualizing Japan courses, I found that posts that were endorsed were expected to score approximately 1.83 percentage points lower, $b = -1.83$, $z(248) = -4.82$, $p < .001$, than those that were not endorsed, while holding other variables constant.

Tentativeness. The “tentativeness” sub-category of the “cognitive processing” category of the LIWC text analysis tool contains 178 words (e.g., maybe, perhaps, guess, unknown, wonder). A summary of results appear in Table 21.

Table 21. Unstandardized Regression Coefficients Estimating Tentativeness

	Hero Courses Coefficient	Super-Earths Courses Coefficient	Visualizing Japan Courses Coefficient
Intercept	3.44***	1.90***	4.30***
(SE)	0.10	0.18	0.34
Self-Paced	0.09	0.33*	0.07
(SE)	0.08	0.17	0.16
Post Descriptor	0.14*	1.27***	-0.05
(SE)	0.06	0.12	0.16
Post Type	-0.16*	-[c]	-[c]
(SE)	0.06	-	-
Word Count	0.07***	0.27***	0.21
(SE)	0.02	0.04	0.11
Analytical Thinking	-0.02***	-0.01***	-0.02***
(SE)	0.00	0.00	0.00
Words Per Sentence	0.20***	0.50***	-0.00
(SE)	0.03	0.06	0.01
Six Letter Words	0.02	-0.04	-0.20**
(SE)	0.03	0.03	0.01

Note: **p* < .05; ***p* < .01; ****p* < .001; [c] = variable dropped because of collinearity.

Pacing condition. In the Super-Earths courses, pacing condition was positively associated with tentativeness, while holding other variables constant.

Post descriptor. In the Ancient Greek Hero courses, I found that, when compared to either an initial post or a response post, a comment as the post descriptor had scores expected to be approximately 0.14 percentage points higher, while holding other variables constant. In the Super-Earths courses, student posts made as initial posts, when compared to response posts or comment posts, were positively associated with tentativeness and were expected to be approximately 1.27 percentage points higher, while holding other variables constant.

Post type. In the Ancient Greek Hero courses, the post type was negatively associated with the tentativeness score. A post type marked as a comment had a tentativeness score that was expected to be approximately 0.01 percentage points lower than one classified as a comment thread, while holding other variables constant.

Word count. In the Ancient Greek Hero and Super-Earths courses, word count was positively associated with tentativeness, while holding other variables constant.

Analytical thinking. In all three courses studied, analytical thinking was negatively associated with tentativeness, while holding other variables constant.

Words per sentence. In the Ancient Greek Hero and Super-Earths courses, words per sentence was positively associated with tentativeness, while holding other variables constant.

Six-letter words. In the Visualizing Japan courses, the number of six-letter words used in a post was negatively associated with tentativeness, while holding other variables constant.

Additional characteristics. As with the other dependent variables, I ran additional models to look at variables that were not applicable to every post. I found several significant predictors in these models. In the Ancient Greek Hero courses, I found that posts marked as discussions were expected to be approximately 0.90 percentage points higher than posts marked as questions, $b = -0.90$, $z = -2.44$, $p < .05$, while holding other variables constant. In the Super-Earths courses, the scores for posts by students who responded to a post originally posted by a student had posts that are expected to be approximately 1.63 percentage points higher, $b = 1.63$, $z = 7.89$, $p < .001$, while holding other variables constant. Additionally, for the posts of students who made posts in response to peer posts, I expected to see scores that were approximately 0.39 percentage points higher, $b = 0.39$, $z = 2.37$, $p < .05$, while holding other variables constant. I also found that the thread descriptor variable was negatively associated with causation and that posts marked as discussions had insight scores that I expected to be approximately 1.30 percentage points lower than posts marked as questions, $b = -1.30$, $z = -5.64$, $p < .001$, while holding other variables constant. No other predictors were found to be significant.

Certainty. The “certainty” sub-category of the “cognitive processing” category of the LIWC text analysis tool contains 113 words (e.g., always, never, clearly, confident, commit, apparent). A summary of results appears in Table 22.

Table 22. Unstandardized Regression Coefficients Estimating Certainty

	Hero Courses Coefficient	Super-Earths Courses Coefficient	Visualizing Japan Courses Coefficient
Intercept	1.69***	3.08***	2.24***
(SE)	0.30	0.11	0.33
Self-Paced	-0.27***	-0.55***	0.23
(SE)	0.08	0.08	0.17
Post Descriptor	0.09	0.26***	-0.30
(SE)	0.08	0.09	0.19
Post Type	-0.13	-[c]	-[c]
(SE)	0.08	-	-
Word Count	-0.02	-0.15***	0.16
(SE)	0.01	0.00	0.09
Analytical Thinking	-0.01***	-0.02***	-0.01***
(SE)	0.00	0.00	0.00
Words Per Sentence	-0.20***	-0.13***	-0.13
(SE)	0.03	0.03	0.07
Six Letter Words	0.64*	0.15***	0.10
(SE)	0.20	0.04	0.12

Note: **p* < .05; ***p* < .01; ****p* < .001; [c] = variable dropped because of collinearity.

Pacing condition. The pacing condition was negatively associated with certainty in both the Ancient Greek Hero and Super-Earths courses, while holding other variables constant.

Post descriptor. In the Super-Earths courses, for student posts made as initial posts, when compared to response posts or comment posts, I expected to see scores approximately 0.26 percentage points higher for certainty, while holding other variables constant.

Word count. The number of words in a post was negatively associated with certainty in the Super-Earths courses, while holding other variables constant.

Analytical thinking. In all three courses studied, analytical thinking was negatively associated with certainty, while holding other variables constant.

Words per sentence. In both the Ancient Greek Hero and Super-Earths courses, the number of words per sentence in a post was negatively associated with certainty, while holding other variables constant.

Six-letter words. In both the Ancient Greek Hero and Super-Earths courses, the number of six-letter words in a post was positively associated with certainty, while holding other variables constant.

Additional Characteristics. The additional models I ran revealed several other significant predictors. In the Ancient Greek Hero courses, I found that the posts of students who responded to peer posts had values for certainty expected to be approximately 0.20 percentage points higher than the posts of students who responded to instructor posts, $b = .20$, $z = 2.12$, $p < .05$, while holding other variables constant. In the Super-Earths courses, the posts of students who responded to posts originally posted by a student had certainty scores expected to be approximately 1.63 percentage points higher, $b = 1.63$, $z = 7.89$, $p < .001$, while holding other variables constant. Additionally, I expected the posts of students who made posts in response to peer posts to certainty score approximately 0.39 percentage points higher, $b = 0.39$, $z = 2.37$, $p < .05$, while holding other variables constant. In the Visualizing Japan courses, I found that threads started as discussions had certainty scores expected to be approximately 0.58 percentage points higher than threads started as questions, $b = 0.58$, $z = 3.55$, $p < .001$, while holding other variables constant. Additionally, posts that were endorsed had certainty scores that were expected to be approximately 2.07

percentage points higher, $b = -2.07$, $z = -3.36$, $p < .001$, than posts that were not endorsed, while holding other variables constant.

Differentiation. The “differentiation” sub-category of the “cognitive processing” category of the LIWC text analysis tool contains 81 words (e.g., hasn’t, but, else, despite, although, except, and exclude). A summary of results appears in Table 23.

Table 23. Unstandardized Regression Coefficients Estimating Differentiation

	Hero Courses Coefficient	Super-Earths Courses Coefficient	Visualizing Japan Courses Coefficient
Intercept	4.00***	1.86***	3.56***
(SE)	0.11	0.14	0.29
Self-Paced	0.26*	-0.14	0.30
(SE)	0.10	0.15	0.15
Post Descriptor	0.16*	1.05***	0.04
(SE)	0.07	0.09	0.14
Post Type	-0.21***	-[c]	-[c]
(SE)	0.06	-	-
Word Count	0.22***	0.73***	0.72***
(SE)	0.02	0.06	0.00
Analytical Thinking	-0.03***	-0.02***	-0.03***
(SE)	0.00	0.00	0.00
Words Per Sentence	0.40***	0.70***	0.30***
(SE)	0.05	0.09	0.06
Six Letter Words	-0.05	-0.04***	0.02
(SE)	0.02	0.02	0.06

Note: $*p < .05$; $**p < .01$; $***p < .001$; [c] = variable dropped because of collinearity.

Pacing condition. In the Ancient Greek Hero courses, the pacing condition was positively associated with differentiation, while holding other variables constant.

Post descriptor. In the Ancient Greek Hero courses, I found that, when compared to either an initial post or a response post, a post that had a comment as the post descriptor had scores expected to be approximately 0.16 percentage points higher with differentiation, while holding other variables constant. In the Super-Earths courses, student posts that were made as initial posts, when compared to response posts or comment posts, had differentiation scores expected to be approximately 1.05 percentage points higher, while holding other variables constant.

Post type. In the Ancient Greek Hero courses, the post type was negatively associated with the differentiation score. A post type marked as a comment had a differentiation score expected to be approximately 0.21 percentage points lower than one classified as a comment thread, while holding other variables constant.

Word count. In all three courses studied, the number of words in a post was positively associated with differentiation. For every additional 100 words added to the post, both the Super-Earths and Visualizing Japan courses had posts expected to have differentiation scores approximately 0.70 percentage points higher, while holding other variables constant.

Analytical thinking. In all three courses studied, the number of words in a post was negatively associated with differentiation, while holding other variables constant.

Words per sentence. In all three courses studied, the number of words in a post was positively associated with differentiation, while holding other variables constant. This expected influence was greatest in the Super-Earths course, where every ten additional words added to a sentence in a post had an expected increase of approximately 0.70 percentage points in the differentiation score, while holding other variables constant.

Six-letter words. In the Super-Earths courses, the number of six-letter words in a post was negatively associated with differentiation, while holding other variables constant.

Additional characteristics. In my models of additional characteristics, several predictors were found to be significant. In the Ancient Greek Hero courses, I found that the posts of students who responded to peer posts (b = .24, z = 2.70, p < .01) had higher differentiation scores than the posts of students who responded to instructor posts, while holding other variables constant. In the Super-Earths courses, the posts of students who responded to a post originally posted by a student had scores expected to be approximately 0.27 percentage points lower in differentiation score, b = -.27, z = -2.45, p < .05, while holding other variables constant. Additionally, the posts of students who made posts in response to peer posts had differentiation scores that were expected to be approximately 0.43 percentage points lower in differentiation scores, b = -.43, z = -2.52, p < .05, while holding other variables constant. I also found that the thread descriptor variable was positively associated with differentiation in the Super-Earths courses and that posts marked as discussions had differentiation scores that were expected to be approximately 0.41 percentage points higher than posts marked as questions, b = .41, z = 2.65, p < .001, while holding other variables constant. No other predictors were found to be significant.

Conclusion

In this chapter, I reported findings that addressed each of my research questions. To answer my first two questions, I used independent sample t-tests. To account for the possibility of unequal variances between my two pacing conditions, I made use of Levene's robust test of equality. Since my sample sizes were so large, many of my variables showed differences as statistically significant. To provide more context and to better explain these differences, I used standardized effect sizes, known as Cohen's *d*, to describe the practical significance. For my third research question, I made use of multi-level mixed-effects regression models that allowed me to account for the clustering effect caused by students who made multiple posts or who may have influenced other student posts. In my concluding chapter, I further explain these results and align them with the literature. I also offer my insights on what these results mean and what implications they have for the study of MOOCs.

CHAPTER FIVE: DISCUSSION

The purpose of this study was to explore the influence of massive open online course (MOOC) pacing conditions (instructor-paced versus self-paced) on course discussion forum activity and cognitive processing. A multiple-case study approach was used to focus on three Harvard University MOOCs, each one offered in both instructor-paced and self-paced versions. The study was framed by the following questions:

1. How does cognitive processing vary between pacing conditions?
2. In what ways does pacing condition affect discussion forum activity and engagement (e.g., number of replies, number of words used)?
3. What aspects of forum activity are predictive of cognitive processing, and do these relationships exist in both pacing conditions?

In Chapter 1, I presented a background and rationale for the study. I provided an overview of the theoretical framework for the study, the CoI Model, as well as an explanation of the statement of the problem and the purpose of the study. I also highlighted the significance of the study and the research questions that framed it. I used Chapter 2 to review the literature examining the connection between cognitive presence and online discussion forums. In this literature review, I provided an overview of MOOCs, including discussions on MOOC dimensions, structures, research directions, and challenges. I then addressed extant research on MOOC discussion forums. I also introduced the conceptual model that guided my analysis.

In Chapter 3, I detailed my study methodology, specifically, how quantitative content analysis (QCA) is used in this study. This chapter also scrutinized the limitations and constraints that may affect findings and implications. My findings were shared in Chapter 4, and this final chapter revisits my hypotheses and provides meanings and implications.

Research Hypotheses Revisited

This section discusses the extent to which each of my hypotheses is supported by the findings reported in Chapter 4. It also covers my hypotheses' context in the literature, as well as the implications my findings present for MOOC research. I will use Table 24 to revisit and explain the evidence for each of my research hypotheses.

Table 24. Evidence Levels for Research Hypotheses

Evidence Level	Meaning
No Evidence	None of the courses provided evidence to support the hypothesis
Minimal Evidence	Only one course provided evidence in support of the hypothesis
Some Evidence	Two courses provided evidence that supported the hypothesis
Strong Evidence	All three courses provided evidence that supported the hypothesis

Hypothesis 1 (H1): Pacing Condition Indirectly Affects Cognitive Processing

- *Students in the self-paced versions of the courses studied will have higher average cognitive processing scores than students in the instructor-paced courses* ***(No Evidence)***

I expected to find that students in the self-paced courses would have higher average cognitive processing scores. I expected this because self-paced students would have more time to read through and consider posts made by their peers. These expectations were informed by the findings of Kanuka and Garrison (2004) and the work by Campbell et al. (2014). I also considered that the lower number of overall posts in a self-paced course might make students feel a bit less overwhelmed generally. In my findings, I discovered that in the Visualizing Japan and Super-Earths course, it was the instructor-paced students who had slightly higher levels of cognitive processing. I had expected that the additional time available to self-paced students to consider posts and content would result in posts indicative of higher levels of cognitive processing, but this was not the case. In fact, in these two courses, the opposite appeared to be happening. These higher levels of cognitive processing may also suggest that there was more knowledge construction in the instructor-paced course forums (Kent et al., 2016). Further, these findings corroborate the study findings of Wang et al. (2016) on the topic of the connection between higher learning gains and students who were active in forums and who made on-topic posts. The differences in cognitive processing scores between the three courses were very small, with only the Super-Earths course having a statistically significant difference, and that difference was of a very small to small practical significance (d = 0.14). Granted that, while this effect size is noticeable, it is certainly minimized by the large sample size of my study.

I found no evidence that the pacing condition was indirectly affecting cognitive processing. However, the "cognitive processing" category of the Linguistic Inquiry Word Count (LIWC) text analysis tool I used contains 797 words, so I considered it useful to explore sub-categories to determine whether the pacing condition was affecting any of them. The effect sizes were small across all of the courses studied, and there was not a consistent pattern between pacing condition and cognitive processing measures. The students in the self-paced Ancient Greek Hero course had posts that demonstrated more insight, causation, tentativeness, and differentiation. This suggests that their posts showed they were questioning or trying to make sense of the course's content. For instance, terms included in the "tentativeness" sub-category—e.g., maybe, perhaps, guess—point to less authoritative posts, which, in turn, could foster more communication among peers (Momeni, Haslhofer, Tao, & Houben, 2015). The students in the self-paced Super-Earths course tended to use words more indicative of causation, discrepancy, and tentativeness. The presence of discrepancy may suggest these that students were making posts that had vague statements and were trying to cover their lack of knowledge about the topic (Williams & D'Mello, 2010).

The outliers for most of the measures were the Visualizing Japan courses. This study did not examine course content or specific prompts but instead focused, for purposes of analysis, on the course discussion forum posts. I think there are several factors that explain why these courses were outliers. The first factor was that these courses had the lowest enrollments of all the courses studied. As explained in other chapters, a small percentage of students engage in discussion forums and, given the lower enrollment numbers for the Visualizing Japan courses, this engagement was especially low, perhaps contributing to the lower scores across the board. Additionally, the content itself may have presented some challenges for learners. A lack of prior or basic knowledge of Japanese history–content that is not a part of traditional K-12 education—may have impacted the ability of students to craft posts high in cognitive processing. While there was not a consistent pattern of specific cognitive processing sub-categories between the self-paced and instructor-paced versions of the Ancient Greek Hero and Super-Earths courses, the Visualizing Japan courses had just one significant category score difference (discrepancy) between the versions, and scores for all but two measures (certainty and differentiation) were higher for the instructor-paced version.

Based on these findings, I found no evidence that clearly showed support for my hypothesis. For two of the courses, Ancient Greek Hero and Super-Earths, the data does not show much difference in cognitive processing between the two pacing conditions. These two courses had the largest number of posts, and this large sample size may have contributed to how many measures showed as being statistically significant. But even in the Visualizing Japan courses, which had a much smaller sample size, there was still not much difference between the two pacing conditions. Within the Visualizing Japan courses, the instructor-paced version did appear to have higher average levels of cognitive processing, although there was not a significant statistical difference for most of the measures.

Hypothesis 2 (H2): Pacing Condition Affects Discussion Forum Activity and Engagement

For this hypothesis, I wanted to take a closer look at specific attributes of discussion forum activity and engagement and to determine how they varied by pacing condition. I considered this to be an important area to understand. By better comprehending what students are doing within discussion forums, we can start to understand how they are participating in online environments such as MOOCs (Chiu & Hew, 2017). I also wanted to provide some information and analysis to more thoroughly describe discussion forum activity in these two pacing conditions, since much is still unknown about the effects of discussion forums within MOOCs (Chiu & Hew, 2017).

- *Students in the instructor-paced versions of the courses will post more frequently* ***(Strong Evidence)***
- *Students in the self-paced versions of the courses will make longer posts (thread length)* ***(No Evidence)***

Students in the instructor-paced versions of the courses studied certainly made more posts than students in the self-paced version. One factor that might explain this is the fact that the student introductions within instructor-paced courses required more information from students, thereby resulting in longer average post threads. Since there were more students overall in the instructor-paced courses, requiring that they post background information and reasons for taking the course in the discussion forums may have skewed the results for "post position" and "amount of posts." In a future exploration of this data, I would be curious to see what these numbers would look like if I were to remove the student introduction posts and run the same analysis.

- *Students in the self-paced versions of the courses will have more reply posts* ***(Minimal Evidence)***

This finding had minimal evidence. In the Ancient Greek Hero self-paced course, there was a statistically significant difference showing that more reply posts came from these students than from students in the instructor-paced version of the course. However, in the Super-Earths courses there was no statistical difference between reply post numbers across the course versions. In the Visualizing Japan courses, the instructor-paced course showed the significant difference on this measure. What is interesting about this finding is that it suggests that Sharif and Magrill (2015) may have been wrong about students not engaging in self-paced discussion forums. Not only were students engaging with each other and with their instructor in the self-paced version of the Ancient Greek Hero course, they were doing it at a higher rate than their peers in the instructor-paced version. The instructor-paced Visualizing Japan course showing more reply posts may again suggest that students were less familiar with the content and, therefore, were more reliant on the instructor to guide them through the content and help them understand it.

- *Students in the self-paced versions of the courses will make longer posts (word count)* ***(Some Evidence)***

The Visualizing Japan courses were again the outliers, which is why this hypothesis is partially supported. I expected to find longer (by word count) posts with the self-paced courses, and the findings for the Ancient Greek Hero and Super-Earths courses bore this out. Students in these two courses made longer posts in terms of word count and more words per sentence, and they were more analytical than their peers in the instructor-paced versions. This all suggests that students were making longer and more meaningful posts (analytical thinking score) in the self-paced versions of these two courses. Also, there were more six-letter words used on average in the posts made in the self-paced Ancient Greek Hero course. I expect that the additional time students in this course had to read and to consider each other's posts may have contributed to these differences.

Overall, my findings suggest that pacing condition may influence specific discussion forum activity measures and student engagement. A clear example is the average post length measure, which was influenced by the required student introduction posts.

Hypothesis (H3): Discussion Forum Activity and Engagement Affect Cognitive Processing

To test this hypothesis, I created several regression models. Findings are presented in Chapter 4 and summarized in Table 25. I created a regression model for each of the dependent variables (DVs). For this hypothesis, I examined the relationship that discussion forum activity has on the demonstration of cognitive processing and the associated sub-scores that make up the overall cognitive processing score. I also used post characteristics (e.g., response post, comment versus comment thread) as independent variables to explore each one's impact on cognitive processing. Using the cognitive processing sub-scores as dependent variables allows for a more refined look at how discussion forum engagement variables serve as predictors of the associated elements of cognitive processing. I expected to find that the posts of students in the self-paced discussion forums would show higher levels of insight, certainty, and differentiation than the posts of students in the instructor-paced discussion forums.

- *Longer (in terms of word count) posts made as discussion posts will have higher levels of cognitive processing* ***(Strong Evidence)***

The findings provided strong evidence in support for this hypothesis. For all three courses studied, there was a positive association between cognitive processing and word count, and that association was significant for the Ancient Greek Hero and Super-Earths courses. This makes sense, since the dependent variable, cognitive processing, is percentage of words used that fall into that LIWC category. Chiu and Hew (2018) link writing posts to higher levels of cognitive processing because doing this requires that a student read, think, understand, and make a statement or provide a reply in the forum. Therefore, the use of more words provides more opportunities for students to use words within the "cognitive processing" category. My findings suggest that word count may be a rough proxy of higher-order thinking. They further suggest that not only were students using more words overall, they were using more words that were indicative of cognitive processing.

- *Students in the self-paced versions of the courses will have posts that have higher levels of insight, certainty, and differentiation than students in the instructor-paced courses.* ***(Minimal Evidence)***

In the Ancient Greek Hero courses, there was a positive association between the pacing condition of "self-paced course" and insight, causation, tentativeness, and differentiation; but no such positive association was observed for the instructor-paced version. The association between self-paced course and tentativeness was not statistically significant when compared to the measure for tentativeness in the instructor-paced version. As previously mentioned, students showed lower levels of certainty in the self-paced version of the Ancient Greek Hero course.

In the Super-Earths self-paced course, tentativeness was the only category statistically significant when it came to measuring positive association. Certainty was negatively associated with tentativeness. For both the Ancient Greek Hero and the Super-Earths self-paced courses, students were making posts with lower levels of certainty. In the Visualizing Japan courses, there were no statistically significant variables discovered by the regression model. This aligns with other findings showing that the Visualizing Japan courses were outliers.

The Super-Earths courses had some of the more interesting results. In these courses, the cognitive processing average score was about 1.92 percentage points lower in the self-paced version of the course. One of the benefits of having done analysis on each of the sub-scores is that I am able to get more granular in describing posts. My findings showed that "insight" may have had the most influence on these lower overall cognitive processing scores for the Super-Earths courses. In the self-paced version of Super-Earths, student posts had insight scores that were about 1.57 percentage points lower than those found in the instructor-paced version. Additionally, student posts in the self-paced course had lower scores (by about 0.55 percentage points) in the certainty category. Conversely, the posts of self-paced Super-Earths students had a statistically significant difference for tentativeness and discrepancy as compared to posts by their instructor-paced peers. As Momeni et al. (2015) found, the use of tentative words is an indicator of a post with fewer claims about correctness and can show that the student poster is wrestling with understanding course concepts. This makes sense, since a course of this nature may present some potentially murky content areas. The post descriptor (whether it was an initial post versus a response or comment post) also influenced several of the sub-scores with the Super-Earths courses. In looking at the tentativeness scores, initial posts were about 1.3 percentage points higher than response or comment posts. For the differentiation score, initial posts were about 1.05 percentage points higher for tentativeness than response or comment posts. Also reflective of tentativeness, a student post responding to another classmate's post scored about 1.63 percentage points higher than the original post.

Table 25. Effect Size of Dependent Variables (DVs), by Course

DV	Ancient Greek Hero	Super-Earths	Visualizing Japan
Cognitive Processing	Self-paced (+0.16)	Instructor-paced (+1.88)	Instructor-paced (+0.19)
Effect Size	-	**Very small to small**	-
Insight	Self-paced (+0.21)	Instructor-paced (+1.93)	Instructor-paced (+0.06)
Effect Size	**Very small to small**	**Very small to small**	-
Causation	Self-paced (+0.05)	Self-paced (+0.23)	Instructor-paced (+0.12)
Effect Size	**Very small to small**	**Very small to small**	-
Discrepancy	Instructor-paced (+0.06)	Self-paced (+0.34)	Instructor-paced (+0.20)
Effect Size	-	**Very small to small**	**Very small to small**
Tentativeness	Self-paced (+0.05)	Self-paced (+0.52)	Instructor-paced (+0.05)
Effect Size	-	**Very small to small**	-
Certainty	Instructor-paced (+0.30)	Instructor-paced (+0.72)	Self-paced (+0.07)
Effect Size	**Very small to small**	**Very small to small**	-
Differentiation	Self-paced (+0.20)	Self-paced (+0.08)	Self-paced (+0.12)
Effect Size	**Very small to small**	-	-

Limitations

One of the limitations of this study was that the analysis was done post-hoc and without much context. For instance, I do not know the specific content of the courses studied, what the instructors' learning objectives were related to discussion forum activity, or what other factors may have influenced students' discussion forum usage. In my study I focused only on students who made discussion forum posts. This limited focus did not consider the text of instructor prompts or the role the instructor had in engaging with students who made posts. This limited my ability to make definitive claims about the full extent of the role of teaching presence in the demonstration of cognitive presence within the courses. However, further exploring the engagement between students and instructors within discussion forums would be an interesting area for follow-up study and analysis. Also, I did not look at students who may have spent time reading discussion posts, nor did I look for relationships between pre-course indications of forum usage and actual discussion forum usage.

The text analysis tool I used presented some limitations. First, LIWC does not look at context. This tool provides a unigram level of analysis and simply counts word frequencies to calculate a percentage-based score. It is also uses pre-determined categories. I did not have the ability to edit words within the LIWC categories. The tool's "analytical thinking" score is a proprietary measure, and I do not know exactly which words or phrases fall within that variable.

Another limitation, one that I had not considered when I started this study, was the size of my dataset. In this case, more may have been a bad thing. Because my sample was so large, it produced statistically significant results for most of the variables, even though they may not have been practically significant. For future research, I will use more demographic information to narrow down observed students. For instance, pre-course surveys asks students how much time they intend to participate in discussion forums. I could use these responses to form sub-groups within each course and then determine what relationships and differences among groups. Making the observations smaller or making the differences in group sizes smaller may allow for a more robust analysis.

Implications

While my study does not allow me to make any definitive conclusions regarding pacing conditions—e.g., self-paced is the best way to influence cognitive processing—it does identify several implications and potential areas of future research. My results provide recommendations for forum structure, as well as implications regarding the role of teaching presence, engagement within self-paced courses, and the creation of a sense of community among online learners. The study also presents a novel approach to using learning analytics with massive datasets.

Forum Structure

This study focused on student engagement within discussion forums and provides insight into how the self-paced and instructor-paced pacing conditions may influence the type and substantive quality of student discussion forum posts. The focus on discussion forums allows me to make several recommendations and to list implications arising from the study. In each of the courses in this study, there was a high number of discussion forum posts. And while the content differed between the courses, the forums used a similar structure. One of the key structural decisions made regarding the forums was that they had pre-defined categories, which forced students to organize and group their threads or replies under appropriate categories. This instructional approach is not typically found in traditional online courses but may be a critical aspect for MOOCs. While it may seem that having pre-determined categories could limit the ability of students to self-direct their learning, something particularly important for a self-paced course (Campbell et al., 2014), it actually may increase the number of posts that foster meaningful discussions within the course (Brooks & Jeong, 2006). The instructors here specifically selected categories that aligned with their course's content. This, in turn, forced students to think about the content before making a post; it also allowed other students to more readily find peers' posts.

Another forum structure decision involved separating discussion posts from question posts. This allowed question posts such as "how to complete a test" or "how to get a certificate" to stay separated from content-specific discussion posts. This helped make sure that students were able to get prompt responses to their questions (e.g., an instructor, staff member, or other student could easily go to the questions section on a forum) without having the question post get lost within a discussion. Anything steps that the instructor can do to help categorize and organize content makes it that much easier for students to find the information they are looking for within a forum. This organization is particularly helpful in that the instructor-paced versions of the courses studied were archived so that students could return to a course after its end date to review materials (in the self-paced versions, students were going through the content and forums at different times). In both course versions, having clearly defined sections within forums for discussions and questions is useful and beneficial for students.

Forum structure decisions such as having pre-defined categories and separate discussion and question areas may influence cognitive presence in a different way. By making content easier to find, students may not make duplicate posts. Students can see what others have already posted and can either add to or use that information to inform their own thinking (instead of just repeating the information in a separate post). The impact of this is likely to be content-specific and driven, ultimately, by how the course is structured. A good example of the potential value of such structuring can be found in the self-paced Super-Earths course. In this course, students were presented in the forums with different problems and equations to solve. As students went through the course's content, if they ran into an issue they could look at the discussion forum and, likely, find an answer and explanation there. They could check their own understanding against that of other students who had previously viewed the content.

Teaching Presence

In an online course, the interaction that occurs between students, typically within the discussion forum, is dictated by the course structures put in place by the instructor and is identified as "teaching presence" within the Community of Inquiry (CoI) model (Garrison, Cleveland-Innes, & Fung, 2010; Joksimovic, Gasevic, Kovanovic, Adesope, & Hatala, 2014). Teaching presence was demonstrated in several ways within the courses examined in this study. The discussion forums all used a similar delivery structure, whereby students created posts using pre-determined instructor categories. In the Visualizing Japan and Super-Earths courses, students were more likely to look to an instructor to make an original post, whereas in the Ancient Greek Hero courses, students were more likely to respond to a post that was originated by a student. This shows a more structured and guided content delivery process in the first two courses. But in all three courses, students were more likely to reply to other student reply posts than to replies by instructors. This suggests that, for a MOOC course, it is important to have a good balance between guided topics and student interaction.

It also suggests that students can and will engage with each other if the instructor provides proper guidance. One of the advantages of having pre-determined categories is that it may help students determine where to make posts, where to find each other's posts, and where to group similar conversations together.

Engagement in the Self-Paced Forums

One of the goals of this study was to explore whether students would feel compelled to engage in the self-paced forums and would create high-quality discussions in those forums. My findings support those of Campbell et al. (2014) that students in a self-paced course will make posts within the discussion forum. In fact, for several measures, such as word count and words per sentence, I found that self-paced students had posts with higher scores on these measures than their peers in the instructor-paced versions of the courses. This has several important implications for course designers. First, it further supports the existence, and the importance, of the connection between teaching presence and cognitive presence. If a course's content and discussion prompts are well designed, this can create opportunities for students to make substantive posts in the course's discussion forum. This also suggests that a student's decision to make a post, and his or her demonstration of cognitive presence, is more predicated upon the content and the forum prompts and structure and less influenced by the pacing condition itself. This means that focusing on how a discussion forum is structured through pre-determined threads or categories and compelling prompts can help to generate critical thinking by students, even in a self-paced course. The findings of this study also suggest that self-paced course discussion forums have the potential to provide the student-to-student and student-to-content interactions that are so important in online learning environments.

Creating a Sense of Community

Maintaining course momentum and keeping learners engaged is a major struggle for MOOCs. The number of learners who complete a MOOC can vary widely from the number who sign up for it, and this study showed that the trend of students enrolling but not engaging with course content still continues. And while studies such as the one completed by Yuan and Powell (2013) find dropout rates of 80–95 percent for MOOCs, Gee (2012) suggests that it is important to understand what the goals for MOOCs are because, if the primary goal is simply to provide access to course content that may not be otherwise accessible, then the dropout rates for MOOCs would not be a concern. Several studies on online learning environments have found that attrition is lower when there is a higher sense of community among learners (e.g., Dueber & Misanchuk, 2001; Moore, 2014, 2016). One way to foster community in online courses is through engagement within the discussion forums. In this study, there appears to be evidence demonstrating that students were interested in engaging with each other's posts regardless of the pacing condition. Even with the unequal variances between the number of discussion forum posts in each version of the courses studied, there were clear levels of student-to-student engagement within discussion forums. Requiring students to make posts introducing themselves and setting out their goals for the course helped create a sense of community. These introductions impacted the length of the posts made, particularly within the instructor-paced versions of the courses. Giving students an opportunity to share their backgrounds and discuss what they hoped to gain in the courses allowed for connections to be made between students.

To further foster the development of community, MOOCs would benefit from employing techniques used by smaller-scale online courses, including providing introductory activities that allow students to introduce themselves to one another or to form and work in learning groups throughout the duration of the course. Each of the courses in the study had these types of introductory posts, and it would be worth further exploration to see what type of follow-up engagement occurred as a result of these introductory posts. These techniques may seem contrary to the MOOC goal of providing self-paced learning, but giving learners an opportunity to develop cohorts can help facilitate learning and create the sense of having a support system as they move through the courses. Also, there was a relationship between course content and student participation within discussion forums, as students who viewed half of course content chapters were more likely to post in discussion forums versus those who had not done this reading. This shows that there was some continuity and connection between students staying engaged both with a course's content and with each other within the course's discussion forum.

Learning Analytics

This study had a massive dataset, and I was able to analyze a lot of information and to extrapolate meaning in a relatively short period of time. In my study, I identified a gap between the widely supported CoI model and MOOC discussion forums. That gap was created from the tedious nature of coding for the CoI model that is simply not realistic when looking at MOOC forums. I was able to use LIWC, a tool that has been used most to look at the psychological factors of words (Tausczik & Pennebaker, 2010), and apply it to the massive dataset created from the Harvard MOOC discussion forums. By using LIWC, I was able to efficiently and consistently analyze discussion forum transcripts across the six courses studied. Even without the full context of the course or any understanding of the course structure, the learning analytics I used were able to provide a rough idea of what was occurring in these courses. If I had the opportunity to interact with the course content, meet with the instructors, or get more information about some of the pedagogical decisions made by the instructors, the value of a tool such as LIWC would only increase. For course designers or instructors, the application of a tool such as LIWC could provide an efficient way to "take the temperature" of their course discussion forums. They would be able to take their intended instructional purposes and compare them to the actual results of the student forum posts. This would be incredibly useful information, both during and after a course, in the ongoing effort to foster the construction of knowledge.

As more and more data are being created by our learning management systems, it has created that much more data that can be mined or explored. For many, these large datasets can be overwhelming and serve as a limitation for analysis. My study has shown that a tool such as LIWC can help cut through some of that wariness and provide valuable information to course designers, instructors, and even administrators. The speed at which LIWC can parse and provide information, coupled with the power of a statistical analysis tool such as Stata, opens up so many research opportunities. My study should provide other researchers with a starting point to expand and explore their own datasets. Another idea involves integrating LIWC into the Learning Management System (LMS) admin area, which would allow the automatic text analysis to run in the background and provide useful statistics, e.g., average word count, words per sentence, to instructors on an on-demand basis. This, in turn, would allow for a tighter integration between the content produced by students and the analysis that can inform an instructor's management of the course discussion forum. This could also help temper or manage the level of involvement the instructor has in a discussion forum through prompts and additional questions or replies to student posts.

Future Research

As I completed my study, I uncovered areas worthy of further research. Based on the parameters of this study, I was not looking at instructor-student interactions outside of determining the type of post to which a student was replying. Future research on the role of the instructor within the discussion forum and on tracking how this role may influence discussion forum activity and subsequent demonstration of cognitive processing would be of interest. I also did not look specifically at course structures or study how those structures and instructional strategies may have influenced the demonstration of cognitive processing.

As mentioned in the implications section, one of the reasons I am even able to consider analyzing this data and looking at different angles on it is because of the significant benefits and advantages of educational data mining. This is an area that I am particularly interested in exploring, as I see it having potential to have major impacts on not only how we organize and structure MOOCs, but also on how we understand student behavior and learning within MOOCs. By leveraging more of the log files, e.g., educational background, indicators of learning preference, I can create more refined groups, which will make the data analysis that much richer. That is a significant step forward in the advancement not only of educational data mining, but also in our understanding of learner behavior in MOOCs. I demonstrated how effective LIWC can be in providing a summary of student discussion posts, and that effectiveness can be improved with more refined groups. My study benefited from having identifiable data from Harvard; many of the other published studies did not have identifiable data. This can allow for more robust post-hoc analysis. Through my study, I have found things that might not work or could work in the future. The power of educational data mining is that it allows for replication and the testing of new hypotheses. I hope to further explore the areas of instructor engagement within forums, the creation of learner profiles to better understand student behavior within MOOCs, and the role of learner motivations as predictors of discussion forum engagement.

Instructor Engagement Within the Forum

As previously mentioned, there are clear indicators that there was instructor presence within the studied courses' discussion forums, as seen through the structure of the content and the forum topics. It would be interesting to explore in greater depth exactly what type of instructor engagement occurred within the forums. For instance, how did the instructor handle replies within the self-paced courses versus in the instructor-paced courses? I would also be interested in looking at the text of the discussion prompts, as it is clear in the Visualizing Japan and Super-Earths courses that students were more likely to respond to an instructor-originated post. I would also be interested in analyzing the types of replies that instructors were making when posted replies. Did they leverage Socratic methods, or did they use different approaches? It would also be of interest to identify the ways, and whether, instructors were facilitating and guiding discussions, outside of setting up the original posts and categories.

Learner Profiles

In this study, I utilized the pacing condition as my primary categorization for the discussion forums. I treated all posters in the courses the same and did not account for differences in prior educational level, motivations, or intentions for the course, or for other factors. In future analysis, I am interested in taking a more refined look at discussion forum activity and making use of more of the identifiable data provided through the log files, e.g., educational background and indicators of learning preference. My study used such large sample sizes that it is difficult to pinpoint truly significant differences. However, by creating learner profiles that group students by factors such as who intended to use the forum, educational level, and final course grade, I may be able to identify more predictive factors for cognitive processing.

Learner Motivations

I am also interested in looking at learner motivations and how they may have influenced engagement within discussion forums. Each of the students in the courses studied were presented with a pre-course survey and registration form, which included questions such as whether the student intended to complete the course, whether he or she wanted to earn a certificate, and how much he or she planned to participate in discussion forums. I have data on actual completed activity for these students, so I would be able to link up expressed intentions with what a student actually did within a course. I could use this information to draw conclusions about what factors may influence a student to participate at higher or lower levels within a discussion forum and why a given student's participation level may have differed from the planned level laid out in the course survey. This would also contribute to the literature about the role of credentialing within MOOCs, as I would be able to look at completion and discussion forum engagement together.

Summary and Conclusions

While massive open online courses (MOOCs) can be viewed as a disruptive force for higher education, this is not necessarily a bad thing. It is often disruptive forces that identify areas for improvement and result in the development of a more effective and efficient method of getting something done. The development of MOOCs presents the higher education establishment with a unique opportunity to re-evaluate both its structure and its business model as society moves toward a more highly connected and digital global world. The current business model for higher education is unsustainable, as more and more students are becoming non-traditional or commuter students, and these non-traditional students are making up an increasingly larger percentage of the enrolled-students pie. Keohane (2013) suggests that “the most important source of funds for many institutions . . . [is] students paying tuition for residential education” (p. 103). But if college costs continue to skyrocket, more learners will be seeking alternative (read: non-residential) ways to gain the skills they need to be successful in today’s economy. Higher education therefore must take a proactive approach to figuring out ways to meet the needs of modern learners and avoid attachments to antiquated structures.

To this end, higher education institutions must be particularly attentive to the risk that faculty members will shift their foci away from developing face-to-face content in favor of exclusively open, online content (Keohane, 2013). Keohane explains that this shift can be detrimental to higher education in that it can start to weaken the importance of residential education, which, again, is the very thing that funds most faculty members’ salaries and gives them the infrastructure through which to develop and deliver online content. Keohane also points out that the openness of online learning could result in students cutting ties to specific institutions and instead looking at education as more of a “shopping mall” experience. In other words, students would view education as being all the same and think that it doesn’t matter where they obtain it (p. 103).

Higher education would benefit by leveraging information gained from assessments and practices of the MOOC model. For example, MOOCs by definition have large enrollments, and enrollment in any given class is purely voluntary. Higher education should examine the factors that contributed to learner enrollment in particular MOOCS, determining just what intrigued learners about the content or the delivery method of certain classes. In doing this, higher education can learn more about what people are looking for education-wise and how they learn best, and the education establishment can then adapt these concepts into a redefined teaching model.

My study addressed several gaps in the current literature. First, there has been a lack of extensive application of the Community of Inquiry (CoI) model to MOOCs. The most significant reason for this is the complexity of coding for CoI presences on the large datasets generated by MOOCs. Second, my research looked at the utility of leveraging quantitative content analysis to extrapolate meaning from large datasets. This study thus served the dual purpose of not only contributing to the literature on the development of cognitive presence in MOOCs, but also furthering the empirical research on the ability to use automated text analysis to measure the demonstration of cognitive processing in MOOCs. The work of Kovanović et al. (2016) is relatively new—published just two summers ago—and has already received accolades within the educational data mining community.

The ability to use an automatic text analysis tool allowed me to efficiently analyze 57,650 discussion posts from 13,495 students. By using a text analysis tool, I was able to consistently apply the same coding parameters, which allowed me to compare between and across many cases. Through this research, I have identified ways to improve the study and further maximize the effectiveness of the LIWC tool. As previously mentioned, refining my groups to be more descriptive (e.g., students who wanted to post in the forum, had similar educational backgrounds) could allow for a deeper level of analysis.

REFERENCES

Akcaoglu, M., & Lee, E. (2016). Increasing social presence in online learning through small group discussions. *The International Review of Research in Open and Distributed Learning*, *17*(3), 1–17. http://doi.org/10.19173/irrodl.v17i3.2293

Akyol, Z., Arbaugh, J. B., Cleveland-innes, M., Garrison, D. R., Ice, P., Richardson, J. C., & Swan, K. (2009). A response to the review of the community of inquiry framework. *International Journal of E-Learning & Distance Education*, *23*(2), 123–136.

Akyol, Z., & Garrison, D. R. (2011a). Assessing metacognition in an online community of inquiry. *Internet and Higher Education*, *14*(3), 183–190. http://doi.org/10.1016/j.iheduc.2011.01.005

Akyol, Z., & Garrison, D. R. (2011b). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Educational Technology*, *42*(2), 233–250. http://doi.org/10.1111/j.1467-8535.2009.01029.x

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. *WWW '14 Proceedings of the 23rd International Conference on World Wide Web*, 687–698. http://doi.org/10.1145/2566486.2568042

Anderson, T., & Dron, J. (2011). Three generations of distance education pedagogy. *International Review of Research in Open and Distance Learning*, *12*(3), 80–97.

Anderson, T., Rourke, L., Garrison, D. R., & Archer, W. (2001). Assessing teacher presence in a computer conferncing context. *Journal of Asynchronous Learning Networks*, *5*(2), 1–17.

Baer, J. (2003). Grouping and achievement in cooperative learning. *College Teaching*, *51*(4), 169–175.

Beckmann, J., & Weber, P. (2016). Cognitive presence in virtual collaborative learning. *Interactive Technology and Smart Education*, *13*(1), 52–70. http://doi.org/10.1108/ITSE-12-2015-0034

Bergner, Y., Kerr, D., & Pritchard, D. E. (2015). Methodological challenges in the analysis of MOOC data for exploring the relationship between discussion forum views and learning outcomes. In *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 234–241).

Breivik, J. (2016). Critical thinking in online educational discussions measured as progress through inquiry phases : A discussion of the cognitive presence construct in the community of inquiry framework. *International Journal of E-Learning & Distance Education*, *32*(1), 1–16.

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom research into edX's first MOOC. *Research & Practice in Assessment*, *8*, 13–25.

Brooks, C. D., & Jeong, A. (2006). Effects of pre-structuring discussion threads on group interaction and group performance in computer-supported collaborative argumentation. *Distance Education*, *27*(3), 371–390. http://doi.org/10.1080/01587910600940448

Campbell, J., Gibbs, A., Najafi, H., & Severinski, C. (2014). DELETE??? A comparison of learner intent and behaviour in live and archived MOOCs. *International Review of Research in Open and Distance Learning*, *15*(5), 235–262. http://doi.org/10.19173/irrodl.v15i5.1854

Cercone, K. (2008). Characteristics of adult learners with implications for online learning design. *AACE Journal*, *16*(2), 137–159. http://doi.org/Article

Chiu, T. K. F., & Hew, T. K. F. (2017). Factors influencing peer learning and performance in MOOC asynchronous online discussion forum. *Australasian Journal of Educational Technology*, *34*(4), 16–28. http://doi.org/10.14742/ajet.3240

Chuang, I., & Ho, A. (2016). HarvardX and MITx: Four years of open online courses- Fall 2012-Summer 2016. *SSRN Electronic Journal*, 1–19. http://doi.org/10.2139/ssrn.2889436

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. http://doi.org/10.1037/0033-2909.112.1.155

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches*. *SAGE Publications* (3rd ed.). Thousand Oaks, CA: SAGE Publications, Inc. http://doi.org/10.1111/1467-9299.00177

Creswell, J. W., & Clark, V. L. . (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing “Course”: Reconceptualizing educational variables for massive open online courses. *Educational Researcher*, *43*(2), 74–84.

Digest of Education Statistics, 2015 (NCES 2016-006). (2016). Retrieved June 26, 2016, from http://nces.ed.gov/fastfacts/display.asp?id=80

Donnelly, R., & Gardner, J. (2011). Content analysis of computer conferencing transcripts. *Journal of Interactive Learning Environments*, *19*(4), 303–315.

http://doi.org/10.1080/10494820903075722

Dowell, N. M. M., & Graesser, A. C. (2014). Modeling learners' cognitive, affective, and social processes through language and discourse. *Journal of Learning Analytics*, *1*(3), 183–186.

Dowell, N. M. M., Graesser, A. C., & Cai, Z. (2015). Language and Discourse Analysis with Coh-Metrix : Applications from Educational Material to Learning Environments at Scale. *Journal of Learning Analytics*, *7750*(October).

Dueber, B., & Misanchuk, M. (2001). Sense of community in a distance education course. In *Proceedings of the Mid South Instructional Technology Conference* (pp. 1–23).

Eynon, R., Hjorth, I., Yasseri, T., & Gillani, N. (2016). Understanding communication patterns in MOOCs: Combining data mining and qualitative methods. In S. ElAtia, D. Ipperciel, & O. R. Zaiane (Eds.), *Data Mining and Learning Analytics: Applications in Educational Research*.

Faasse, K., Chatman, C. J., & Martin, L. R. (2016). A comparison of language use in pro- and anti-vaccination comments in response to a high profile Facebook post. *Vaccine*, *34*(47), 5808–5814. http://doi.org/10.1016/j.vaccine.2016.09.029

Fini, A. (2009). The technological dimension of a massive open online course: The case of the CCK08 course tools. *The International Review of Research in Open and Distributed Learning*, *10*(5), 1–26. http://doi.org/10.19173/irrodl.v10i5.643

Gallagher, S. E., & Savage, T. (2016). Comparing learner community behavior in multiple presentations of a massive open online course. *Journal of Computing in Higher Education*, *28*(3), 358–369. http://doi.org/10.1007/s12528-016-9124-y

Gallagher, S., & LaBrie, J. (2012). Online learning 2.0: Strategies for a mature market.

Continuing Higher Education Review, *76*, 65–73.

Garrison, D. R. (2007). Online community of inquiry review: social, cognitive, and teaching presence issues. *Journal of Asynchronous Learning Networks*, *11*(1), 61–72. http://doi.org/10.1128/JB.05513-11

Garrison, D. R., & Anderson, T. (2003). *E-Learning in the 21st century: A framework for research and practice*. New York, NY: RoutledgeFalmer. http://doi.org/10.4324/9780203166093

Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2–3), 87–105. http://doi.org/0.1016/S1096-7516(00)00016-6

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, *15*(1), 7–23. http://doi.org/10.1080/08923640109527071

Garrison, D. R., Anderson, T., & Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education*, *13*(1–2), 5–9. http://doi.org/10.1016/j.iheduc.2009.10.003

Garrison, D. R., Cleveland-Innes, M., & Fung, T. S. (2010). Exploring causal relationships among teaching, cognitive and social presence: Student perceptions of the community of inquiry framework. *The Internet and Higher Education*, *13*(1–2), 31–36. http://doi.org/10.1016/j.iheduc.2009.10.002

Garrison, D. R., Cleveland-Innes, M., & Fung, T. S. (2010). Exploring causal relationships among teaching, cognitive and social presence: Student perceptions of the community of inquiry framework. *Internet and Higher Education*, *13*, 31–36.

http://doi.org/10.1016/j.iheduc.2009.10.002

Gašević, D., Adesope, O., Joksimović, S., & Kovanović, V. (2015). Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. *Internet and Higher Education, 24*, 53–65. http://doi.org/10.1016/j.iheduc.2014.09.006

Gee, S. (2012). MITx - the fallout rate. Retrieved November 26, 2017, from http://www.i-programmer.info/news/150-training-a-education/4372-mitx-the-fallout-rate.html

Gibson, A. M., Ice, P., Mitchell, R., & Kupczynski, L. (2012). An inquiry into relationships between demographic factors and teaching, social, and cognitive presence. *Internet Learning*, *1*(1), 7–16.

Gilbert, P. K., & Dabbagh, N. (2005). How to structure online discussions for meaningful discourse: A case study. *British Journal of Educational Technology, 36*(1), 5–18.

Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education*, *23*, 18–26. http://doi.org/10.1016/j.iheduc.2014.05.004

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40*(5), 223–234. http://doi.org/10.3102/0013189X11413260

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 193–202. http://doi.org/10.3758/BF03195564

Gunawardena, C. N., Lowe, C. A., & Anderson, T. (1997). Analysis of a global online debate and the development of an interaction analysis model for examining social construction of knowledge in computer conferening. *Journal of Educational Computing Research*, *17*, 397–431.

Gwynne, P. (2013). A Fresh Twist on Online Learning. *Research Technology Management*, *56*(1), 6–7.

Hara, N., Bonk, C., & Anjeli, C. (2000). Content analysis of online discussions in an applied educational pyschology course. *Instructional Science*, *28*, 115–152.

Hecking, T., Chounta, I.-A., & Hoppe, H. U. (2015). Analysis of user roles and the emergence of themes in discussion forums. In *2015 Second European Network Intelligence Conference* (pp. 114–121). IEEE. http://doi.org/10.1109/ENIC.2015.24

Hecking, T., Hoppe, H. U., & Harrer, A. (2015). Uncovering the structure of knowledge exchange in a MOOC discussion forum. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 - ASONAM '15*, 1614–1615. http://doi.org/10.1145/2808797.2809359

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53.

Hewitt, J. (2003). How habitual online practices affect the development of asynchronous discussion threads. *Journal of Educational Computing Research*, *28*(1), 31–45.

Joksimovic, S., Gasevic, D., Kovanovic, V., Adesope, O., & Hatala, M. (2014). Psychological characteristics in cognitive presence of communities of inquiry: A linguistic analysis of online discussions. *Internet and Higher Education*, *22*, 1–10. http://doi.org/10.1016/j.iheduc.2014.03.001

Jordan, K. (2015). Massive Open Online Course Completion Rates Revisited : Assessment , Length and Attrition. *International Review of Research in Open and Distributed*

Learning, *16*(3), 341–358. http://doi.org/10.13140/RG.2.1.2119.6963

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the linguistic inquiry and word count. *The American Journal of Psychology*, *120*(2), 263–286. http://doi.org/10.2307/20445398

Kanuka, H., & Garrison, D. R. (2004). Cognitive Presence in Online Learning. *Journal of Computing in Higher Education Spring*, *15*(2), 21–39. http://doi.org/10.1007/BF02940928

Ke, F. (2010). Examining online teaching, cognitive, and social presence for adult students. *Computers & Education*, *55*(2), 808–820. http://doi.org/10.1016/j.compedu.2010.03.013

Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (2nd ed.). New York, NY: Routledge.

Kent, C., Laslo, E., & Rafaeli, S. (2016). Interactivity in online discussions and learning outcomes. *Computers & Education*, *97*, 116–128. http://doi.org/10.1016/j.compedu.2016.03.002

Keohane, N. O. (2013). Higher education in the twenty-first century: Innovation, adaption, preservation. *PS: Political Science & Politics*, *46*(1), 102–105. http://doi.org/10.1017/S1049096512001734

Khazaei, T., Lu, X., & Mercer, R. (2017). Writing to persuade: Analysis and detection of persuasive discourse. In *Proceedings of the iConference 2017* (pp. 203–215). http://doi.org/10.9776/17022

Kim, J., Kwon, Y., & Cho, D. (2011). Investigating factors that influence social presence and learning outcomes in distance higher education. *Computers & Education*, *57*(2), 1512–1520. http://doi.org/10.1016/j.compedu.2011.02.005

Kizilcec, R. F., Schneider, E., Cohen, G. L., & McFarland, D. A. (2014). Encouraging forum participation in online courses with collectivist, individualist, and neutral motivational framings. *eLearning Papers*, *37*, 13–22.

Kovanović, V., Gasevic, D., Hatala, M., & Siemens, G. (2017). *A novel model of cognitive presence assessment using automated learning analytics methods*. *Analytics for Learning (A4L) Network*.

Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., … Siemens, G. (2016). Towards automated content analysis of discussion transcripts: A cognitive presence case. *Learning Analytics and Knowledge (LAK'16)*, 15–24. http://doi.org/10.1145/2883851.2883950

Kozan, K. (2016). The incremental predictive validity of teaching, cognitive and social presence on cognitive load. *Internet and Higher Education*, *31*, 11–19. http://doi.org/10.1016/j.iheduc.2016.05.003

Kreijns, K., Van Acker, F., Vermeulen, M., & Van Buuren, H. (2014). Community of Inquiry: Social presence revisited. *E-Learning and Digital Media*, *11*(1), 5–18. http://doi.org/10.2304/elea.2014.11.1.5

Lai, E. R. (2011). Critical thinking: A literature review. *Pearson's Research Reports*, *6*, 40–41.

Lambert, J., & Fisher, J. (2013). Community of inquiry framework: Establishing community in an online course. *Journal of Interactive Online Learning*, *12*(1), 1–16.

Lee, S. M. (2014). The relationships between higher order thinking skills, cognitive density, and social presence in online learning. *Internet and Higher Education*, *21*, 41–52. http://doi.org/10.1016/j.iheduc.2013.12.002

Littlejohn, A., Hood, N., Milligan, C., & Mustain, P. (2016). Learning in MOOCs: Motivations and self-regulated learning in MOOCs. *Internet and Higher Education*, *29*, 40–48. http://doi.org/10.1016/j.iheduc.2015.12.003

Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distance Learning*, *14*(3), 202–227.

Major, C. H., & Blackmon, S. J. (2016). Massive Open Online Courses: Variations on a new instructional form. *New Directions for Institutional Research*, (167), 11–25. http://doi.org/10.1002/ir.20151

Martin, F., & Ndoye, A. (2016). Using learning analytics to assess student learning in online courses. *Journal of University Teaching & Learning Practice*, *13*(3), 1–20. http://doi.org/10.1177/0047239516656369

McLoughlin, D., & Mynard, J. (2009). An analysis of higher order thinking in online discussions. *Innovations in Education and Teaching International*, *46*(2), 147–160. http://doi.org/10.1080/14703290902843778

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. New York: Cambridge University Press. http://doi.org/10.1017/CBO9780511894664

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation. The Jossey-Bass higher and adult education series* (4th ed.). Hoboken, NJ: Jossey-Bass.

Meyer, K. A. (2003). Face-to-face versus threaded discussions: The role of time and higher-order thinking. *Journal of Asynchronous Learning Networks*, *7*(3), 55–65.

Momeni, E., Haslhofer, B., Tao, K., & Houben, G.-J. (2015). Sifting useful comments from Flickr Commons and YouTube. *International Journal on Digital Libraries*, *16*(2), 161–179. http://doi.org/10.1007/s00799-014-0123-1

Moore, R. L. (2014). Importance of developing community in distance education courses. *TechTrends*, *58*(2), 20–24. http://doi.org/10.1007/s11528-014-0733-x

Moore, R. L. (2016). Interacting at a distance: Creating engagement in online learning environments. In L. Kyei-Blankson, J. Blankson, E. Ntuli, & C. Agyeman (Eds.), *Handbook of Research on Strategic Management of Interaction, Presence, and Participation in Online Courses* (pp. 401–425). Hershey, PA: IGI Global. http://doi.org/10.4018/978-1-4666-9582-5.ch016

Nawrot, I., & Doucet, A. (2014). Building engagement for MOOC students: Introducing support for time management on online learning platforms. In *23rd International World Wide Web Conference (WWW '14)* (pp. 1–7). Seoul, South Korea.

Onah, D. F. O., Sinclair, J. E., & Boyatt, R. (2014). Exploring the Use of MOOC Discussion Forums. *Proceedings of London International Conference on Education*, 1–4.

Oztok, M., Zingaro, D., Brett, C., & Hewitt, J. (2013). Exploring asynchronous and synchronous tool use in online courses. *Computers & Education*, *60*(1), 87–94. http://doi.org/10.1016/j.compedu.2012.08.007

Paterson, L., & Goldstein, H. (2010). New statistical methods for analysing social structures: An introduction to multilevel models. In A. Skrondal & S. Rabe-Hesketh (Eds.), *Multilevel Modelling* (Vol. 1, pp. 11–18). Thousand Oaks, CA: SAGE Publications, Inc.

Pence, H. (2012). When will college truly leave the building: If MOOCs are the answer, what is the question? *Journal of Educational Technology Systems*, *41*(1), 25–33.

http://doi.org/doi: http://dx.doi.org/10.2190/ET.41.1.c

Pennebaker, J., Boyd, R., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. *Austin, TX: University of Texas at Austin*. Austin, TX: University of Texas at Austin.

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, *16*(3), 9–20. http://doi.org/10.24059/olj.v16i3.267

Raffaghelli, J. E., Cucchiara, S., & Persico, D. (2015). Methodological approaches in MOOC research: Retracing the myth of Proteus. *British Journal of Educational Technology*, *46*(3), 488–509. http://doi.org/10.1111/bjet.12279

Rodriguez, C. O. (2012). MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses. *European Journal of Open, Distance and E-Learning*, *2012*(1).

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (1999). Assessing social presence in asynchronous text-based computer conferencing. *International Journal of E-Learning & Distance Education*, *14*(2), 50–71. http://doi.org/Article

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 597–599. http://doi.org/10.22237/jmasm/1257035100

Schrire, S. (2004). Interaction and cognition in asynchronous computer conferencing. *Instructional Science*, *32*(6), 475–502.

Sharif, A., & Magrill, B. (2015). Discussion forums in MOOCs. *International Journal of Learnings, Teaching and Educational Research*, *12*(1), 119–132.

Shea, P., & Bidjerano, T. (2009). Community of inquiry as a theoretical framework to foster "epistemic engagement" and "cognitive presence" in online education. *Computers and Education*, *52*(3), 543–553. http://doi.org/10.1016/j.compedu.2008.10.007

Siemens, G. (2005). Connectivism: a learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, *2*(1), 3–10.

Simms, T., Ramstedt, C., Rich, M., Richards, M., Martinez, T., & Giraud-Carrier, C. (2017). Detecting cognitive distortions through machine learning text analytics. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 508–512). IEEE. http://doi.org/10.1109/ICHI.2017.39

Slotter, E. B., & Ward, D. E. (2015). Finding the silver lining: The relative roles of redemptive narratives and cognitive reappraisal in individuals' emotional distress after the end of a romantic relationship. *Journal of Social and Personal Relationships*, *32*(6), 737–756. http://doi.org/10.1177/0265407514546978

Song, M., & Yuan, R. (2015). Beyond social presence: Increasing cognitive presence through meaningful interaction. In *Global Learn 2015* (pp. 731–736). Berlin, Germany.

Steenbergen, M. R., & Jones, B. S. (2010). Modeling multilevel data structures. In A. Skrondal & S. Rabe-Hesketh (Eds.), *Multilevel Modelling* (Vol. 1, pp. 19–54). Thousand Oaks, CA: SAGE Publications, Inc.

Stevens, V. (2013). What's with the MOOCs? *The Electronic Journal for English as a Second Language*, *16*(4), 1–14.

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of Graduate Medical Education*, *4*(3), 279–282. http://doi.org/10.4300/JGME-D-12-00156.1

Sun, C., Li, S., & Lin, L. (2016). Thread structure prediction for MOOC discussion forum. In

W. Che, Q. Han, H. Wang, W. Jing, S. Peng, J. Lin, … Z. Lu (Eds.), *International Conference of Young Computer Scientists, Engineers and Educators* (Vol. 624, pp. 92–101). Harbin, China: Springer, Singapore. http://doi.org/10.1007/978-981-10-2098-8_13

Swan, K., Shea, P., Fredericksen, E. E., Pickett, A. M., & Pelz, W. E. (2000). Course design factors influencing the success of online learning. In *Proceedings of the WebNet 2000 World Conference on the WWW and Internet* (pp. 513–518). USA.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. http://doi.org/10.1177/0261927X09351676

Touati, A. (2016). Self-directed learning in MOOCs: A disconnect between theory and practice. *Middle Eastern & African Journal of Educational Research*, (19), 15–30.

Veletsianos, G., & Shepherdson, P. (2015). Who studies MOOCs? Interdisciplinarity in MOOC research and its changes over time. *The International Review of Research in Open and Distributed Learning*, *16*(3), 1–17. http://doi.org/10.19173/irrodl.v16i3.2202

Veletsianos, G., & Shepherdson, P. (2016). A systematic analysis and synthesis of the empirical MOOC literature published in 2013-2015. *International Review of Research in Open and Distributed Learning*, *17*(2), 198–221. http://doi.org/10.19173/irrodl.v17i2.2448

Vonderwell, S. (2003). An examination of asynchronous communication experiences and perspectives of students in an online course: A case study. *Internet and Higher Education*, *6*(1), 77–90.

Vrasidas, C., & McIsaac, M. S. (1999). Factors influencing interaction in an online courses. *American Journal of Distance Education*, *13*(3), 22–36.

Wang, Q., Woo, H. L., & Zhao, J. (2009). Investigating critical thinking and knowledge construction in an interactive learning environment. *Interactive Learning Environments*, *17*(1), 95–104.

Wang, X., Wen, M., & Rosé, C. P. (2016). Towards triggering higher-order thinking behaviors in MOOCs. *Proceedings of LAK16 6th International Conference on Analytics and Knowledge 2016*, 398–407. http://doi.org/http://dx.doi.org/10.1145/2883851.2883964

Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating how student ' s cognitive behavior in MOOC discussion forums affect learning gains. *Proceedings of the 8th International Conference on Educational Data Mining*, 226–233.

Webb, E., Jones, A., Barker, P., & van Schaik, P. (2004). Using e-learning dialogues in higher education. *Innovations in Education and Teaching International*, *41*(1), 93–103.

Wen, M., Yang, D., & Rosé, C. P. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? In *Proceedings of Educational Data Mining* (pp. 1–8).

Whipp, J. (2003). Scaffolding critical reflection in online discussions: Helping prospective teachers think deeply about field experiences in urban schools. *Journal of Teacher Education*, *54*, 321–333.

Wiebe, E., Thompson, I., & Behrend, T. (2015). MOOCs from the viewpoint of the learner: A response to Perna et al. (2014). *Educational Researcher*, *44*(4), 252–254. http://doi.org/10.3102/0013189X15584774

Williams, C., & D'Mello, S. (2010). Predicting student knowledge level from domain-independent functiona and content words. In *Proceedings of 10th International Conference on Intelligent Tutoring Systems* (pp. 62–71). http://doi.org/10.1007/978-3-

642-13437-1_7

Wise, A. F., Hausknecht, S. N., & Zhao, Y. (2014). Attending to others' posts in asynchronous discussions: Learners' online "listening" and its relationship to speaking. *International Journal of Computer-Supported Collaborative Learning*, *9*(2), 185–209. http://doi.org/10.1007/s11412-014-9192-9

Wong, J.-S., Pursel, B., Divinsky, A., & Jansen, B. J. (2015). An analysis of MOOC discussion forum interactions from the most active users. In N. Agarwal, K. Xu, & N. Osgood (Eds.), *Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 452–457). Switzerland: Springer International Publishing. http://doi.org/10.1007/978-3-319-16268-3_58

Wong, J.-S., Pursel, B., Divinsky, A., & Jansen, B. J. (2016). An analysis of cognitive learning context in MOOC forum messages. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16* (pp. 1315–1321). San Jose, CA: ACM Press. http://doi.org/10.1145/2851581.2892324

Yang, D., Wen, M., Kumar, A., Xing, E. P., & Rosé, C. P. (2014). Towards an integration of text and graph clustering methods as a lens for studying social interaction in MOOCs. *International Review of Research in Open and Distance Learning*, *15*(5), 215–234.

Yang, Q. (2014). Students motivation in asynchronous online discussions with MOOC Mode. *American Journal of Educational Research*, *2*(5), 325–330.

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). London: SAGE Publications.

You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *Internet and Higher Education*, *29*, 23–30.

http://doi.org/10.1016/j.iheduc.2015.11.003

Yuan, L., & Powell, S. (2013). *MOOCs and open education: Implications for higher education*.

APPENDIX

Table 26. Means and Effect Sizes of Dependent Variables, by Course

	Ancient Greek Hero		Super-Earths		Visualizing Japan	
	Instructor	Self	Instructor	Self	Instructor	Self
Cognitive Processing	*M* = 11.72, *SD* = 8.56	*M* = 11.88, *SD* = 7.42	*M* = 13.42, *SD* = 14.16	*M* = 11.54, *SD* = 8.42	*M* = 12.84, *SD* = 6.87	*M* = 12.65, *SD* = 7.21
Difference	Not significant		**Very small to small (*d* = 0.14)**		Not significant	
Insight	*M* = 3.34, *SD* = 4.17	*M* = 3.55, *SD* = 3.96	*M* = 5.07, *SD* = 13.16	*M* = 3.14, *SD* = 4.01	*M* = 3.98, *SD* = 4.05	*M* = 3.92, *SD* = 3.66
Difference	**Very small to small (*d* = 0.05)**		**Very small to small (*d* = 0.15)**		Not significant	
Causation	*M* = 1.31, *SD* = 2.25	*M* = 1.43, *SD* = 2.14	*M* = 1.35, *SD* = 2.66	*M* = 1.58, *SD* = 2.73	*M* = 2.07, *SD* = 2.87	*M* = 1.95, *SD* = 2.56
Difference	**Very small to small (d = 0.05)**		**Very small to small (*d* = 0.10)**		Not significant	
Discrepancy	*M* = 1.41, *SD* = 2.59	*M* = 1.35, *SD* = 2.08	*M* = 1.33, *SD* = 2.67	*M* = 1.67, *SD* = 2.50	*M* = 1.49, *SD* = 2.21	*M* = 1.29, *SD* = 1.99
Difference	Not significant		**Very small to small (*d* = 0.13)**		**Very small to small (*d* = 0.10)**	
Tentativeness	*M* = 2.68, *SD* = 3.62	*M* = 2.73, *SD* = 3.48	*M* = 2.49, *SD* = 3.81	*M* = 3.01, *SD* = 4.28	*M* = 2.85, *SD* = 3.18	*M* = 2.80, *SD* = 3.28
Difference	Not significant		**Very small to small (*d* = 0.13)**		Not significant	
Certainty	*M* = 1.75, *SD* = 4.03	*M* = 1.45, *SD* = 2.69	*M* = 2.20, *SD* = 3.92	*M* = 1.48, *SD* = 2.69	*M* = 1.29, *SD* = 2.28	*M* = 1.36, *SD* = 3.58
Difference	**Very small to small (*d* = 0.08)**		**Very small to small (*d* = 0.20)**		Not significant	
Differentiation	*M* = 2.82, *SD* = 3.54	*M* = 3.02, *SD* = 3.25	*M* = 2.35, *SD* = 3.39	*M* = 2.43, *SD* = 3.24	*M* = 2.65, *SD* = 2.96	*M* = 2.77, *SD* = 3.15
Difference	**Very small to small (d = 0.06)**		Not significant		Not significant	

Table 27. Means and Effect Sizes of Post Characteristics, by Course

	Ancient Greek Hero		Super-Earths		Visualizing Japan	
	Instructor	Self	Instructor	Self	Instructor	Self
Post Position	*M* = 27.43, *SD* = 61.97	*M* = 20.07, *SD* = 29.66	*M* = 934.34, *SD* = 1397.69	*M* = 197.49, *SD* = 229.02	*M* = 40.94, *SD* = 39.16	*M* = 3.40, *SD* = 5.68
Difference	**Very small to small (d = 0.13)**		**Medium (d = 0.60)**		**Large to very large (d = 1.12)**	
# of Comments	*M* = 2.14, *SD* = 7.33	*M* = 3.56, *SD* = 5.61	*M* = 1.52, *SD* = 4.58	*M* = 1.34, *SD* = 1.79	*M* = 0.56, *SD* = 0.98	*M* = 0.39, *SD* = 1.18
Difference	**Very small to small (d = 0.10)**		Not significant		**Very small to small (d = 0.11)**	
Word Count	*M* = 78.39, *SD* = 125.75	*M* = 84.04, *SD* = 99.92	*M* = 47.44, *SD* = 69.67	*M* = 54.77, *SD* = 68.22	*M* = 60.00, *SD* = 50.01	*M* = 53.00, *SD* = 42.07
Difference	**Very small to small (d = 0.05)**		**Very small to small (d = 0.11)**		**Small (d = 0.20)**	
Words per Sentence	*M* = 15.64, *SD* = 11.67	*M* = 17.12, *SD* = 9.85	*M* = 12.57, *SD* = 9.13	*M* = 14.38, *SD* = 9.92	*M* = 19.98, *SD* = 10.50	*M* = 19.91, *SD* = 9.66
Difference	**Very small to small (d = 0.12)**		**Small (d = 0.20)**		Not significant	
# of Six-Letter Words	*M* = 18.74, *SD* = 10.34	*M* = 19.54, *SD* = 8.90	*M* = 20.29, *SD* = 14.84	*M* = 20.25, *SD* = 10.11	*M* = 24.25, *SD* = 9.69	*M* = 24.85, *SD* = 9.71
Difference	**Very small to small (d = 0.10)**		Not significant		Not significant	
Analytical Thinking	*M* = 62.55, *SD* = 30.97	*M* = 66.19, *SD* = 28.95	*M* = 54.18, *SD* = 32.07	*M* = 59.33, *SD* = 30.44	*M* = 72.86, *SD* = 26.46	*M* = 74.59, *SD* = 26.14
Difference	**Very small to small (d = 0.12)**		**Small (d = 0.20)**		Not significant	