THE UNIVERSITY OF CHICAGO

The Effects of Innovative Online Learning Features on Students' Educational Outcomes

By

Xin (Liz) Feng

June 2020

A paper submitted in partial fulfillment of the requirements for the
Master of Arts degree in the
Master of Arts Program in the Social Sciences

Advisor: Stephen Raudenbush

Preceptor: Amit Anshumali

Abstract

As the use of online learning resources to provide and support instruction is on the rise, a major question for scholars and educationists to address is how we can use technology to enact more effective instruction in this new online environment rather than designing online courses based on the traditional classroom model. In learning science and practice, one general challenge is how to reliably determine what are the most effective methods for supporting learning and under what circumstances do those methods work. While learning scientists are used to conduct randomized experiment trials in schools to test different education theories, the massive amount of data being collected from educational technologies may provide us an opportunity to generate new insights. In this paper, I measure the effectiveness of two innovative instruction features of an online course – the Open Learning Initiative (OLI) environment and the discussion forum – to test the application of cognitive and social cognitive learning theories in the online environment. we use a hierarchical linear model to evaluate the impact of students' use of OLI features on their learning outcomes and get a positive result for the effectiveness of OLI. After conducting computational content analysis and centrality measurement on the discussion forum, we find that academic discussion plays an important role among students' conversations online.

INTRODUCTION

When the free online course on artificial intelligence offered by Professor Andrew Ng at the Stanford University attracted more than 580,000 students around the globe in 2011, people became excited about a foreseeable future where advanced knowledge will be distributed to a larger population at a much lower cost (Markoff 2011). However, besides the scale of audience, better education is also about improving the quality of instruction. While there are notable stories of online courses that were claimed to be more effective than traditional classroom-based instruction, the more typical situation is that, at best, online courses achieve the same outcomes Chirikov 2020). The question to address in the development of online education is how we can use technology to enact instruction in this new online environment so that we can arrive at the same or better outcomes as the in-person instruction.

Most of the Massive Open Online Courses (MOOCs) are still based on the traditional teacher-centered classroom model. This dominant approach to teaching is based on behaviorism, developed from pedagogical research by Thorndike (1911), Pavlov (1927), and Skinner (1971). In a behaviorist classroom, teachers follow a stimulus-response framework that relies heavily on repetition and positive reinforcement such as praise, good grades, and prize. This method has proven successful in areas where there is a “correct” response or easily memorized material (Skinner 1976). However, this approach is designed to teach small classes of students with similar background knowledge and future goals. When the context changes to thousands of students coming from different educational backgrounds, we need to reevaluate the method of designing an online course (Strader 2012; Ferguson 2019).

In addition to the teacher-centered behaviorist theory, cognitivist learning theory has been on the rise since the mid-1950s (Yilmaz 2011). This shift is due to the fact that behaviorists fail to explain the mental process of learning and advances in technology offer new opportunities for creative implements of psychology theories (Atkins 1993, Deubal 2003). Cognitive learning theories try to explain how knowledge is acquired, processed, stored, retrieved, and activated by learners during the different phases of the learning process (Anderson, Reder, and Simon 1997). Within the cognitive school, researchers majorly focus on two main areas: the individual cognitive trends derived from Piaget's research and the sociocultural trend based on Vygotsky's work (Deubel 2003; Duffy and Cunningham 1996; Gillani 2003). Piaget's theory describes and explains the changes in logical thinking of children and adolescents while Vygotsky's theory focuses on the role of cultural and social interaction in students' learning process.

In this paper, we are interested in two innovative online course features - OLI and discussion forum. The Open Learning Initiative (OLI) is based on the cognitive framework in the line of Piaget's. It is a learning environment developed by Carnegie Mellon University based on the cognitive principles of problem-solving (Koedinger 2016). On the other hand, the discussion forum is based on the social cognitive theory developed from Vygotsky's (Dommett 2019). It is an asynchronous forum that allows students to discuss logistic and academic issues online (Marra 2004). The comparison between the OLI and the discussion forum is meaningful to pedagogical research in online education. They represent teaching methods different from the behaviorist's classroom model. By testing their effectiveness, we examine the implementation of two dominant theoretical frameworks of cognitive learning science in an online learning environment. Next, we are going to discuss different measurements of the effectiveness of the OLI and the discussion forum.

The goal of the study is to test the effectiveness of the online course features. Specifically, how the OLI feature is helping students improve their score and how the discussion forum is increasing students' willingness to participate in academic discussions online. Different cognitive approaches to online learning design lead to different definitions of effective education. In our research of OLI, we are going to use grades as the measurement of effective learning. According to Piaget's school of cognitivism, students' cognitive development is transforming with time. Thus, we need a longitudinal measurement of students' performance. Grades are collected in between a certain period of time and it is a relatively objective measure of learning outcomes. Despite the common criticism of grades for their lack of reliability and common standards, it is the most appropriate measurement in our research context (Caspersen 2017). For the research on the discussion forum, we are going to use the centrality in network theory as a measurement of effective learning. Vygotsky's school of social cognitivism considers effective education coming from interactions among peers. The discussion forum is designed to encourage students to learn from each other through conversations (Marra 2004). Thus, to test the effectiveness of the discussion forum, we are evaluating whether students are willing to engage in academic discussion online. Centrality measures things that can be transmitted and diffused through certain paths in a network. For example, it can measure the influence of a single person's attitude in an organization or the importance of certain words in a corpus. The higher the centrality, the larger the unit's influence on the whole network (Borgatti 2005). We measure the centrality of words in students' online conversations to see whether academic discussion is influential in the virtual community of students.

The dataset of this research comes from the DataShop – the world's largest repository of learning interaction data. It is collected from a 12-week online course *Introduction to Psychology as a Science* offered collaboratively by Georgia Institute of Technology and Carnegie Mellon University (CMU) via the Coursera platform. This online course provided general course structure such as registration and syllabus, video lectures and slides, discussion forum, writing assignments, quizzes, and a final exam. On top of that, it incorporates the OLI learning environment into its course. Students voluntarily sign up for "MOOC only" or "MOOC + OLI" learning environment. This setup makes the dataset suited for quasi-experiment research where the “MOOC only” group is the control group, “MOOC + OLI” group is the treatment group and the OLI is the treatment. In our dataset, we have 753 “MOOC+OLI” students and 137 “MOOC only” students who complete the entire courses by taking all the quizzes and the final exam. To test the effectiveness of OLI, we employ a two-level hierarchical linear model and find out that students do get better learning outcomes when they use OLI. Besides students’ test scores, we also have 3864 discussion forum’s posts in our dataset. To test the effectiveness of the online discussion forum, we use content analysis methods to transform the text data. By conducting network analysis, we find that academic discussion plays an important role among students’ conversations online.

In part I of this paper, I lay out the previous literature on the cognitive theory, the social cognitive theory in online course design, and the testing of effectiveness in an online learning environment. In Part II, I describe the data in my analysis and outline the hierarchical linear model and the computational content analysis method I use to assess the effectiveness of OLI and discussion forum course features. In Part III, I present the result of my analysis. Part IV will summarize the findings of my study and exploring how my research could lead to further discussion in online education research in the future.

PART I: LITERATURE REVIEW

Learning theories are fundamental for effective online teaching because they shed light on different aspects of the learning process. The advance of computer technology provides us an opportunity to systematically incorporate advances in learning science into the virtual classroom and to test different theories at the same time. In this paper, I test the effectiveness of two innovative online course features – OLI and the discussion forum. The science behind their design comes from the cognitive school of learning science.

Cognitive Theory & The Science Behind OLI

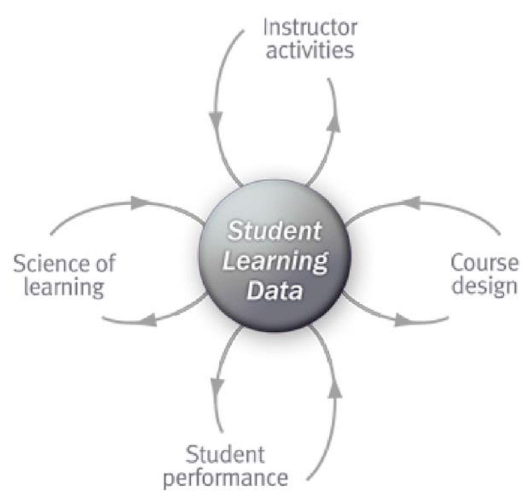
The largest difference between cognitivism and behaviorism is that behaviorism focuses on monitoring the behavior while cognitivism concentrates on tracing the conceptual change and the mental activity of learning. The cognitive school views (1) learning as an active process "involving the acquisition or reorganization of the cognitive structures through which humans process and store information" and (2) the learner as an active participant in the process of knowledge acquisition and integration (Good and Brophy 1990, 187; Merriam and Caffarella 1999, 254; Simon 2001, 201). This implies that acquiring knowledge is a mental activity that involved internal coding and structuring by the learner (Derry 1996, Spiro et al. 1992). Thus, effective instruction is based on cognitive science that aligned with a student's knowledge acquisition process. Jean Piaget – the leading scholar of the cognitive school – pushes the theoretical framework one step forward. He argues that children are born with a very basic mental structure, genetically inherited and evolved, on which all subsequent learning and knowledge is based (Gillani 2003). This assertion challenges the behaviorist's view of effective education as teacher-centered and leads to the development of children-centered classroom and individualized education.

Since the mid-1980s, education scientists had been trying to put cognitive school's theories into practice. They found inspiration in tutoring practice. One-to-one tutoring of expert human tutors is more effective than typical one-to-many classroom instruction, however, it is not practical to assume that every child could have an individual tutor (Bloom 1984). Harnessing the advancement of technology, researchers aim to develop a computing tutor based on cognitive theories so that one-to-one tutoring can be realized in a virtual environment. John R. Anderson and colleagues coded cognitive theories as computational models and borrowed the method from artificial intelligence to create the first Intelligent Tutoring System (Newell & Simon 1972, Anderson, Boyle, & Reiser 1985). This system improved students' learning experience relative to conventional teacher-centered instruction (Corbett 1995) Based on this original system, researchers developed computer tutor for mathematics teaching and was in use in more than 2000 schools across the United States in 2004 and 2005 (Koedinger 2006).

These initial trials of individualized education that incorporated cognitive learning theories into the design of online learning lay the foundation for the OLI learning environment. However, to achieve an optimal result, researchers need to embrace Piaget's theory - children are born with a genetically determined mental structure – into the design. In 2006, CMU developed a cognitive tutor that employs a simple Bayesian method of estimating the student's knowledge and assigns each individual a prior state (Koedinger 2006). This prior state is the default setting of a student's mental structure. It represents a stage when a student hasn't started the learning process. Based on this state, the cognitive tutor selects appropriate practice problems for students and collect a database of information on individual students' performance. The system can make inferences about any change in students' knowledge structure and suggest additional exercises that could remediate any apparent knowledge gap (Meyer 2002). Furthermore, for each practice problem, the cognitive tutor runs a cognitive model step-by-step along with the student to follow the student's individual path in solving a complicated problem (Koedinger 2006). This generated tailored feedbacks that accurately point out students' weakness in the cognitive process of problem-solving and knowledge integration.

In 2012, Carnegie Mellon University put together a team of scientists and improved the functioning of the cognitive tutor. The result was a new online course feature the Open Learning Initiative that provided a learning environment that is student-centered and based on cognitive theory besides the MOOC learning environment designed by Coursera. Instruction on every concept starts with one or more observable learning objectives. Then, it presents expository content in the form of text, images, simulations, short videos, and worked examples. In between the exposition are interactive tasks that support students to engage in authentic practice with the concepts and skills they are learning. The tasks are presented in a supported environment with hints available to the students if they are struggling. They receive feedback that reinforces correct responses and targets common student misconceptions (Strader 2012). The functioning of OLI is shown in Graph 1. The OLI feature was first tested on a statistics course and has proved to improve students' learning outcomes (Koedinger 2016). In this study, we are testing the OLI feature for a psychology course. We want to see whether the effects of OLI could be generalized to any course subject.

Graph 1: OLI Functioning Loop (The Open Learning Initiative)



Social Cognitive Theory & the Science Behind Discussion Forum

The cognitive school of learning science could be divided into two branches. Besides the cognitive theory developed from Piaget's work, the social cognitive theory derived from Vygotsky's research also has a great influence on online course design. In the social cognitive view, learning is neither driven by inner cognitive forces nor buffeted helplessly by environmental influences. Vygotsky advocated that learning should be a social process. His study of learning concentrated on the interplay between the individual and society, and how social interaction and language come into play in affecting learning (Fosnot 1996; Gredler 1997; Jarvis, Holford, and Griffin 2003).

By the 1960s, the social cognitive theory was further developed by psychologist Albert Bandura. He listed out many mechanisms through which individuals may learn from others. This includes how individuals learn from each other through direct communications. Moreover, he posited the theory of observational learning – people learned through observation, imitation, and modeling. They pay more attention to examples that are attractive, similar to them, or prestigious and are rewarded for their behaviors (Bandura 1971). One of the implications of Bandura's theory is that students adopt positive learning methods from role models in the student body. In the long run, this could change the culture of a learning community. This theory lays the foundation for our evaluation of the discussion forum. When we evaluate the most influential talk on the discussion forum, we are assuming students learn from observing it and cater to its topic in their discussion.

The significance of learning community in online learning is highlighted in many studies. Developing a sense of learning community among distance learners is especially important as distance learners by nature work in isolation (Mohamad 2014). In a face-to-face environment, students construct experiences and knowledge via synchronous, interactive discussions and problem-solving sessions. Web-based learning courses have limited opportunities and thus tend to rely on asynchronous online discussion forums to create them (Garrison, Anderson & Archer 2001). The type of forum mentioned here is the primitive one where open learners participate freely on a loosely guided agenda. The design of the discussion forum is aligned to social cognitive learning theories including collaborative learning, resource-based learning and problem-based learning (Hammond 2005; Macdonald and Twining 2002) and conversational learning (Laurillard 1999). Forum activity is a good indicator of students' level of engagement as well as a good predictor of academic performance (Macfadyen & Dawson 2010; Dawson 2010). Even though an online discussion forum does not provide face-to-face communication, the value of written communication as used in online discussion forums comes from the necessity of preciseness, organization of thought, and clear expression (McCreary 1990).

Although researchers acknowledge the critical role the online discussion forum plays in online courses, the empirical evidence to indicate that text-based communication can facilitate learning is just emerging (Wang 2015). One reason for the relatively small number of studies addressing meaningful learning via online discussion forums may be that not enough research has been done using computational content analysis in this domain. In this study, we are going to use the computational content analysis method to see the role of academic discussion in the online learning community.

Testing of Effectiveness in Online Learning Environment

Testing the effectiveness of learning methods and under what circumstances do these methods work has always been a challenge for education researchers. Evaluation of educational innovations has evolved from qualitative method to quantitative inference and eventually to computational implications. Before 1990s, educational program evaluation were predominantly surveys, small-scale, in-depth qualitative case studies (Cook 2002). In 1950s, Frederick Mosteller and Howard Hiatt established the randomized trial as a foundation for causal inference in medicine and thus ushered in the era of quantitative evaluation method. In 1999, the American Academy of Arts and Sciences organized a conference and advocated for more randomized controlled trials (RCTs) in education (Raudenbush 2020). Since then, randomized controlled experimentation has been considered a gold standard for determining whether a method of learning is effective.

The RCT is a type of scientific experiment with a control group and a treatment group. The treatment group will be given a treatment. The difference between the outcomes of the control group and the treatment group is the treatment effect**.** Drawing from the treatment effect, researchers can make causal claims. Experiments provide strong internal validity[1] for causal inference but may not provide external validity for generalization of the method to contexts not samples in that experiment. Analysis of associations of method and outcome threatens internal validity but doing so across many naturally-occurred contexts provides external validity[2] at a much lower cost than doing experiments in all these contexts.

RCTs of education research does not usually provide evidence on whether the learning method will generalize to other courses and course contexts (Koedinger 2016). Classroom studies can add ecological validity by being performed with real students and in the context of real courses. Even here the few experiments in a limited number of course contexts may leave a critic wondering whether these testing effects generalize across all course content or might be limited to certain kinds of content or contexts. In fact, many of these studies have focused on the learning of facts and verbally communicated content. Perhaps the testing effect is less relevant to the learning of skills or principles.

It should be clear that determining causal relationships is important for scientific and practical reasons because causal relationships provide a path toward explanatory theory and a path

[1] Internal validity refers to the degree of confidence that the causal relationship being tested is trustworthy and not influenced by other factors or variables.
[2] External validity refers to extent to which results from a study can be applied (generalized) to other situations, groups or events.

toward reliable and replicable practical application. We need evidence for causes that generalize across a wide variety instructional contexts and course content. If we can be increasingly certain a learning method is causally related to more optimal learning across a wide variety of contexts and content, then that method should be used to guide course design and students should be encouraged to use it. Coupling evidence from both experiments and analysis of naturally occurring high-volume data appears an effective way to increase generalizable certainty (Koedinger 2016).

Causal inference makes deductive claims while in recent years, we see more and more inductive research design in social science. This is partially due to the hype triggered by a larger amount of data available on the internet and increasingly sophisticated computing algorithms. Computational social science has almost influenced all subject fields in social science (Peng 2019). This is the same for education research. Education data mining, learning analytics, and computational educational research all refer to this new trend. However, it also has acquired a bad reputation in the social sciences. Many researchers see it as synonymous with the practice of using algorithms to sift through data for associations and then falsely reporting them as if confirmations of single-test hypotheses (Evans 2016). Thus, computational social science, more than ever, needs rigorous theorization, research design, statistical analysis, and results presentation. In education research, we should be more careful about implement statistically accountable data mining on education data. More often, we still need RCTs as guidance for designing the research strategy.

PART II: METHOD, STRATEGY & DATA

In this part of the paper, I am going to elaborate on the methods I use to test the effectiveness of the OLI and the discussion forum. First, I will briefly introduce the hierarchical linear model, the computational content analysis, and the network centrality measurement. I am going to explain their conceptual framework and why it is appropriate to use these methods in my research. Next, I will give an exploratory analysis of my data – the variables I am using, basic patterns I find in early exploration and some sacrifices I make because of the lack of access to data.

METHODS

Hierarchical Linear Model

Hierarchical linear models are also referred to as multilevel linear models in sociological research and mixed-effects models or random-effects models in biometric applications. Researchers oftentimes use it on a nested data structure. For example, we can have social data of persons nested within classrooms, and the classroom units themselves are nested within schools. With hierarchical linear models, each of the levels in this structure has its submodel. These submodels describe relationships among variables within a given level and specify how variables at one level influence relations occurring at another. (Raudenbush and Bryk 2002) Besides the nested data describe above, a hierarchical linear model can also be applied to analyze individual change over time. In this scenario, the multiple observations on each individual are seen as nested with in the person. This is called the linear growth model. To be more specific, let's look at the two-level growth model formula below.

Level-1 Model

$$Y_{ti} = \beta_{0i} + \beta_{1i}*(a_{ti}) + e_{ti}$$

Level-2 Model

$$\beta_{0i} = \gamma_{00} + \gamma_{01}*(X_i) + u_{0i}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}*(X_i) + u_{1i}$$

The goal of this model is to assess each person's rate of learning and predict future learning status. We can compare the growth rate of each individual. At level 1, each person's development is represented by an individual growth trajectory that depends on a unique set of parameters. In our level-1 model above, a_{ti} could be viewed as a person's age or other time measurements for person i at time t. β_{0i} is the true ability of person i at $a_{ti} = 0$. β_{1i} is the growth rate for person i over the data-collection period and represents the expected change during a fixed unit of time. Y_{ti} is the observed status at time t for individual i. e_{ti} is the random error of level 1. We assume the error is independently and normally distributed with a mean of 0 and constant variance σ^2. At level 2, individual growth parameters β_{0i}, β_{1i} become the outcome variables, where they may depend on some person-level characteristics. X_i is either a measured characteristic of the individual's background (e.g., sex or social class) or of experimental treatment (e.g., type of curriculum employed). γ_{00} is the mean of β_{0i} and γ_{10} is the mean of β_{1i}. γ_{01} is the effect of X_i on the parameter β_{0i}. γ_{11} is the effect of X_i on the parameter β_{1i}. u_{0i} and u_{1i} are random effects with a mean of 0.

This model is suited for our data. We want to evaluate individual's change over the course period. The quiz scores and final exam scores are collected during a fixed time period. Our level 1 model will be a set of growth trajectories that showcase student scores or learning trends. In the level 2 model, our X_i will be the treatment OLI. From this model, we can see OLI's effect on students' learning outcomes. A linear growth model is most appropriate for situations when the number of observations per individual is few (e.g., three or four occasions). We have 12 occasions (11 quizzes and 1 final score) in our data. This will cause low reliability of prediction in the result. Besides that, our data meets the conceptual framework of a two-level linear growth model.

Computational Content Analysis

We live in an online social world nowadays. We generate electronic text data when we are sharing our thoughts on social media, messaging our friends, and participating in online transactions. Digitized libraries have also become our major reading resources. This supply of text has elicited demand for natural language processing (NLP) and machine learning tools to filter, search, and translate text into valuable data.

The history of text analysis in social science could be traced back to 1910 when Max Weber proposed a large-scale analysis of the German press to trace temporal shifts in cultural values. Later, sociologists engaging in content analysis often begin by qualitatively coding text to meaningful categories. In the last decade, however, statistical NLP has become dramatically more accurate and powerful at recovering linguistic structures and semantic associations. Computation becomes capable of augmenting our fine perception of patterns in language and their links to the social world beneath (Evans 2016).

In computational content analysis, a document is a “bag of words”. We called our unit of analysis "token". It can be referred to as individual words or meaningful phrases. For example, the sentence "a good wine is a wine that you like" contains nine tokens. Words or tokens are often stemmed from related roots (e.g., pray, prayer, and prayed all collapse to pray). Richer tokenization also becomes possible by including frequent n-grams, which are common word sequences of length n. For example, a bi-grams token may a two-word phrase such as “peer review” or “off-topic”. Each document can then be represented as a sparse vector of counts for each token in the vocabulary, with such counts often normalized by the number of tokens within the document (Evans 2016). When we have our text data in hand, the first step is usually to tokenize and normalize the data for later analysis.

After transforming our text data (a document or a corpus) into a bag of tokens, researchers may study the patterns of communication or underneath social relationships using supervised or unsupervised machine learning methods. In our study, we are only concerned about unsupervised methods. These methods usually begin with a corpus of unannotated text and then discover novel structures for interpretation (Evans 2016). Four of the most common: clustering, network analysis, topic modeling, and vector space embedding. We will be focusing on network analysis for our research purpose.

Network Analysis & Centrality

Social network analysis (SNA) is the use of network theory to analyze social networks. This analysis applied network theory to social relationships consisting of nodes which represent the individual actors in the network and ties or edged depicting the relationship between the individuals. The network is then visualized as a social network diagram where the nodes are represented as dots or circles and the ties as lines (Suraj 2015). The unit of social network analysis is oftentimes a person or an organization. To apply social network techniques in the content analysis framework, researchers develop the semantic network approach. A semantic network links words or phrases collocated within documents, sentences, clauses, and dependency parse trees. This approach can be considered the fine-grained corollary of document clustering. This unsupervised and largely model-free approach reveal the fine structure of cognitive and cultural associations between entities through calculations of their network positions, such as word centrality, influence, structural equivalence, and constraint (Carley 1993, Schank & Colby 1973, van Atteveldt 2008).

Centrality is one of the most studied concepts in network analysis. In graph theory, indicators of centrality identify the most important vertices within a graph. In social network theory, we use centrality to identify the most influential person in a network. Numerous measures of centrality have been developed, including degree centrality, closeness, betweenness, eigenvector centrality, and information centrality, etc. The difference between these centralities is the implicit assumptions about the manner in which things flow in a network (Borgatti 2005). For semantic networks, we usually use eigenvalue centrality, betweenness centrality, and closeness centrality because they are good at representing how word influences the whole corpus.

DATA

The secondary dataset of the online course *Introduction to Psychology as a Science* is drawn from the DataShop, a repository of educational data at the Carnegie Mellon University. The dataset is quasi-experimental and includes pre-test and post-test. It has a treatment group (students participate in both OLI features and MOOC) and a control group (students who only use MOOC features). This dataset also has a student background survey which compensate for the wide variety of background of students and reduce the confounding effect.

For the OLI evaluation, I select the following variables to build a 2-level growth model.

- **Level-1 Outcome variable (*QUIZCORR*)**: quiz scores and final exam scores, the correct percentage each student gets on an exam. Its numeric value is between 0 and 100.
- **Level-1 predictor (*QUIZNUM*)**: the number of quiz, we are using it as a time measurement since each quiz is taken in a seven days roll. We have pretest, eleven quizzes, and one final exam. Its numeric value is between 0 and 13. 0 refers to the pretest, 1 to 12 refers to the eleven quizzes and 13 refers to the final exam.
- **Level-2 predictor (*GROUP*)**: group is a binary variable. It has two values. When group = 1, this means the student is in the treatment group (MOOC + OLI). When group = 0, this means the student is in the control group (MOOC only).
- **Pers**: this variable doesn't show up in our final model but we use it to denote each student. We end up having 890 students' records in our dataset. Its numeric value is 0 to 890.

For our dataset, there are initially 9075 students, who registered to use the OLI and MOOC materials. MOOC only students amount to 18645. However, because this is an online course and many students drop out in the process, among students who took the final exam, we have 939 MOOC+OLI students and 215 MOOC only students. The ideal situation is to get data for each student. However, we only have access to data of students who took the final exam in the data repository. It is difficult to get access to the whole dataset when researchers are using private instead of public data. Because of this lack of access to data, our result is influenced by selection bias. Students who don't drop out until the end of the course are those who are diligent and devoted. Thus, when we are analyzing the OLI treatment effect, we would not be sure whether their progress in this course is due to the OLI or their diligent nature. After cleaning the dataset, we notice that there are some students who miss some quiz scores. We clean those data out because we want to compare students' growth without too much noise. The final dataset we use consists of 753 MOOC+OLI students and 137 MOOC only students.

For the discussion forum research, we use the 3864 posts of 491 users in our dataset

Table 1: Glimpse of the Data

post number	comment	post	code	thread starter	previous post
1	0	Hi! My name is Julia, I am from Spain and...	O	My name is NAME509825 NAME509825. I'm 17 and w...	Hello My name is NAME1017864 I am From Utah. G...
1	0	The Quiz is still open. Once it closes tomorro...	A2	I did not give the correct answer for these tw...	I did not give the correct answer for these tw...
1	0	This assignment was like bowling with a curtai...	A2	My experience with this course so far has been...	In my humble opiniÃ³n "the data would be analy...
1	0	I cannot find a link to input my written assig...	A2	Hi, I did not understand how assignments works...	Hi, I did not understand how assignments works...
1	0	My "theory" is that it's just a technical prob...	A2	the grader marked question 9's right answer of...	I experienced the same thing - can ysomebody c...
1	0	Oh, and Christine, You were right t...	O	I've done several Coursera classes and I'm int...	Christine and Eric, You two, and on...
1	0	I am not involved in any age gap relationship,...	A2	According to the info in OIL; younger couples ...	It will end up in cheating by 20 on 40. J...
1	0	Will the final exam be timed? 	A2	Hello ! I wonder what means: <span>F...	Great cause I saw it and was wondering <b...

In table 1, we are going to use the post column for content analysis. Thread starter of table 2 is the discussion thread the post is categorized under. Previous post is the post before current post.

For an exploratory data analysis, I am implementing counting words and phrases method on my corpus. First, we can look at the superlative adjectives. Here, we see that some of the most frequently appeared adjectives include “good” and “great”. That is a good sign. This shows that students have a positive feedback for this course in general.

Table 2: Frequently Appear Adjective

```
FreqDist({'other': 432, 'good': 376, 'final': 333, 'same': 314, 'first': 310, 'many': 304, 'd
ifferent': 191, 'great': 186, 'oli': 172, 'right': 156, ...})
```

Then, we can turn to look at the 5 most frequently appeared noun. These are “courses”, “assignment”, “time”, “%”, and “psychology”. Notice here that “%” is simply a result of dirty data. I have cleaned the data. However, because these data are directly scraped from the online forum, they contain many html language codes, making it nearly impossible to create a perfectly clean dataset. Words such as “courses”, “assignment” and “psychology” suggests that my corpus is quite academic in terms of users’ linguistic patterns.

Table 3: 5 Most Frequently Appear Noun

```
[('course', 1021),
 ('assignment', 508),
 ('time', 455),
 ('%', 448),
 ('psychology', 396)]
```

PART III: RESULT

Two-Level Growth Model for the OLI

Level-1 Model

$$QUIZCORR_{ij} = \beta_{0j} + \beta_{1j}*(QUIZNUM_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(GROUP_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}*(GROUP_j) + u_{1j}$$

In the level-1 model, $QUIZCORR_{ij}$ is the outcome variable of each student, representing the score of person j at quiz (exam) i. $QUIZNUM_{ij}$ is the time variable, it is the quiz number of person j. Because the first quiz is the pretest, which is taken at week 0, $QUIZNUM_{ij}$ could also be viewed as the amount of time in weeks that had elapsed from the first data-collection point since the quiz is given out seven days in a roll. β_{0j} represents the pretest score. It can be interpreted as the true ability of student j at $QUIZNUM_{ij} = 0$. β_{1j} is the growth rate of student j over the data collection period.

In the level-2 model, $GROUP_j = 0$ if students only use MOOC and $GROUP_j = 1$ if students use both MOOC and OLI features. γ_{11} represents the effect of $GROUP_j$ on the student j's growth parameter. By identifying β_{1j}, we will have a growth curve for each student j. Also, knowing γ_{11} will help us understand how does using interactive learning feature OLI affect students' educational outcomes.

Table 4: Reliability Result

Random level-1 coefficient	Reliability estimate
INTRCPT1, β_0	0.644
QUIZNUM, β_1	0.173

As we can see from table 5, intercept β_0 has a relatively good reliability of 64.4%. However, the reliability of β_1 is only 17.3%.

Table 5: Coefficient Result

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	72.104986	1.471067	49.015	888	<0.001
GROUP, γ_{01}	3.364898	1.533953	2.194	888	0.029
For QUIZNUM slope, β_1					

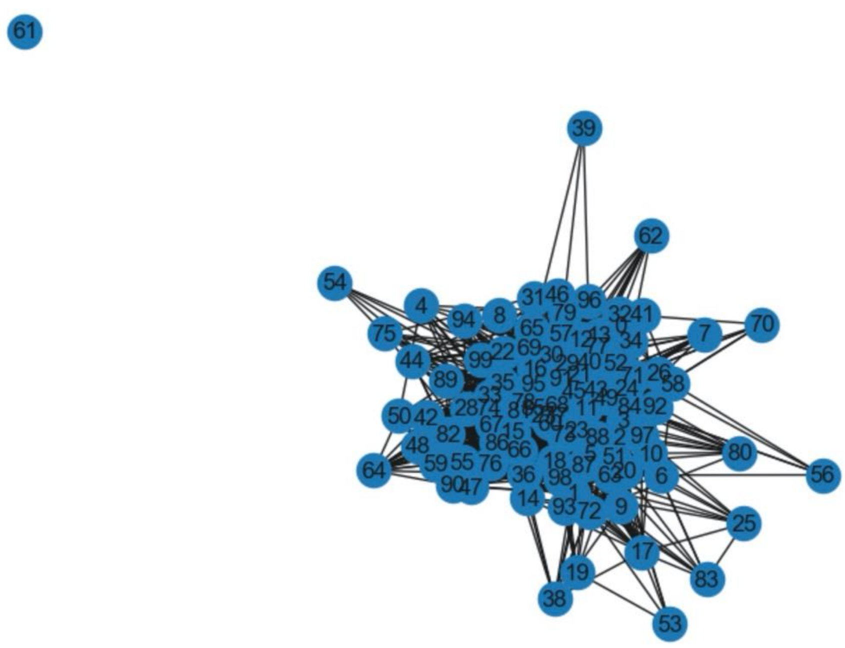
INTRCPT2, γ_{10}	0.557300	0.123963	4.496	888	<0.001
GROUP, γ_{11}	0.441604	0.130768	3.377	888	<0.001

We can see from above that all the coefficients are statistically significant with p-values less than 0.05. γ_{01} equals 3.365. This means that students in Group = 1, who are choosing to use both OLI and MOOC features, will have a higher pretest score from the start. This showcase our result is affected by the selection bias problem. γ_{11} equals 0.44. This indicates that students who use MOOC + OLI get better learning outcomes than students who only use MOOC. To improve the research accuracy, future research may include students' background characteristics such as education level and age into the level-1 model.

Semantic Network Centrality for the Discussion Forum

First, we try to find any post that is not connected to or influenced by other post. Here, we are only dealing with 100 texts because if we include the whole corpus, the network would be too complicated to see any clear pattern.

Graph 2: Network of Posts



We can clearly see that post 61 is an outlier. After extracting the post from the corpus, the content of the post is “I meni je drago sto mozemo suradivati (niko ovdje nije strucan iz psihologij e cini mi se - zato smo je valjda upisali) u svakom slucaju cestitam na prvom polozenom testu sretn o!!!!”. This is Croatian, not English. This is why it is not linked to other posts. After translating, thi s post means “And I’m glad we can work together…”.

Then, I go on to examine some of the most influential words in the database. Here, I use both closeness centrality and eigenvalue centrality to find the most and least 10 influential words.

Graph 3: Closeness Centrality (Left is Top 10, Right is Bottom 10)

```
[('>', 0.4807692307692308),
 ('people', 0.45045045045045046),
 ('<', 0.4329004329004329),
 ('course', 0.4065040650406504),
 ('disorder', 0.390625),
 ('group', 0.3875968992248062),
 ('find', 0.37174721189591076),
 ('affect', 0.37037037037037035),
 ('b', 0.36231884057971014),
 ('high', 0.3597122302158273)]
```

```
[('certificate', 0.2364066193853428),
 ('final', 0.23529411764705882),
 ('correct', 0.211864406779661),
 ('explanation', 0.211864406779661),
 ('right', 0.205761316872428),
 ('teacher', 0.19455252918287938),
 ('hello', 0.19455252918287938),
 ('have', 0.1926782273603083),
 ('write', 0.1926782273603083),
 ('come', 0.19157088122605365)]
```

Graph 4: Eigenvalue Centrality (Left is Top 10, Right is Bottom 10)

```
[('high', 0.3586133793780562),
 ('drive', 0.31672050535809476),
 ('people', 0.3167195978648418),
 ('iq', 0.3086112445794588),
 ('sex', 0.27058929071669097),
 ('>', 0.2466399778788388),
 ('find', 0.2390322470256942),
 ('group', 0.23709219627848374),
 ('prove', 0.23289673131378064),
 ('interest', 0.19498623739239865)]
```

```
[('assignment', 0.0014141971126021466),
 ('rubric', 0.0013852556826784762),
 ('certificate', 0.0012489233455172076),
 ('final', 0.0012347840492300918),
 ('right', 0.0007511790049585707),
 ('teacher', 0.00033004388904151805),
 ('hello', 0.00033004388904151805),
 ('have', 0.0001504716334483842),
 ('write', 0.0001504716334483842),
 ('come', 0.000132886205551115)]
```

From the above, we notice that, words such as “course”, “disorder”, “iq”, “sex”, group”, which are related to course materials and could be viewed as embedded in academic phrases, are most likely to send a signal with the most coverage to the rest of the network. While words describing logistic or casual issues such as “certificate”, “final”, “hello”, “assignment”, “rubic” have less influence in the corpus. Thus, I conclude that academic talk has more influence than logistic and casual talk in our database.

PART IV CONCLUSION & DISCUSSION

In this research, we have shown that the OLI and the discussion forum are effective online course features in terms of improving students' learning outcomes and encouraging students to participate in academic discussion. However, there are some caveats that need to be mentioned. First, when we are evaluating the effectiveness of the OLI, the ideal situation is to use the whole dataset for the hierarchical linear model. We are not able to include every student's data point in our research because we are using a private dataset and we only have access to a portion of the data. Second, since we are only using data points of students who attend the final exam, this makes our result subject to selection bias. It is hard to tell whether the positive learning outcomes of students are due to the OLI treatment or because they are the most diligent people among their peers. Third, when we conclude the effectiveness of the discussion forum, we are making an inductive claim rather than a deductive claim. The computational content analysis does not lead to causal statements. It only detects patterns in the data. Thus, the result is questionable under the RCTs framework.

Our research contributes to the pedagogical research of online education as we examine the cognitive framework behind the OLI and the discussion forum. Methodologically, it also explores the methods of effectiveness research. There are many exciting research opportunities for educationists to look at in online education. Better education requires better theories. Better theories require vastly more data than we have; online educational technology can provide that data. Careful research design and analysis of online course data could reveal more about effective instruction in the future.

References

Anderson, J. A., Reder, L. M., & Simon, H. A. (1997). Situative versus cognitive perspectives: Form versus substance. *Educational Researcher* 26(1): 18-21.

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science* 228, 456-468.

Ally, M. (2004) Foundations of educational theory for online learning, *The theory and practice of. online learning* (2nd ed), Athabasca University, pp. 3-31

Atkins, M. J. (1993). Theories of learning and multimedia applications: An overview. *Research Papers in Education*, 8(2), 251-271.

Bandura, A. (1971). *Social Learning Theory*. General Learning Corporation.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher* (13):3-16.

Borgatti, S. P., (2005). Centrality and Network Flow. *Social networks* 27(1): 55-71.

Carley, K. (1993). Coding choices for textual analysis: a comparison of content analysis and map analysis. *Sociol. Methodol*. 23:75–126.

Caspersen, J., et al. (2017). "Measuring Learning Outcome." *European Journal of Education Research, Development and Policy*, vol. 52, no. 1, pp. 20–30.

Chirikov, I., Semenova, T., Maloshonok, N., Bettinger, E., & Kizilcec, R. F. (2020). Online education platforms scale college STEM instruction with equivalent learning outcomes at lower cost. *Science Advances*, *6*(15), eaay5324.

Cook T. D. (2002). Randomized experiments in educational policy research: a critical examination of the reasons the educational evaluation community has offered for not doing them. *Educ. Eval. Policy Anal.* 24(3):175-99.

Corbett, A. T., Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural. knowledge. User modeling and user-adapted interaction, 4, 253-278.

Cumming, S., Ono, T. (1997). Discourse and grammar. In T. A. van Dijk Ed., *Discourse as. structure and process*, pp. 112-137. London: Sage

Dawson, S. (2010). 'Seeing' the learning community: An exploration of the development of a. resource for monitoring online student networking. *British Journal of Educational Technology*, 41(5), 736-752.

Derry, S. (1996). Cognitive schema theory in the constructivist debate. *Educational Psychologist* 3: 163-47.

Deubel, P. (2003). An investigation of behaviorist and cognitive approaches to instructional multimedia design. *Journal of Educational Multimedia and Hypermedia* 12(1): 63-90.

Dommett, E. J. (2019). Understanding student use of twitter and online forums in higher education. *Education and Information Technologies*, *24*(1), 325–343.

Duffy, T. M., Cunningham, D. J. (1996). Constructivism: Implications for the design and delivery of instruction. *Handbook of research for educational communications and technology*, ed. D. Jonassen, 170-98. New York: Simon and Schuster Macmillan.

Evans, J., Pedro A. (2016). Machine Translation: Mining Text for Social Theory. *Annual Review of Sociology* 42:21-50.

Ferguson, R. (2019). Teaching and learning at scale: futures. In R. Ferguson, A. Jones, & E. Scanlon (Eds.), *Educational visions: The lessons from 40 years of innovation* (pp. 33–50). Ubiquity Press.

Fosnot, C. T. (1996). Preface. In Constructivism: Theory, perspectives and practice. New York: Teachers College Press, Colombia University.

Gredler, M. E. (1997). *Learning and instruction: Theory into practice*, 3rd ed. Upper Saddle River, NJ: Prentice.

Garrison, D., Anderson, T., & Archer, W. (2001). Critical Thinking, Cognitive Presence, and Computer Conferencing in Distance Education. *American Journal of Distance Education* Vol 15.

Gillani, B. B. (2003). *Learning theories and the design of e-learning environments*. Lanham, MD: University Press of America.

Good, T. L., Brophy, J. E. (1990). *Educational psychology: A realistic approach*, 4th ed. White Plains, NY: Longman.

Hammond, M. (2005). A review of recent papers on online discussion in teaching and learning in. higher education. *Journal of Asynchronous Learning Networks*, 9(3), 9–23.

Jarvis, P., J. Holford, and C. Griffin. (2003). *The theory and practice of learning*, 2nd ed. Sterling, VA: Kogan Page

Koedinger, K. R., Corbett, A. (2006). Cognitive Tutors: Technology Bringing Learning Science to the Classroom.

Koedinger, K. R., (2015). Learning Is Not a Spectator Sport. Proceedings of the Second ACM Conference on Learning.

Koedinger, K. R., (2016). et al. Is the Doer Effect a Causal Relationship? Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK 16, 2016.

Laurillard, D. (1999). A conversational framework for individual learning applied to the "learning. organization" and the "learning society". *Systems Research and Behavioral Science: The Official Journal of the International Federation for Systems Research*, 16(2), 113–122.

Marra, R. M., et al. (2004) "Content Analysis of Online Discussion Forums: A Comparative Analysis of Protocols." Educational Technology Research and Development, vol. 52, no. 2, pp. 23–40.

Macdonald, J., Twining, P. (2002). Assessing activity–based learning for a networked course. *British Journal of Educational Technology*, 33(5), 603–618.

Macfadyen, L. P., Dawson, S. (2010). Mining LMS data to develop an "early warning system" for. educators: A proof of concept. *Computers & Education*, 54(2), 588–599.

Markoff, J. (2011). Virtual and artificial, but 58,000 want course. *The New York Times*. https://www.nytimes.com/2011/08/16/science/16stanford.html.

McCreary, E.K. (1990). Three behavioral models for computer-mediated communication. In L.M. Harasim (Ed.), *Online education: perspectives on a new environment*, pp. 117-130. New York: Praeger.

Merriam, S. B., Caffarella, R. S. (1999). *Learning in adulthood: A comprehensive guide*, 2nd ed. San Francisco: Jossey-Bass.

Meyer, O., & Lovett, M. C. (2002). Implementing a computerized tutor in a statistical reasoning. Course: Getting the big picture. B. Phillips (Ed.) Proc. of the Sixth International Conference on Teaching Statistics.

Mohamad M., Shaharuddin S. (2014). Online Forum Discussion to Promote Sense of Learning Community among the Group Members. *International Education Studies*, Vol 7, No. 13.

Newell, A., Simon, H. A. (1972). *Human Problem Solving.* Englewood Cliffs, NJ: Prentice-Hall.

Pavlov I.P. (1927). Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. Oxford University Press.

Peng, T. Q., Liang, H. & Zhu, J. H. (2019). Introducing computational social science for Asia-Pacific communication research, *Asian Journal of Communication*, 29:3, 205-216.

Raudenbush, S. W. (2020). Randomized Experiments in Education, with Implications for Multilevel Causal Inference. *Annual Review of Statistics and Its Application* Vol 7:177-208.

Raudenbush, S. W., Bryk A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd Edition. Sage Publication.
Schank, R. C, Colby, K. M. (1973). *Computer Models of Thought and Language.* San Francisco: Freeman.
Simon, H. A. (2001). Learning to research about learning. *Cognition and instruction*, ed. S.M. Carver and D. Klahr, 205-26. Mahwah: NJ: Lawrence Erlbaum.
Skinner, B. F. (1971). *Beyond freedom and dignity*. New York: Knopf.
Skinner, B. F. (1976). *About Behaviorism*. New York: Vintage Books.
Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1992). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Constructivism in education*, ed. L. P. Steffe and J. Gale. Hillsdale, NY: Lawrence Erlbaum Associates.
Strader, R. & Thille, C. (2012). The Open Learning Initiative: Enacting instruction online. In Oblinger, D.G. (Ed.) *Game Changers: Education Information Technologies (201-213). Educause*.
Suraj, P., & Kumari Roshni, V. S. (2015). Social network analysis in student online discussion forums. *2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, 134–138.
Thorndike, Edward Lee. (1911). *Individuality*. The Riverside Press, Cambridge.
Van Atteveldt, W. H. (2008). *Semantic Network Analysis: Techniques for Extracting, Representing, and Querying Media Content*. Charleston, SC: BookSurge.
Wang, X., et al. (2015). Investigating How Student's Cognitive Behavior in MOOC Discussion. Forum Affect Learning Gains. EDM.
Yilmaz, K. (2011). The cognitive perspective on learning: Its theoretical underpinnings and implications for classroom practices. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 84*(5), 204–212.