ATTITUDES, BEHAVIORS, AND LEARNING OUTCOMES FROM USING CLASSTRANSCRIBE, A UDL-FEATURED VIDEO-BASED ONLINE LEARNING PLATFORM WITH LEARNERSOURCED TEXT-SEARCHABLE CAPTIONS

BY

ZHILIN ZHANG

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Computer Science
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Advisors:

Teaching Professor Lawrence Angrave
Professor Karrie Karahalios

ABSTRACT

This thesis consisted of a series of three studies on students' attitudes, behaviors, and learning outcomes from using ClassTranscribe, a Universal Design for Learning (UDL) featured video-based online learning platform. ClassTranscribe provided accurate accessible transcriptions and captioning plus a custom text-searchable interface to rapidly find relevant video moments from the entire course. Users could edit the machine-generated captions in a crowdsourcing way. The system logged student viewing, searching, and editing behaviors as fine-grained web browser interaction events including full-screen-switching, loss-of-focus, caption searching and editing events, and continued-video-watching events with the latter at 15-second granularity.

In Study I, lecture material of a sophomore large-enrollment (N=271) system programming 15-week class in Spring 2019 was delivered solely online using a new video-based web platform - ClassTranscribe. Student learning behaviors and findings from four research questions were presented using individual-level performance and interaction data. Firstly, we reported on learning outcomes from alternative learning paths that arose from the course's application of Universal Design for Learning principles. Secondly, final exam performance was equal or better to prior semesters that utilized traditional in-person live lectures. Thirdly, learning outcomes of low and high performing students were analyzed independently by grouping students into four quartiles based on their non-final-exam course performance of programming assignments and quizzes. We introduced and justified an empirically-defined qualification threshold for sufficient video minutes viewed for each group. In all quartiles, students who watched an above-threshold of video minutes improved their in-group final exam performance (ranging from +6% to +14%) with the largest gain for the lowest-performing quartile. The improvement was similar in magnitude for all groups when expressed as a fraction of unrewarded final exam points. Finally, we found that using ClassTranscribe caption-based video search significantly predicted improvement in final exam scores. Overall, the study presented and evaluated how learner use of online video using ClassTranscribe predicted course performance and positive learning outcomes.

In Study II, we further explored learner's searching behavior, which was shown to be correlated with improved final exam scores in the first study. From Fall 2019 to Summer 2020, engineering students used ClassTranscribe in engineering courses to view course videos and search for video content. The tool collected detailed timestamped student behavioral data from $1,894$ students across 25 engineering courses that included what individual students

searched for and when. As the first study showed that using ClassTranscribe caption search significantly predicted improvement in final exam scores in a computer science course, in this study, we presented how students used the search functionality based on a more detailed analysis of the log data. The search functionality of ClassTranscribe used the timestamped caption data to find specific video moments both within the current video or across the entire course. The number of search activities per person ranged from zero to 186 events. An in-depth analysis of the students (N=167) who performed $1,022$ searches was conducted to gain insight into student search needs and behaviors. Based on the total number of searches performed, students were grouped into "Infrequent Searcher" (< 18 searches) and "Frequent Searcher" (18 to 110 searches) using clustering algorithms. The search queries used by each group were found to follow the Zipf's Law and were categorized into STEM-related terms, course logistics and others. Our study reported on students' search context, behaviors, strategies, and optimizations. Using Universal Design for Learning as a foundation, we discussed the implications for educators, designers, and developers who are interested in providing new learning pathways to support and enhance video-based learning environments.

In Study III, we investigated students' attitudes towards learnersourced captioning for lecture videos. We deployed ClassTranscribe in a large (N=387) text retrieval and mining course where 58 learners participated in editing captions of 89 lecture videos, and each lecture video was edited by two editors sequentially. In the following semester, 18 editors participated in follow-up interviews to discuss their experience of using and editing captions in the class. Our study showed how students use captions to learn, and shed light on students' attitudes, motivations, and strategies in collaborating with other learners to fix captions in a learnersourced way.

ACKNOWLEDGMENTS

I wish to thank everyone who has helped me. Words cannot express my gratitude and appreciation for your support.

- First and foremost, my advisors, Professor Lawrence Angrave and Professor Karrie Karahalios, thank you for always being so caring, thoughtful, and supportive. You brought me into the world of CS research, where I found the true passion for my career. You significantly influenced me with your excellence in teaching and research, and more importantly, how you care for others and promote equal opportunities for all.

- My mentor, Professor ChengXiang Zhai, who gave me a lot of great advice and inspired me on my research projects. I got the inspiration for Study III of this thesis when serving as a teaching assistant for your class *CS410: Text Information Systems*. Thank you for always encouraging me to "dream big and achieve big", and to pursue my true passion. I would not be where I am without you.

- More mentors and collaborators, Professor Ranjitha Kumar, Professor Yang Wang, Professor Hongye Liu, Professor Jenny Amos, Professor Craig Zilles, Professor Alex Kirlik, Professor Abdussalam Alawini, Professor Eric Shaffer, Professor Wade Fagen-Ulmschneider, Professor John Popovics, Professor Yun Huang, and Professor Sauvik Das, who provided me with valuable advice and guidance on research.

- Even more mentors and friends, Bhavya, Tiffany Wenting Li, and Silas Hsu. You taught me how to do research by showing me. I have learned a lot by working with you.

- Friends and co-authors whom I was fortunate to work with: Zhuohao Zhang, Si Chen, Ruihua Sui, Chirantan Mahipal, Jinda Han, Deniz Arsan, Yuan Shen, Shanduojiao Jiang, Natã Barbosa, Haolin Yuan, Ian Ludden, Patrick Lin, Yiyin Shen, Kusum Vanwani, and the members of the Social Spaces Group, ClassTranscribe Team, Text Information Management and Analysis Group, Social Computing Systems Lab, and Data Driven Design Group.

- A special thank you to Professor Jiawei Han for mentoring me during my undergraduate study. You gave me key suggestions on how to get started with CS research.

In addition, I wish to acknowledge the University of Illinois at Urbana-Champaign (UIUC) IT staff, the National Center for Supercomputing Applications (NCSA), the College of Engineering, current and former undergraduate and graduate students, Director of Digital Media John Tubbs, and Professor Mark Hasegawa-Johnson, who have contributed to the development, support and direction of the ClassTranscribe project. I wish to acknowledge members of the Illinois Computer Science Education group, and the Illinois iLearn Group who helped in the preparation of the research papers. I appreciate the invaluable technical support from University of Illinois students, staff, and faculty, including Rob Kooper, Dave Mussulman, Scott Cimarusti, Tim Yang, and Jon Gunderson. I wish to thank the UIUC Institutional Review Board (IRB) and the Center for Innovation in Teaching & Learning (CITL) Data Analytics Group for their input and help.

The research reported here was supported by a Microsoft Corporation gift to the University of Illinois as part of the 2019 and 2020 Lighthouse Accessibility Microsoft-Illinois partnership, an award from Center for Innovative Teaching and Learning, and the Institute of Education Sciences, U.S. Department of Education through Grant R305A180211 to the Board of Trustees of the University of Illinois. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Finally, thank you to my family and friends. Your love and support always give me the strength and courage to continue the journey.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

As the popularity of video-based online learning, for example, Massive Open Online Courses (MOOCs), continues to grow, it is crucial to provide accessible videos to all, including but not limited to, students who are non-native speakers and students with physical or mental disabilities. The COVID-19 pandemic has necessitated a rapid shift to online education, increasing the urgency of providing accessible online educational videos [1]. Prior research has shown that accurate captions play an important role in making lecture videos accessible [2], especially for students who are Deaf or Hard of Hearing (DHH). Further, following the principles of Universal Design for Learning (UDL), captions are found to be beneficial to a wide range of students [3] including students with other disabilities, such as Attention Deficit Hyperactivity Disorder (ADHD), dyslexia, and students who prefer to learn by reading.

However, creating accurate captions is challenging. Commercial captioning services are slow and expensive [4], while Automatic Speech Recognition (ASR) algorithms are faster and cheaper but tend to be less accurate and hard to satisfy students' needs in learning with educational videos [5]. Generating domain-specific captions, for example, captions for Science, Technology, Engineering, and Math (STEM) courses, presents additional challenges due to the use of scientific terms and jargon, and various kinds of accents of instructors. [6, 7]

This thesis included a series of three studies, aiming to investigate (i) how the introduction of ClassTranscribe, a UDL-featured video-based online learning platform with text-searchable captions affected students' learning outcomes; (ii) what behaviors did students present in caption-based video search; and (iii) what attitudes did students hold towards learnersourced captioning for online lecture videos.

The primary goal of the first study was to report on and evaluate the replacement of physical lectures (that used a blend of active learning and instructor-led live-coding examples) with online content designed for online viewing. These videos were delivered using a new custom web-based, text-searchable video player system, ClassTranscribe [8, 9, 10], that met accessibility standards, applied Universal Design for Learning (UDL) principles, and allowed students to search for relevant content by indexing transcribed video. The ClassTranscribe application generated low-cost accurate captions by combining automated and crowd-sourced techniques. The latter included editing by both course staff and students within the course. Learning outcomes of the Spring 2019 students were compared to those of previous comparable semesters. By combining behavioral interaction data captured by

ClassTranscribe with student "gradebook" assessment data, the learning benefits of the tool and online videos were examined by class quartile within the course and also compared to three previous comparable semesters.

The second study was a continuation of the first study. We presented the findings from behavioral analyses of students' caption search activities that were recorded in the user activity logs of ClassTranscribe. Study I and previous research showed that caption search was correlated with improved student exam scores in a sophomore-level computer science class [9] and was reported as useful for learning by students [8, 11]. In this study, we investigated how students use caption search in video-based learning and discussed implications for improving video-based learning environments based on Universal Design for Learning (UDL) principles.

The third study aimed to investigate students' attitudes towards learnersourced captioning. This study was motivated by students' feedback in a senior-level large enrollment (N=387) computer science class (*CS410: Text Information Systems*) on Text Retrieval and Mining. In Fall 2020, CS410 was offered completely online. Students watched lecture videos, worked on assignments, and took tests on Coursera [1], a MOOC platform. However, in the middle of the semester, students reported that the caption quality of the lecture videos was bad and hindered learning. Several other students agreed in the follow-up discussion and offered that they are willing to help with improving the captions of the lecture videos. Since Coursera did not support caption editing on student side, we deployed ClassTranscribe for students to fix captions in a crowdsourcing way. We then offered an extra-credit activity for students to fix the captions, where 58 students edited 89 STEM lecture videos, and then conducted semi-structured interviews with 18 participants to investigate their attitudes.

This thesis is organized as follows: In the background section, we review the use of lecture videos and Universal Design for Learning (UDL) principles from both the CS education and educational literature, and we briefly discuss students' use of videos and searching for content as a means of regulating their metacognitive awareness. Then we give a brief overview of ClassTranscribe with comparison to other similar tools and studies, and we discuss related works on caption search and learnersourced captioning. The methods section covers the data collection and analysis process. The result section shows findings on learning outcomes, search behaviors, and attitudes towards learnersourced captions. Finally, we conclude the paper with the discussion section where we summarize the findings and discuss implications for future research.

[1]https://www.coursera.org/

CHAPTER 2: BACKGROUND

2.1 VIDEOS AND ONLINE LEARNING

Online learning environments constitute an important application area of human-computer interaction and are ubiquitous in today's educational landscape. A key instructional element in these learning environments is the set of video lectures that often serve as the unique portal connecting the instructor to the learner, apart from serving as primary portals of content delivery [12, 13, 14]. Not surprisingly, studies have found that it is the course component with which students spend most of their time [15, 16]. Other studies, especially those on massively open online courses (MOOCs), have found that a significant number of students primarily watch only videos, while skipping over assessment problems and other interactive course components [17, 18]. Students report that the videos make them feel more engaged in the course and motivate them [19], and that they also feel more connected to the instructor [20, 21]. Additionally, students perceive that the videos help them learn [22]. However evaluating if the videos have actual effects on learning is a necessary criteria for most practitioners, and Study I investigates video watching beyond affective measures and perceptions to better understand learning outcomes, specifically in videos that are closed-captioned and transcribed.

Modern creation and delivery of educational online lecture videos includes a multitude of presentation formats, sources, and delivery tools. Sources include lecture-room multi-view recording systems to simultaneously capture live lectures from room camera and display sources, screen recording applications (e.g. Camtasia, OBS Studio and native support in Microsoft Windows 10), and content professionally edited for MOOCS and large audiences. Delivery mechanisms span emailing a hyperlink to the class ("Here's the mp4 file on my shared directory") to sophisticated commercial systems that optimize video playback for large audiences (e.g. YouTube, Echo360, Kaltura, Coursera).

Recent work on crowdsourcing transcriptions include [10, 23, 24]. The system presented here overcomes limitations from existing open-source and commercial alternatives. Namely, it is a system that enables all of the following features:

- Has inexpensive and accurate captions. Commercial caption systems offered free-but-low-accuracy transcriptions, or professional transcriptions services at $1 per video-minute.

- Has fine-grained server-side logging of user interaction events to support educational research, student engagement and course rewards based on students exhibiting desired behaviors. The former includes both insights into how and when students are engaging with course videos resources and, by joining with course performance data, provides insights into how different learning behaviors benefit each student or student sub-populations.

- Deploys easily on university hardware or the commercial cloud.
- Is open source and extendable. The source-code is available at `https://github.com/classtranscribe/`.
- Supports modern accessibility web standards (e.g. closed-captions with configurable display settings, Aria-tag support for blind users)
- Supports multi-stream and viewer-controlled playback speed.
- Enables the finding of relevant content by indexing transcriptions within the current video and across the entire course that is relevant to the student's current activity or the student's pre-exam review knowledge-seeking question.

In the tool presented here, audio is initially transcribed using a modern high-accuracy speech-to-text cloud service (Azure Cognitive Services Speech-To-Text) at a cost of approximately $1 per hour. This service also supports improved transcription accuracy of domain-specific speech (though this feature was not enabled or evaluated in this study, and is an opportunity for future tool improvements and research).

2.2 CAPTIONS FOR VIDEO-BASED LEARNING

Previous research has shown that to make educational videos accessible to the widest audience possible, it is important to improve the readability of the text and captions of the videos [2]. Following the principles of Universal Design Learning (UDL), captions are found to be beneficial to a wide range of students [3] including students who are deaf or hard of hearing and students who are non-native speakers. However, many educational videos still have no captions or machine-generated captions with high error rates [25].

Compared to general videos, it requires more domain knowledge to transcribe educational videos accurately. For instance, videos for STEM courses are hard to transcribe because of challenges such as complicated technical terminologies and slides with significant graphical content [6] while captions can be useful for understanding these technical terms in STEM classes. Therefore, it is helpful to explore how domain-related terms and equations can be transcribed for lecture videos. By gaining such understandings, new mechanism and design could be proposed to improve caption quality for educational videos online.

2.3 UNIVERSAL DESIGN FOR LEARNING PRINCIPLES

Videos used in the study were closed-captioned and had accompanying transcriptions, which adhere to Universal Design for Learning (UDL) principles. Following these principles ensures that students' diverse abilities and needs are met. A UDL framework [26] suggests that instructional materials encourage (1) equitable use, (2) flexibility in use, (3) simple and intuitive use, (4) perceptible information, (5) tolerance for error, (6) low physical effort, and (7) adequate size and space for use. By providing materials that incorporate these principles, students of all backgrounds have fewer barriers to learning [27]. Although research has found that students enjoy courses that adhere to UDL principles (e.g., [28]) and that they perceive their instructor's teaching abilities to be better [29], there do not seem to be significant differences in learning outcomes for students with and without disabilities. For example, researchers [30] found no significant difference in both disabled and non-disabled students' learning between a course taught using UDL principles compared to a course taught in a regular manner. Another research [31] showed improvements in learning outcomes for some types of UDL materials but not all.

One reason that we hypothesize that videos adhering to UDL principles – and videos that have a text-searchable feature in particular – will be beneficial for all students is because of their ability to assist in students' metacognitive awareness. Metacognition is awareness of one's own thinking; more specifically, it enables a person to monitor, assess, and regulate one's understanding and thought process [32]. Thus students who are metacognitively aware are able to recognize what they do not know and then figure out how to extend their knowledge [33]. As such, engaging in metacognitive thinking and strategies improves learning outcomes [34, 35, 36], and this effect on learning outcomes also holds true in online environments [37]. Engaging in metacognitive strategies is effective for improving low-achievers' learning outcomes [38], and it also explains the difference between high and low achievers [39, 40].

Instructors must be deliberate in their support of metacognitive awareness [41]. In this particular study, students had access to a unique text-searchable interface to rapidly find relevant video moments from the entire course. Thus, they were able to reflect on what they did not understand and could easily locate material to assist them in clarifying that understanding. Researchers at the University of Houston [24] developed a similar tool, although no research to date has looked at the relationship between the tool's use and student performance–a gap in which this study attempts to fill. In the next section we outline related tools and discuss the utility of the tool presented here.

In addition, Study II includes the UDL materials, in the form of videos accompanied with text, and secondary learning opportunities of the online course book, to understand if this material promotes better learning outcomes for all, and if alternative learners paths that utilize these materials result in positive learner outcomes. We researched how students interacted with ClassTranscribe, a UDL-featured online learning tool and presented caption-based video searching and indexing as a new learning pathway. We studied students' behaviors in interacting with this learning pathway and proposed ways to support and enhance students' learning experiences. Previous research [42, 43, 44] has shown that UDL can provide opportunities for enhancing learning experience for students with disabilities and promote an inclusive learning environment. Our study adds to that body of research by discussing how a searchable video system can not only help learners with physical or cognitive disabilities, and learners who are non-native English speakers, but ultimately provide an inclusive enhanced video-based learning experience for all learners.

2.4 CLASSTRANSCRIBE - A UDL-FEATURED ONLINE LEARNING PLATFORM

This research presents findings from data analysis of students' usage, behaviors, and learning outcomes in using ClassTranscribe [8, 9, 10], a UDL-inspired tool that can automatically generate text-searchable captions to lecture videos uploaded by instructors in a web interface that includes accessibility support (e.g., support for users who use a screen reader), or require a low-distraction interface. ClassTranscribe uses Automatic Speech Recognition (ASR) to generate captions for lecture videos and indexes them to facilitate keyword-based search. Students can search by keywords across all lecture videos of the same course and retrieve the relevant videos and moments in the videos. They can directly jump to a specific moment by clicking on the search result. Figures 2.1 and 2.2 provide sample screenshots of the video and searching interface of ClassTranscribe.

Figure 2.3 summarizes a mapping of ClassTranscribe features to the UDL framework. The authors' opinion is that the caption-based video search functionality should be mapped into the "Provide options for Perception" item in the "Provide multiple means of Representation" category because it offers a new way to "customize the display of information" by enabling students to search and gather information they need.

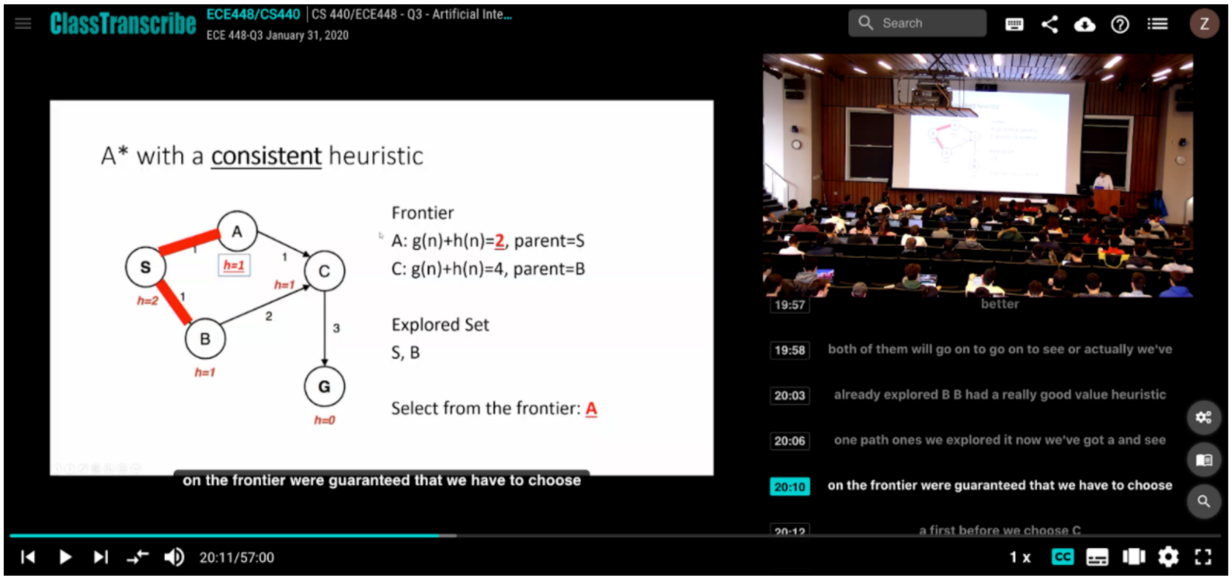
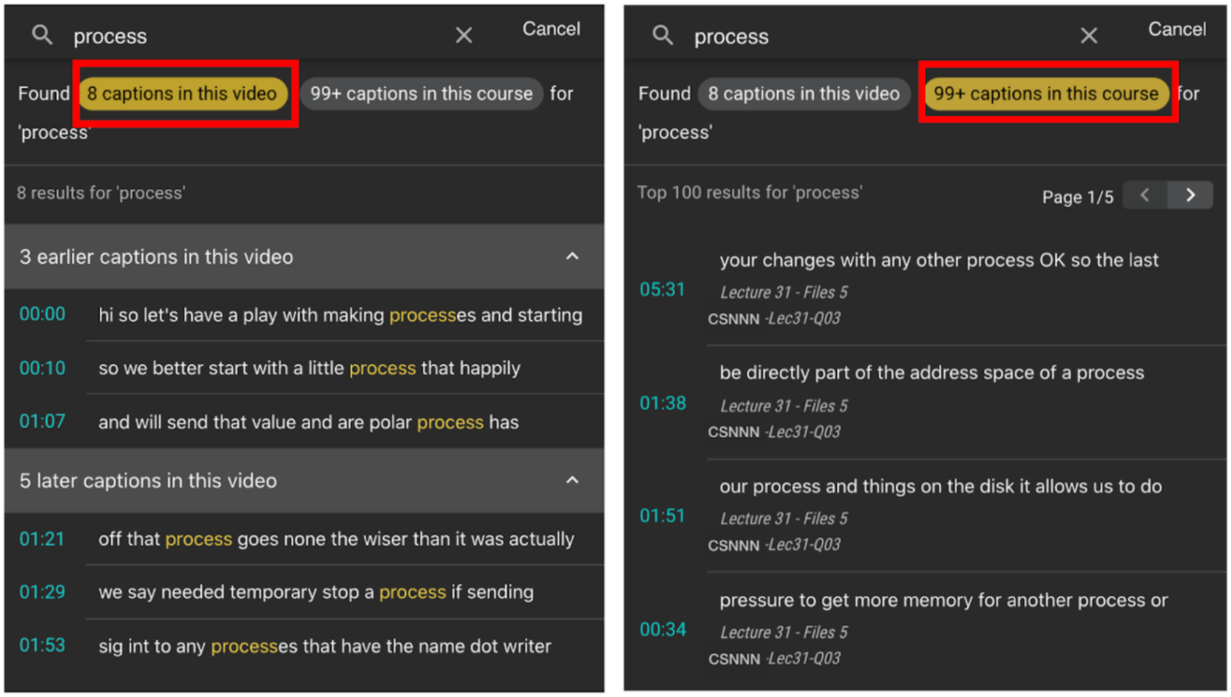


Figure 2.1: Video Interface of ClassTranscribe. Closed captions can be turned on and configured. Captions on the side are provided with spacing to facilitate reading by users with dyslexia.

2.5 CAPTION-BASED VIDEO SEARCHING AND INDEXING

A previous study found an enhanced video-based learning environment (i.e., with embedded note-taking, supplemental resources, and practice questions) significantly improved recall test scores over a common video environment limited to play, pause, rewind, and forward operations [45]. Researchers have experimented with video-lectures in a traditional mathematics course (Calculus). The students' feedback and usage analysis showed that video resources are considered useful and are correlated with improved academic results [46]. Based on the positive learning outcomes from video-based learning environments, researchers have explored new features in video-based learning environments and are creating new learning pathways for students. One of these new features is caption-based video searching and indexing.

In Study I, we show that the introduction of searchable video lectures in a sophomore-level system programming course, which complemented the equivalent online book content, saw an increase in course performance for all students [9]. Similarly, in a freshman-level introduction to electronics course, translated searchable class videos led to improved course scores among students that used them [8]. With the demonstrated improved performance in CS and ECE courses, we were curious to better understand how students used the search functionality as part of their authentic university course experience (as opposed to a simulated learning environment). Study II is based on student usage data from 25 engineering courses to



(a) (b)

Figure 2.2: The Searching Interface of ClassTranscribe; captions appear on the right side of a lecture video. (a) matching captions found in the same lecture video as the one being currently viewed. (b) matching captions found in other lecture videos.

explore how students used video-based index search, and provide insights so educators can learn about student behaviors and learning outcomes related to their use of the search and indexing features.

A similar comparable system is the Indexed Captioned Searchable Videos ("ICS") system used at the University of Houston [24] and [11]. Recent evaluation of ICS in an introductory psychology course [47] reported increased instructor and student satisfaction but "confounding variables in the structuring of the course prevent us from making firm conclusions about student performance." However, They conducted a student survey that found students considered indexing and search features to be helpful [11]. Another study [48] explored the design space of video navigation. The study created a prototype video tool, lectureScape, and used it in a simulated learning environment. The tool provided several UI navigation mechanisms including visual and keyword and the study participants performed pre-specified navigation tasks to search and watch a 15-minute video.

To the best of our knowledge, this is the first study to analyze students' usage and behaviors in caption-based video searching and indexing in multiple engineering courses and in an authentic engineering setting.

2.6 LEARNERSOURCED CAPTIONING

Following the principles of Universal Design Learning (UDL), captions are found to be beneficial to a wide range of students [3]. Research has shown it is important to improve the readability of the text and captions of the videos [2]. Yet many videos remain uncaptioned or have machine-generated captions with high error rates [25].

There exist educational systems, such as ClassTranscribe [23] and ICS Videos [49], that can generate captions for lecture videos in a learnersourcing fashion. Previous studies [8, 49, 50] have shown that learnersourcing tools are effective for captioning lecture videos and have considerable value in educational practice. Previous research has shown that involving students in fixing captions in foreign language educational videos does not impair learning and also helps reduce errors in the captions [51]. However, despite the fact that these learnersourcing tools could be beneficial for educational purposes, none of these works studied how learners worked together in the learnersourcing experience. To the best of our knowledge, this is the first study to touch on students' behaviors and attitudes towards collaborating with each other to help with improving the captions of lecture videos in a STEM online class.

UDL Guideline	**UDL Guideline item(s)**	**ClassTranscribe Feature**
Provide multiple means of Engagement	Minimize threats and distractions	- Distraction/stress-free learning interface
	Provide options for Sustaining Effort & Persistence	- Student personal usage analytic reports based on interaction with the platform
Provide multiple means of Representation	Provide options for Perception Offer ways of customizing the display of information Offer alternatives for auditory information Offer alternatives for visual information	- Captions & transcriptions available - Caption-based video search to filter caption results - Multiple cameras angles - Configurable playback interface, caption font size and color, background color, background transparency - Can generate textbook from video in multiple formats (epub, pdf, html, etc) from videos
	Provide options for Language & Symbols	- Multiple languages for captions/transcriptions
Provide multiple means of Action & Expression	Provide options for Physical Action	Accessible design - Multiple keyboard shortcuts - Support screen readers for users who are blind or have low-vision - Continue watching a partially completed video
	Provide options for Expression & Communication	- Create shareable link for specific video moments

Figure 2.3: Features of ClassTranscribe classified using UDL. (Guideline items reproduced from `udlguidelines.cast.org`)

CHAPTER 3: METHODS

3.1 STUDY I: ANALYTICS OF ONLINE LECTURE VIDEO VIEWING

In Spring 2019, Monday-Wednesday-Friday lectures for a sophomore system programming class for CS majors were scheduled at 8am. The instructor suggested this early time would severely impact attendance and learning. Rather than penalizing absence, the instructor offered to record equivalent lecture content specifically designed for online viewing. The scheduled time slots were used three times: once for the first lecture, once for a guest lecture, and once for the closing lecture. The weekly Wednesday lecture time was re-purposed for instructor office hours, which were utilized by two students over the semester, and the Monday and Friday meeting times were not used. This change was announced only on the first day of class. The instructor offered to provide in-person 8 am lectures for students who requested an in-class experience, but no student requested it. Previous semesters also offered online video lectures, but the content consisted of simple automated recordings of the classroom lectures. For the Spring 2019 semester, only the new recorded video lectures were used by the course and delivered by ClassTranscribe.

3.1.1 Data Collected

Table 3.1: A summary of the 1198413 events by user action.

Action	Count	Fraction
timeupdate	413912	0.345383
play	164428	0.137205
seeking	160924	0.134281
pause	153034	0.127697
userinactive	134455	0.112194
seeked	102531	0.085556
changevideo	24750	0.020652
changedspeed	23151	0.019318
fullscreenchange	17195	0.014348
selectcourse	2850	0.002378
filtertrans	1087	0.000907
edittrans	83	0.000069
sharelink	13	0.000011

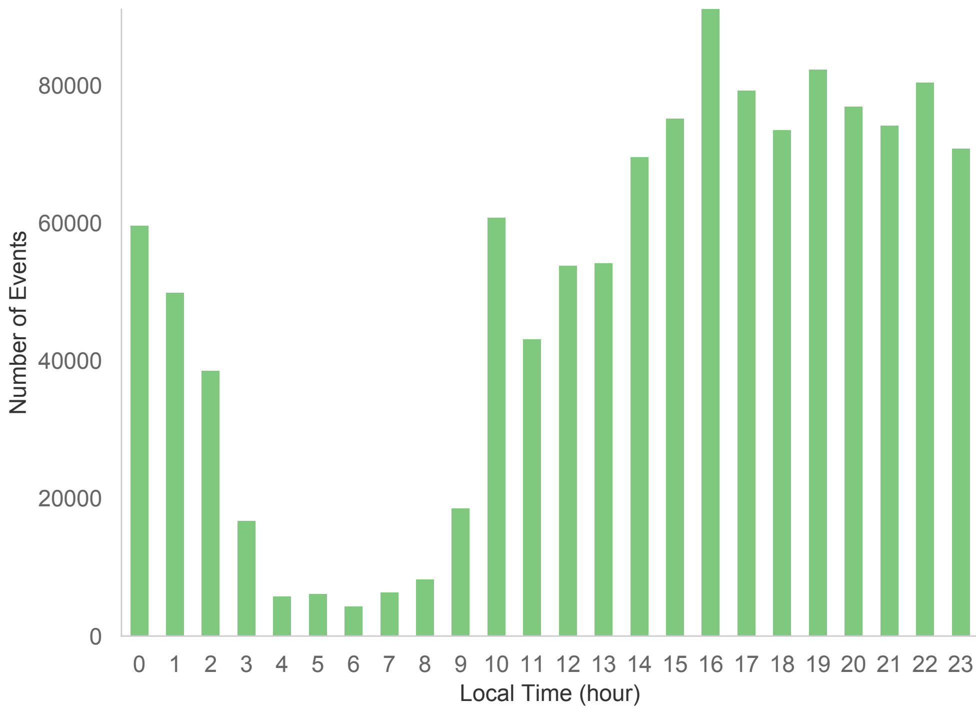


Figure 3.1: Events captured by hour

Event logging of student interactions was available in approximately the last 7 weeks prior to the final exam, of which 6 weeks contained lecture video content. Table 3.1 presents a summary of the events collected for the system programming course by event type. A *timeupdate* event was generated whenever a user watched 15 continuous seconds of video, which could take fewer wall-time seconds if the video was played at a higher speed. A total of $1,725$ video hours was served over the last 7 weeks of the course. The small number of *selectcourse* events suggest students remained logged in between viewings. The small number of times that students shared a URL link to a specific moment in the video suggests students may have been unaware that they could easily share and discuss a particular video moment in the course's discussion forum external to this tool. This suggests a future version of the tool should present user tips to enable more pedagogically valuable utilization of the tool. Students used the tool throughout the day, particularly in the afternoon and evening (See Figure 3.1). Only 0.6 % students chose to learn at the scheduled 8am time slot.

3.1.2 Evaluation

We use behavioral and gradebook data to understand how students learn and thrive in the course. We present four student-outcome questions -

- **RQ1.1:** "Are students exploiting multiple learning paths?" Are students utilizing the same learning resources in similar proportions, or are they exploiting multiple learning paths available and thriving?
- **RQ1.2:** "First, do no harm" - Did the cancellation of traditional in-person lecture and sole use of the ClassTranscribe for lecture content, affect exam performance when compared to similar previous semesters?
- **RQ1.3:** "Who benefits?" Do online lecture videos help all students, or only the strongest students? Are more video minutes viewed associated with improved exam scores?
- **RQ1.4:** "What benefits?" What features can benefit student learning? And how much is the benefit?

3.2 STUDY II: ANALYSIS OF CAPTION-BASED VIDEO SEARCH

From Fall 2019 to Summer 2020, engineering students used ClassTranscribe in multiple engineering courses to view course videos and search for video content. The tool collected detailed timestamped student behavioral data from $1,894$ students across 25 engineering courses that included what individual students searched for and when. Study II is an analysis of caption search based on students' interaction data logged by ClassTranscribe.

3.2.1 Event Types

Every user interaction (event) on ClassTranscribe was logged by the server for later analysis. The event types logged by the system are listed and described in Table 3.2. The *filtertrans* event is logged when a user performed keyword-based search on captions. From the search results, when a user clicked on a found caption in the same lecture video, the *seeking* event was triggered and the tool automatically adjusted (scrubbed) the playback position to jump to the corresponding video moment (Figure 2.2). Once the scrubbing completed, the *seeked* event was generated. The *seeking* and *seeked* events could also be generated when the user manually adjusted the video playback position. When the user navigated to another lecture video (e.g., by clicking on a caption line in the search results), the *changevideo* event was generated. Thus, *seeking*, *seeked*, and *changevideo* events after a search indicated that the learner clicked on a caption matching the search query.

Table 3.2: The Event types logged by ClassTranscribe

Event Type	Event Description
timeupdate	Watched another 15 seconds of video
play	Played a video
seeking	Adjusted to a video playback position
seeked	Finished loading to a video playback position
pause	Paused the video
userinactive	Switched focus away to another webpage
changevideo	Changed to a different video
filtertrans	Updated search filter of the transcription interface
edittrans	Edited the transcript; updated a caption line of text

3.2.2 Evaluation

Using the event data logged by the video system, our goals were to better understand students' interactions with this video-based learning platform and propose methods to support and enhance it for a more effective learning experience. We asked the following research questions about searching captions in educational videos.

- **RQ2.1:** What do searchers search for using caption-based video search?
- **RQ2.2:** What collective behaviors do searchers exhibit before, during and after search?
- **RQ2.3:** Is caption-search used to review previously viewed content or to find new content?

The interaction of each student with ClassTranscribe was recorded by the server as a time series of event data stored in a database. This included data that represented student actions e.g., starting a search or playing and pausing a video. Each event included a student identifier, a timestamp and details about the specific event. To address the above research questions we aggregated by student identifier the event data and analyzed the event sequence and event types (e.g., search, seeking, video play) performed by each student, as described in the following section.

3.2.3 Mining System Logs

To answer our research questions, we mined the system event logs. Log mining is commonly used to analyze search behavior [52, 53]. We defined search behavior in terms of frequency of searches, types of search queries, frequency and types of other events (Table

3.2) around search performed by students. Previous work [54] used a similar definition of search behavior for educational video search, but our analysis was more detailed (e.g., we also studied the types of search queries).

Figure 3.2 shows an overview of the study design. With approval from the Institutional Review Board (IRB), we collected anonymized student interaction data logged by ClassTranscribe. The data included interaction events from 1,894 students across 25 engineering courses during 09/2019 to 07/2020. For each user interaction with the tool, the event logs contain information about the user (anonymized UserID), the Event Type (Table 3.2), the timestamp of the interaction, the lecture video being viewed at the time of interaction (Video ID), and additional details relevant to the interaction (e.g., search query in case of *filtertrans* event).

Event logs were extracted from the system's Postgres database using a SQL query and subsequent data processing and analysis was performed in Python using the Pandas, numpy and scipy modules. Search data was grouped by view using an anonymous identifier. Widely used statistical techniques (e.g., Chi-square tests [55]), data mining techniques (e.g., k-means clustering [56]) and qualitative analysis techniques (e.g., manual coding using a grounded approach [57]) were used for further analysis.

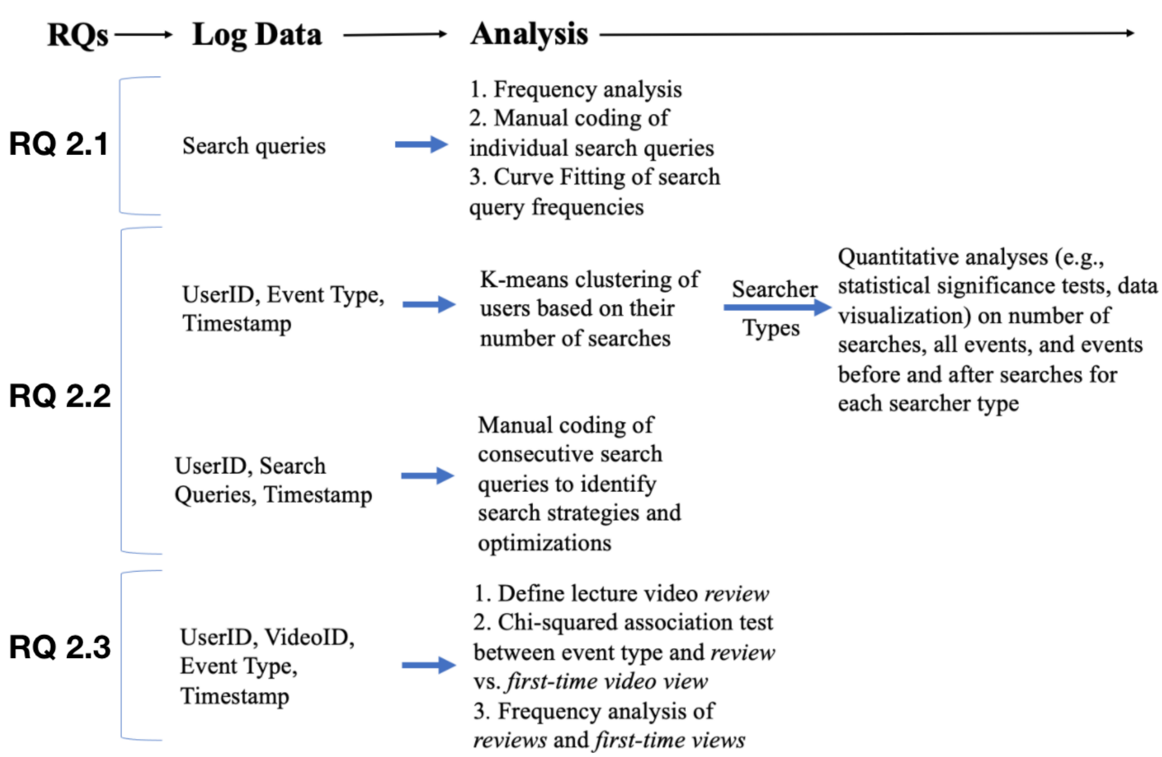


Figure 3.2: Overview of the analytical architecture for Study II

3.3 STUDY III: INTERVIEWS ON ATTITUDES TOWARDS LEARNERSOURCED CAPTIONING

In this section, the study design and methodology of Study III will be presented as follow. First, We will introduce the caption editing interface of ClassTranscribe. Then, information on the activity where 58 students were involved in fixing the captions of 89 out of 93 lecture videos with two rounds of editing in a senior-level text mining course will be given. We will then discuss the settings of the follow-up interview after the caption-fixing activity. Lastly, the demographic information of the 18 participants in the interview will be presented.

3.3.1 The Caption Editing Interface of ClassTranscribe

Figure 3.3 shows the user interface for editing captions on ClassTranscribe. The captions to the video are displayed on the right side of the video. Each caption has a corresponding video timestamp on the left. By clicking on the timestamp, users can jump to the corresponding moment of the video. Clicking on a line of caption will open up the "edit mode" of the caption clicked. In the "edit mode", users can modify the caption as they want. After editing the caption, users can either click on the "Save" button or hit "return/enter" on the keyboard to save their edits. Any edits made are immediately reflected on the interface and stored into the database. Therefore other users of the system can also see the change simultaneously. Users can search captions based on keywords within the current video or across all lecture videos in the same course. If the user feels that having captions is distraction, they can also turn it off. All user activities on the system, including searching, watching a video, seeking, editing a caption are logged in the SQL database and can be retrieved for later analysis.

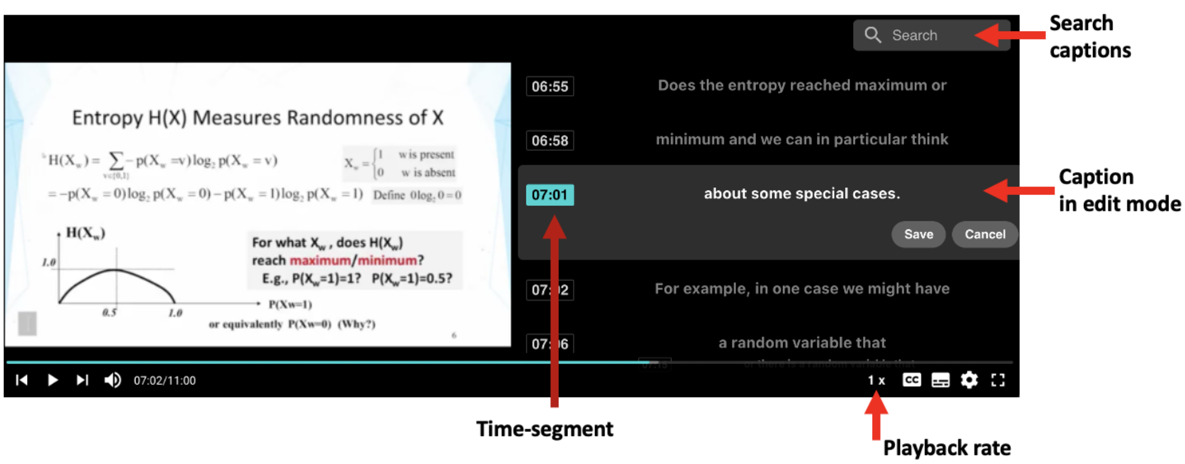


Figure 3.3: Interface for Editing Captions

3.3.2 Caption Editing Activity Design

During the Fall 2020 semester, a senior-level computer science class on Text Mining and Analytic at a large public university in the US was given on Coursera. There were a total of 93 lecture videos in the course. Each video was on average 12 minutes long. Each lecture video transcript has about $1,500$ words. There were about 400 students in the class. The course was given entirely online: students watched the lecture videos, did the assignments, quizzes, and took the exams on the Coursera platform. However, in the middle of the semester, some students complained in the discussion forum (Piazza) that the captions to the lecture videos were inaccurate to the point that it could hinder their learning experience. In the follow-up discussion, students offered that they would be willing to help with fixing the captions and improving the course content. Therefore, we introduced ClassTranscribe as the platform for students to edit the captions. We announced an activity where each student could sign up to edit videos and receive extra credits. 58 students participated in the activity, in which 50 students officially signed up for the activity (can get the extra credit) and 8 students volunteered to help.

The caption editing activity consisted of two rounds. The task of *Editor One* was to take a first pass at fixing errors in captions. After *Editor One* finished their task, *Editor Two* then reviewed the captions to address any remaining errors. In this way, *Editor One* and *Editor Two* sequentially edited the captions to fix the errors. Students in this activity knew who signed up for which role for each video, but they could only see the edited text by other editors without knowing what the exact change was.

The system described above was released to students towards the end of the course only for this activity. Before that, students in this course used Coursera to watch lecture videos and complete assignments. For the extra-credit activity on fixing captions, students were given an opportunity to sign up as *Editor One* and *Editor Two* for one lecture video each to get 1% extra-credit. To support participation for as many students as possible, each student could only register for a maximum of two lecture videos each as *Editor One* and *Editor Two* for a total of 2% extra-credit. Students had two weeks to complete the *Editor One* task. After that, they had another two weeks for completing the *Editor Two* task. Although we limit the maximum number of videos that a student could sign up for, there were still not enough videos for every student in the class to register for the extra-credit activity. Therefore, to provide equal opportunity for all students, other extra-credit activities were also released simultaneously. In addition, we did not set any minimum number of edits for getting the extra credit. Students were completely free on deciding what they wanted to edit and how many edits they wanted to do.

The system captured every caption edit made by editors with the information on who made the edit, the time of edit, the caption *before* and *after* the edit, the corresponding lecture video time-segment, and the lecture name. Figure 3.4 shows an example of captions *before* (left) and *after* (right) edits. The edited parts are highlighted in the figure. We later showed edit logs in this format to editors who participated in the follow-up interview.

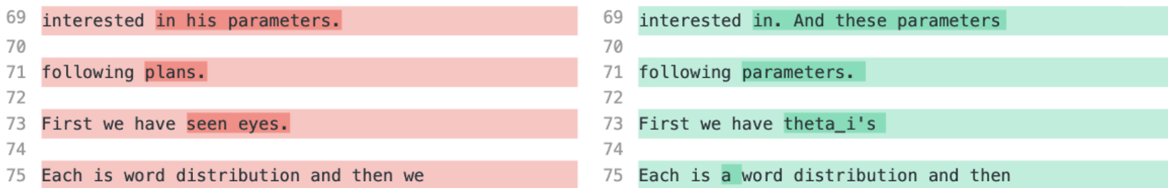


Figure 3.4: Sample captions *before* and *after* an edit

3.3.3 Follow-up Interview

In Spring 2021, we recruited students who had previously participated in the caption editing activity to participate in an online semi-structured interview with approval from the University IRB. All interviews were conducted entirely online using Zoom and were recorded with consent from participants. 18 participants were recruited via email. Each participant was paid a compensation at $20/hour (pro-rated). and each interview was approximately 45 minutes on average. Table 3.3 summarizes the demographic information of the interviewees. None of the interview participants reported having any physical or mental chronic conditions that would prevent them from understanding the speech in lecture videos. We collected interview data from 18 participants and interaction log data containing 10,378 rows of word-level edits generated by 58 learner editors from the ClassTranscribe system. The author used a grounded approach [58, 59] and conducted a thematic analysis [60] based on the interview notes and transcribed text of interview recording.

This study involved students in fixing captions. We conducted this research to investigate students' attitudes and behaviors towards fixing captions in a learnersourced way. Specifically, we addressed the following research questions:

- **RQ3.1:** How do students use captions for learning?
- **RQ3.2:** How do students edit the captions?
- **RQ3.3:** What explicit suggestions do students have for learnersourced captioning?

Table 3.3: Interviewee demographics. There were eight Females (F), nine Males (M) and one who preferred not to disclose their gender. Fifteen participants identified as Asian or Asian American (A), two as White (W), and one as Black (B). Thirteen participants were from Computer Science (CS). Other majors included Computer Engineering (CE), Aerospace Engineering (AE), Cognitive Psychology (CP), Civil and Environmental Engineering (CEE), and Computer Science Statistics. Ten participants were working professionals in the Data Science graduate degree program (DSG), six other graduate students (G) and two undergraduates (UG). Thirteen participants were non-native English speakers but reported no problems in understanding or speaking English (NNV2) and five native English speakers (NV). Three participants performed some voluntary edits without officially signing up in the caption editing activity.

PID	Gender	Age	Race	Major	Program	English Proficiency	Voluntary Edits
P1	F	18-24	A	CS & Stats	UG	NNV2	No
P2	M	35-44	W	CS	DSG	NV	Yes
P3	F	18-24	A	CS	DSG	NNV2	No
P4	M	25-34	A	CS	DSG	NV	No
P5	F	18-24	A	CP	UG	NNV2	No
P6	F	18-24	A	CS	G	NNV2	No
P7	F	18-24	A	CS	G	NNV2	No
P8	M	35-44	A	CS	DSG	NNV2	Yes
P9	M	25-34	A	CS	DSG	NV	No
P10	M	18-24	A	CS	UG	NNV2	No
P11	F	18-24	A	CE	UG	NNV2	No
P12	M	18-24	A	CEE	UG	NNV2	No
P13	M	25-34	A	CS	DSG	NNV2	No
P14	F	18-24	A	CS	DSG	NNV2	No
P15	F	18-24	A	AE	UG	NV	No
P16	-	25-34	A	CS	DSG	NNV2	No
P17	M	35-44	W	CS	DSG	NNV2	Yes
P18	M	25-34	B	CS	DSG	NV	No

CHAPTER 4: STUDY I: LEARNING OUTCOMES

4.1 LEARNING OUTCOMES OF MULTIPLE LEARNING PATHS

Students may learn about a particular system programming topic by attending a lecture (2015-2018), viewing a recorded lecture from a commercial capture system (2015-2018), viewing lecture video recorded for online use with ClassTranscribe (in Spring 2019 semester only), searching and/or reading lecture transcript (Spring 2019 semester only), reading the free course book in a variety of formats (epub,pdf,html; in all semesters), or viewing recordings of classroom lectures (previous semesters only). This is an example of application of UDL principles, where multiple learning activities and resources can be utilized by the student to achieve the same learning goal. Using the behavioral data captured in Spring 2019, it became possible to compare learning outcomes of alternative learning activities and validate student use of UDL materials in the course's design.

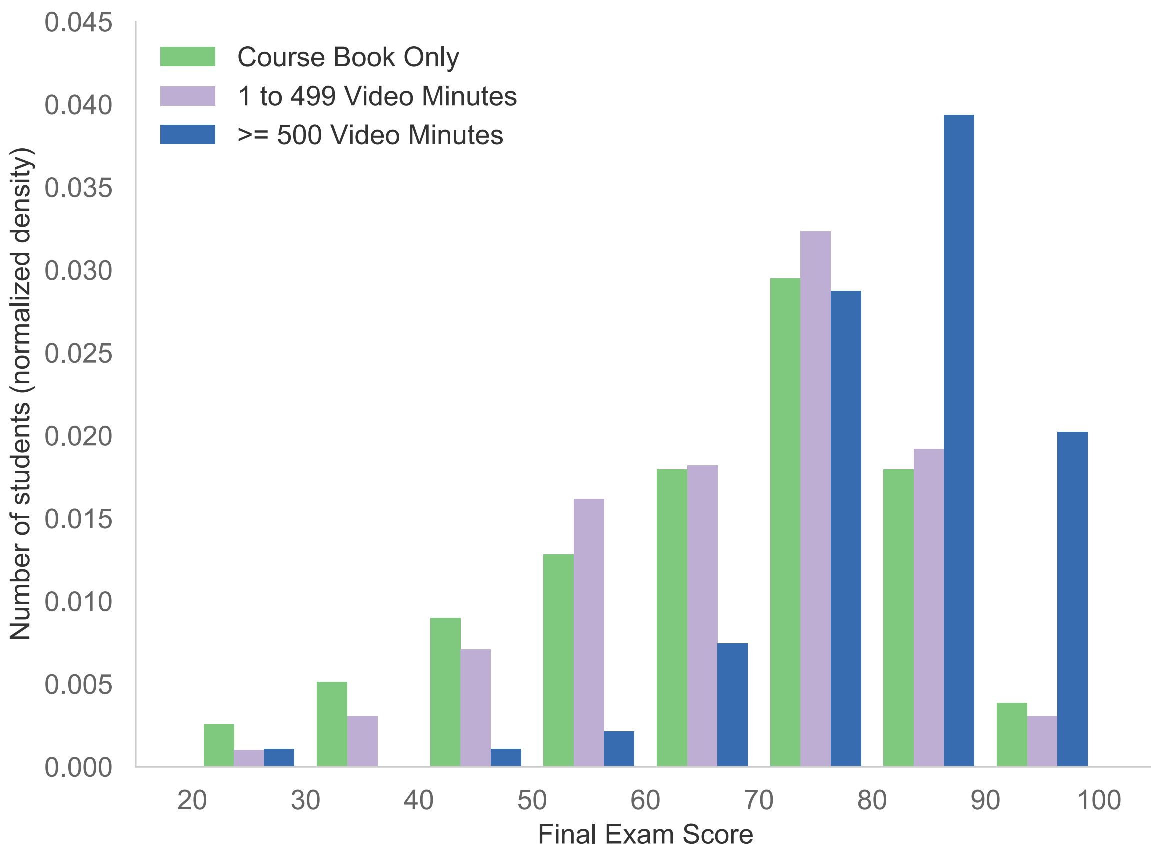


Figure 4.1: Histogram of final exam scores for different learning paths.

Table 4.1: Final exam score (out of 100) for students who chose to learn without using ClassTranscribe (Path-I), watched below average duration of video content in ClassTranscribe (Path-II), or above average duration (Path-III).

	Learner Path-I $n = 78(28.8\%)$ 0 minutes	Learner Path-II $n = 99(36.5\%)$ 1 to 499 minutes	Learner Path-III $n = 94(34.7\%)$ ≥ 500 minutes
Q_1	57.0	59.8	76.5
Q_2	70.4	71.4	82.4
Q_3	77.8	79.4	88.8

In Spring 2019 semester, out of 271 students taking the final exam, 194 (71.6%) students used ClassTranscribe and 77 students never viewed videos in ClassTranscribe (28.4%). This group was identified as the "Course Book" learners, as this was the most likely source that the students used to learn the content. It is possible a subset of students found and viewed previous semester lecture videos on the commercial lecture capture system, or co-watched the intended videos while another had logged in. The online course book is the most plausible learning path because it was positively mentioned in course feedback forms and linked from the course web pages. The 194 students who viewed at least 15 seconds of continuous video in ClassTranscribe (i.e. generated at least one *timeupdate* event) watched an average of 500.1 video minutes (equivalent to 8.3 video hours, and was rounded to 500 video minutes in subsequent analyses). Behavioral data were captured over 7 weeks, of which 6 contained content, corresponding to 83 minutes viewed per week. This time-on-task is lower than 3 live lectures total 150 minutes per week.

Students' primary choice of learning activity were characterized as 3 learning paths: i) No use of video lectures (i.e., the course book readers) ii) Below average use of video lectures iii) Above average use of video lectures. Reading events of the course book were not available. A limitation of this analysis is students may have further supplemented these expected learning activities by engaging in other relevant but untracked resources (for example, web searches, discussion forums, and discussions with their peers and course staff).

The final exam quartiles and histogram of the learning outcomes of the 3 learner paths are summarized in Table 4.1 and Figure 4.1 respectively. The primary finding was above average usage of ClassTranscribe led to improved exam performance. For example the median of Path-III learners was ≥ 11.0 point improvement on the final exam compared to learners choosing Path-I or Path-II.

A secondary finding was that a significant fraction of students can still earn competent – albeit on average lower – scores in a final exam by choosing alternative and secondary learning resources (Learner Path-I). This is encouraging feedback and validation for Universal Design advocates and practitioners who have created, or are considering creating, equivalent course resources in alternative formats.

4.2 FIRST, DO NO HARM

The final exam over all previous semesters used 45 randomized multiple choice questions (with randomized variants) that covered all topic areas of the course, and included concept questions and skilled application of a system programming idea within a programming context. Long form questions varied by semester, were graded manually using a rubric, and were excluded in this analysis. The exam scores of the Spring 2019 semester were compared to previous Spring semesters with the same instructor. Though programming and lab assignments have varied incrementally over time, the overall content, structure, lecture delivery and exam assessments have remained similar. The quartiles of multiple choice final exam scores are presented in Table 4.2. All quartiles of the Spring 2019 semester showed a modest improvement compared to previous semesters. A counter argument is to observe that quartiles show a mostly slow rising trend, and we hypothesize that active and blended learning in physical lectures or use of lectures videos has an insignificant role in this course. In either perspective, we can still conclude that the *dropping of lectures caused no significant harm* to exam performance in a lower-curriculum required class of the CS undergraduate program.

Can we therefore continue and conclude that lectures have no other intrinsic or extrinsic additional value? No. A limitation of this finding is that we have reduced learning outcomes and lecture value to the numerical performance in a final exam setting and not measured a student's experience and affect for Computer Science e.g. long-term interest, motivation, perseverance, and the ability to thrive in future academic and non-academic settings.

Table 4.2: Final exam multiple choice quartiles and 50% spread ($Q_3 - Q_1$), for students in Spring semesters of 2015,2016,2018, and 2019.

Statistic	2015 $N = 195$	2016 $N = 325$	2018 $N = 247$	2019 $N = 271$
Q_1	60.1	58.7	60.0	**63.6**
Q_2 (median)	68.1	69.6	71.1	**72.7**
Q_3	76.6	78.3	80.0	**83.3**
$Q_3 - Q_1$	16.5	19.6	20.0	**19.7**

Table 4.3: Within-quartile final exam improvement for each course performance quartile. For each quartile the exam mean and standard deviation are reported for below and above average minutes-viewed.

Course Performance Q_i (Rank)	Threshold τ_i (hours)	Exam $\langle score_i(t < \tau_i)\rangle$	Exam $\langle score_i(t \geq \tau)\rangle$	Mann Whitney p	Peer Improvement Absolute Δ_i	Fractional Reduction $\frac{\Delta_i}{100-\langle score_i\rangle}$
Q_1 (Rank 0% − 25%)	5.6	55.4 ± 16.3	69.9 ± 16.9	0.005	14.4	0.38
Q_2 (Rank 25% − 50%)	6.2	68.6 ± 12.2	78.6 ± 8.6	0.002	10.1	0.39
Q_3 (Rank 50% − 75%)	9.2	74.1 ± 10.4	80.4 ± 7.7	0.008	6.3	0.30
Q_4 (Rank 75% − 100%)	11.3	81.1 ± 10.1	86.2 ± 8.5	0.029	5.2	0.38

Lower performing students (Q_1 Q_2) may struggle to succeed due to diverse course-related, affect, and acquisition reasons (e.g. missing prerequisite knowledge or skills, incomplete, ineffective or inefficient learning strategies), as well as external non-course hardships (e.g. financial stress, limited time, reduced peer support, lower confidence, imposter syndrome). Apriori, i) It was unclear if ClassTranscribe could be used effectively by lower-performing students (compared to their peers) to improve their course-exit performance as measured by their final exam score; ii) However, the tool provided an opportunity for students to find and review relevant lecture material on any day and any hour of the day. Would this help ameliorate difficulties experienced by lower performing students? Similarly, could upper-performing students also gain from its use, compared to their peers?

This analysis explored if learning benefits were possible for both lower performing and upper performing students in the Spring 2019 semester. Students were grouped into performance quartiles based on their total weighted course score (with the final exam score excluded). The course score was comprised of numerous assessments that is typical of a CS programming-heavy course: challenging autograded weekly programming assignments, quizzes and programming problems given under exam conditions, and lab programming exercises. All assessments were individual work.

The total class enrollment (N=271 exam takers) was sufficiently large to evaluate 4 quartile groups each split into 2 subgroup conditions (below vs. above threshold of total-video-minutes-viewed). To compare ClassTranscribe use within each group, the average of video-minutes-watched-by-that-group-only was an appropriate threshold because it i) is a simple general definition that can be implemented in future analyses and replication studies; ii) is statistically expedient and likely to ensure a reasonable subgroup population size; iii) allows the comparison of student outcomes for students of a similar competency. Using a threshold based on total class average or fraction based on total video minutes were considered but rejected in favor of a per-quartile threshold due to the above 3 reasons.

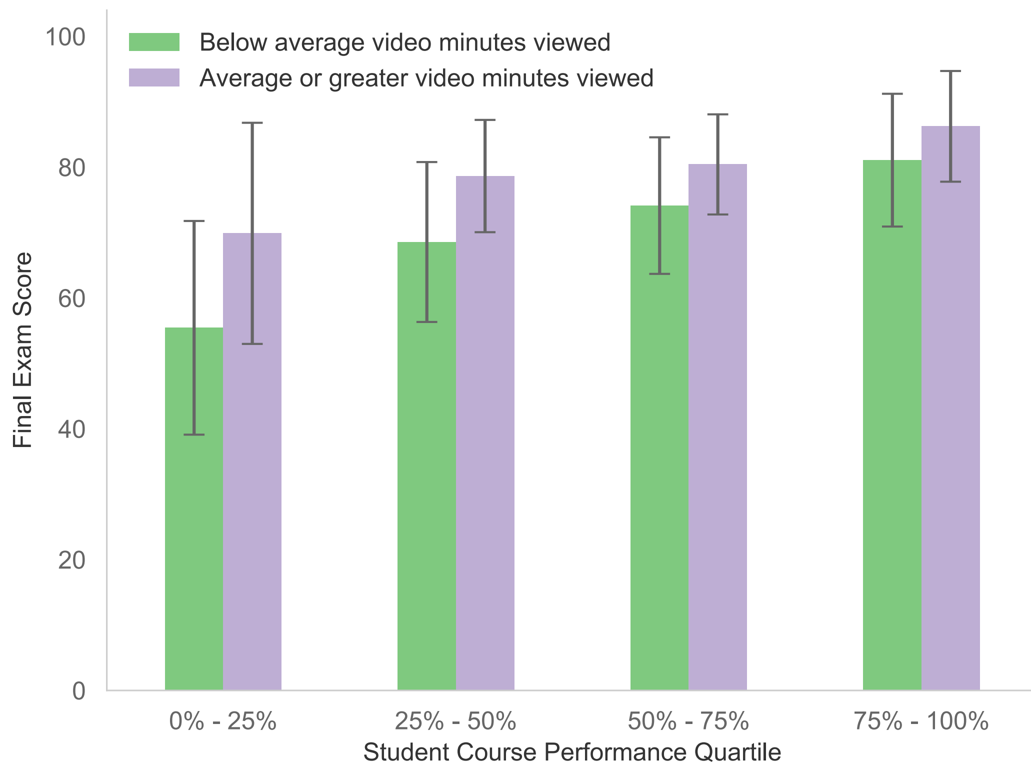


Figure 4.2: Overall Course Performance partitioned by Quartile Rank

For each quartile, students who used ClassTranscribe more than average for that quartile, surpassed their peers' exam performance in the same quartile by a significant amount (Table 4.3 and Figure 4.2). For example, on average, in the lowest performing quartile, a student could increase their exam score by 14.4 absolute points (out of 100) compared to another student in the same quartile who did not significantly use the tool. A Mann-Whitney U test [61], which does not require the normal distribution assumption, rejected the null-hypothesis. The exam performance increase for every quartile is significant at the $p < 0.05$ level.

The largest increase in exam improvement was exhibited by the lowest performing students (Q_1) using least additional studying (5.6 hours over the 7 week period). This supports the finding that use of ClassTranscribe is an effective learning technique that is open to students of all competency levels.

A second finding was that the benefit to all students was similar in range ($30-40\%$) when expressed as fractional reduction of average points lost on the final exam for each quartile. Replication of this fractional effect, understanding its magnitude, cause and scope, is worthy of future research. Further, a reasonable message to students of all abilities is, "Want to do well in the final? Take time to watch the videos over the semester and you can reduce your exam points you would have lost by a third."

4.4 WHAT BENEFITS?

We hypothesized that some events (e.g., using full screen playback) might be a proxy for a student committing their full attention to the content, and thus changing to fullscreen would be predictive of improved student outcomes. Other behaviors (e.g., pausing the video or switching to another activity) might be indicative of either partial attention and disengagement, or integrated behaviors where a student is using the video player in conjunction with another course-related activity (e.g., working on an assignment). No statistically-significant event measures (criteria Mann-Whitney U test, $p < 0.05$) were found that impaired exam performance. In addition to measuring the total video minutes viewed, searching for content, changing playback speed and using full screen were behaviors rejected by the null hypothesis (Mann-Whitney U test, $p < 0.01$) and predicted further improvement of final exam performance (see Table 4.4 and Figure 4.3). Other event types had no predictive power (i.e., could not exclude the null hypothesis). For example, there was no evidence that pausing the video led to either improvement or impairment of final exam performance.

Table 4.4: Behaviors that were predictive of exam performance. The mean and standard deviation for the final exam scores for students that included above threshold (group “Y1”) or below or equal threshold (group “Y0”) of the described behavior. Searching captions, changing playback speed and using full screen options were correlated with improved final exam score (approximately 4 - 8 points improvement).

Behavior	N(Y0)	N(Y1)	mean(Y0)	mean(Y1)	threshold	p-value
Search Captions	185	86	79.5 ± 13.4	83.4 ± 12.5	0.0	< 0.01
Change Speed	199	72	78.4 ± 13.7	87.2 ± 9.4	113.2	< 0.01
Full Screen	212	59	78.9 ± 13.7	87.5 ± 8.8	81.5	< 0.01

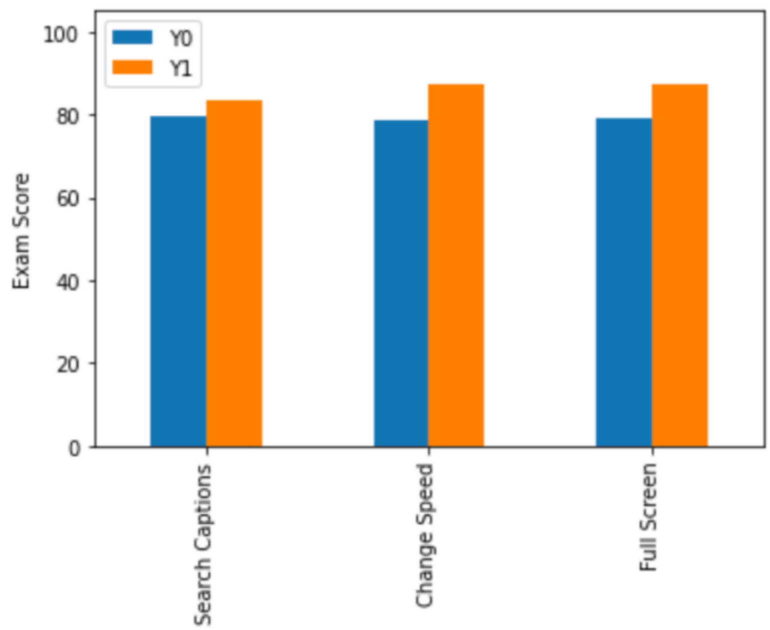


Figure 4.3: Student behaviors that were predictive of improved learning. Students who searched the transcriptions, changed video playback or used the full screen features more than average (Group Y1) than other students (Group Y0), had higher final exam scores than those who did not. See Table 4.4 for details.

CHAPTER 5: STUDY II: SEARCHING BEHAVIORS

5.1 WHAT DO SEARCHERS SEARCH FOR USING CAPTION-BASED VIDEO SEARCH?

5.1.1 Number of Searches

Among the $1,894$ students, the number of search activities per person ranged from 0 to 186 events. A time-series analysis of the students (N=167) who performed $1,022$ searches found search was used throughout the semester with peak usage occurring in the middle of the semester. One possible reason for the relatively low 9% (167/1894) usage rate of the searching function was that students were unaware that the search functionality existed. This is supported by survey results previously reported in [8], where only 40% (96/242) responded affirmatively to "Were you aware of ClassTranscribe's ability to search for course content using text search?". Thus an estimated usage rate for the students who were aware of this functionality would be 22% (167/758). These results suggest user interface designers of digital learning aids should consider the importance of awareness of the feature when tools are used in authentic settings that extend beyond a simulated classroom experience.

5.1.2 Categories of Search Queries

A manual analysis of the search queries data found three major categories of search terms: *STEM-related keywords*, *Logistics*, and *Others* (see Table 5.1). Among the 589 unique search queries, 427 (72.5%) were *STEM-related keywords*, 45 (7.6%) were on logistics of the class, and 117 (19.9%) fell into the *Others* category. In terms of the percentage of total searches ($1,022$), *STEM-related keywords* constituted 71.7%, *Logistics* constituted 9.5%, and *Others* constituted the remaining 18.8%. As the majority, *STEM-related keywords* were primarily domain words used in the class. We speculate that students were searching for these keywords to learn about specific concepts in the class. *STEM-related keywords* constituted 37 (74%) of the top 50 most frequently searched keywords. Another frequently searched keyword category was *Logistics*. We observed that students searched for keywords about academic integrity and course policies on collaboration for assignments throughout the semester. These topics are normally specified on the course website. However, a collaboration policy can vary significantly between courses, the same course in different semesters and even vary between instructors of different sections of the same course. Thus it is unsurprising that we found searches related to collaboration and cheating.

The *Others* category consisted of more general search queries, which we further categorized into i) *search-functionality exploration*, ii) *direct navigation*, and iii) *indirect navigation* behaviors that we discuss here. Queries that explored the search functionality used words that were known to be in the transcript (e.g. “hello” and “i”). *Direct navigation* searching used a specific topic or course item. For example, some students searched for “L25” (lecture 25), “mp 7” (machine problem 7), “question eight”, and “csNNN” (course number anonymized) to directly navigate to the relevant point in the video. Lastly we defined *indirect navigation* where students recalled that the instructor uttered a key or unique phrase and used this as a proxy for a specific moment in a video search. Example search queries included, “essentially”, “actual”, “i thought i knew”, “i don’t have the operator.”

Table 5.1: Categories of the search queries with corresponding counts.

Category	**Sub-category**	**Example queries and number of search occurrences**
STEM-Related Keywords (72.7%)		malloc (31), process (18), intro to diodes (17), deadlock (16), Bresen (10)
Logistics (7.7%)		cheating (17), collaboration (10), plagiarism (9), academic (5), lab (3)
Others (19.6%)	Exploring Search Functionality	i (8), your (5), hello (5), little (4), everything (3)
	Direct Navigation	csNNN (5), L25 (2), mp seven (2), problem/question eight (2)
	Indirect Navigation	Essentially (1), actual (1), i thought i knew (1), i don’t have the operator (1)

5.1.3 Zipf’s Law of Search Keywords

Zipf’s law is an empirical law proposed by linguist George Kingsley Zipf [62, 63], who observed that in a large corpus of written words, the frequency of any word was inversely proportional to its rank in a frequency-of-use table. This is often roughly described as “there will be a few common words and a large number of rare words,” however, Zipf’s law is a stronger empirical observation that the frequency of utterances follows a power law distribution. Previous research shows that search queries follow Zipf’s law, (e.g., on domain-specific queries on Google [64]) . So, we aimed to investigate whether this law also holds on queries in educational videos.

As shown in Figure 5.1, we plotted the top 200 search keywords by its frequency rank and number of occurrences, and utilized the scipy package to generate the fitted curve following the Zipf's law (kx^{-s}), where x is the frequency rank of and k,s are parameters. The best fit was found at $k = 32.1$, $s = 0.58$. A two-tailed Kolmogorov–Smirnov test [65] between the actual and fitted curves indicated that the number of occurrences of keywords follows a Zipfian distribution, $D(592) = 0.41$, $p < 0.005$. We concluded that Zipf's Law also holds for video-based searching in an educational context. We expect this to hold true especially in case of searches within domain-specific courses (e.g., engineering) because the queries are generally coherent and coherence is an important requirement for Zipf's law [66]. For example, as described in Section 5.1.2, STEM-related keywords constituted the majority of searches in our dataset.

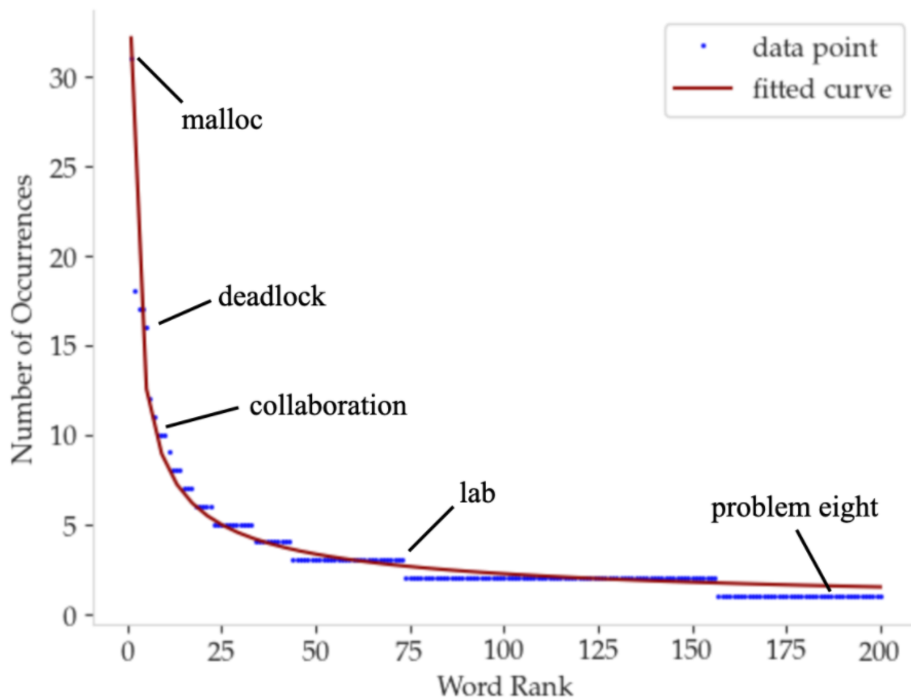


Figure 5.1: Zipf's Law in Search Keywords

5.2 WHAT COLLECTIVE BEHAVIORS DO SEARCHERS EXHIBIT BEFORE, DURING AND AFTER SEARCH?

5.2.1 Classification of Users Based on Search Frequency

As stated in Section 5.1.1 above, the total number of searches per student has a large range of [0, 186]. One hypothesis is that more active users, i.e., users who have more events overall on the tool, performed more searches as well. To check this hypothesis, we plotted the number of searches vs. total number events for all users as shown in Figure 5.2. As we can see from the plot, this was not generally true; the students who performed no searches are spread across all activity levels. There is one outlier point with > 1 million activities and 186 searches, that is not shown in the plot. After removing the outlier, we computed the Pearson correlation coefficient and found no linear correlation $r(1893) = 0.05$, $p < 0.05$. This suggests that there can be variances in search usage among students with similar activity levels. To further study these variances, we performed k-means clustering [56] of users based on their number of searches as described below.

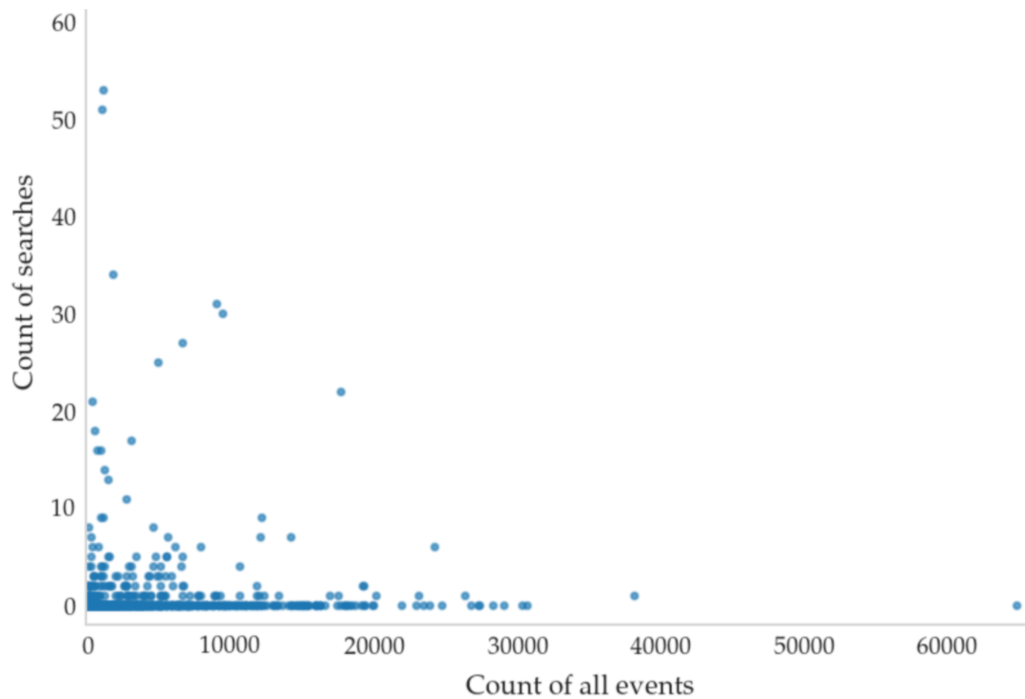


Figure 5.2: Total searches vs. the total logged records. Each dot represents one student.

Firstly, to determine the optimal number of clusters, we used the Elbow method [67]. Figure 5.3 shows the With-in-Sum-of-Squares (WSS) distance vs. number of clusters. Based on this plot, we selected 3 as the optimal number. Next, since k-means can sometimes give suboptimal results, we performed k-means ten times with different centroid seeds and selected the results with the best WSS. Figure 5.4 shows the results of the k-means algorithm. From the results, we identified two main types of students based on their counts of searches: "Infrequent Searcher" (≤ 17 searches) and "Frequent Searcher" (18 to 110 searches). These thresholds are the midpoints between two (sorted) cluster centroids. There was only one student with > 110 searches, who is identified as an outlier and excluded in further analyses. After removing the outlier, there were 166 students with 837 total searches. Further, since our focus is on students who performed searches, we used “Infrequent Searcher” for students with 1 to 17 searches and “Frequent Searcher” for students with 18 to 110 searches. Using this categorization, we found 155 Infrequent Searchers and 11 Frequent Searcher users.

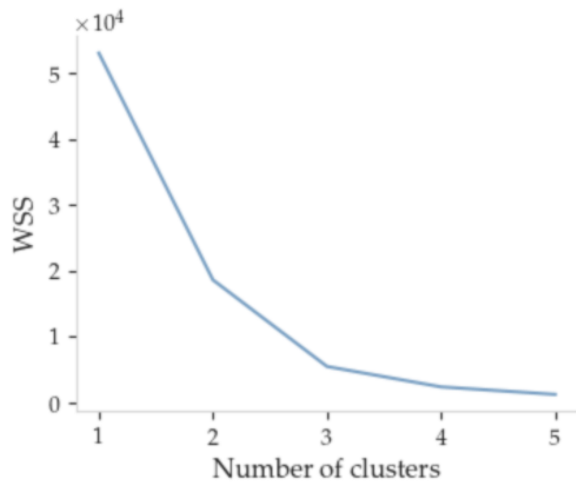


Figure 5.3: Determining the optimal number of clusters using the Elbow Method

Figure 5.4: k-means clusters of students based on their counts of searches. Colored dots represent the students clustered into 3 clusters. Black bubbles are cluster centroids; cluster threshold boundaries (not shown) were equidistant between neighboring centroids.

To further investigate the differences between the groups, we additionally looked at their number of all activities performed and their durations on the tool. Here, duration refers to the time difference (number of days) between a user's first logged activity and their last activity on the tool. A Mann-Whitney U test [68, 69] was used to compare these distributions for Frequent Searchers and Infrequent Searchers, i.e., the two extremes. We used this test because, unlike the student t-test, it does not require the assumption that the data are normally distributed. Based on the test, we found the following: (1) the **duration** of Frequent Searchers ($mean$ = 157.2 days, $std.$ = 106.8) is significantly longer than the duration of Infrequent Searchers (*mean* = 66.3 days, *std.* = 70.6); $z = 3.22$, $p < 0.005$; (2) the **number of all activities** performed by Frequent Searchers (*mean* = 5073.4, *std.* = 5330.4) and Infrequent Searchers (*mean* = 4349.8, *std.* = 5839.1) is not found to be significantly different; $z = 0.9$, $p = 0.36$. Further, using a Chi-square test[55], there was a significant relationship between the searcher type (i.e., Infrequent Searchers or Frequent Searchers) and event type (i.e., Search or Not Search) , $\chi^2(1, N = 4,574,297) = 13,816.88$, $p < 0.01$. This suggests that Frequent Searchers not only performed more searches but were more likely to perform Search over other Event Types and vice versa.

We note that the two groups had different sample sizes and the number of Infrequent Searchers is small. However, we used statistical tests (Mann-Whitney U and Chi-square tests) to compare them and found significant p-values < 0.01 suggesting that they are different. Both the tests are shown to work well with unequal sample sizes [68, 70]. The Mann-Whitney U is also known to be robust for reasonably small sample sizes [68]. For the Chi-square test performed above, the small number of Infrequent Searchers was not a concern because the corresponding expected frequencies (cells in contingency table) were not small (> 300).

5.2.2 Activities Before and After Searches

To investigate searchers' activities in terms of their interaction with the system immediately before and after a search event, we aggregated events surrounding the search activity separately for different groups of searchers (Figure 5.5). In this figure, "I" stands for Infrequent Searchers, and "F" stands for Frequent Searchers. The y-axis represents the event offset from the search activity event. For example, "+1" identifies the exact next event after the search, and "−2" identifies the second event before the search event. Each colored bar represents the sum of counts of a particular activity for a given group and their corresponding percentages are shown.

By comparing the events surrounding the search activity, we found that Infrequent Searchers tended to continue watching (e.g., mostly performing events like *timeupdate*, *seeking*, *changevideo*), while Frequent Searchers tended to continue searching (i.e, *filtertrans*). The continous/consecutive search behavior is discussed in more detail in Section 5.2.3. We noticed that Infrequent Searchers were more likely to start searching from the state of *userinactive* (8.5% at -1 offset) compared to Frequent Searchers (under 4% at -1 offset), which means they switched to ClassTranscribe from another webpage to perform the search. Our hypothesis was that these students were probably working on homework assignments and trying to look for information. However, we would need more information to support this.

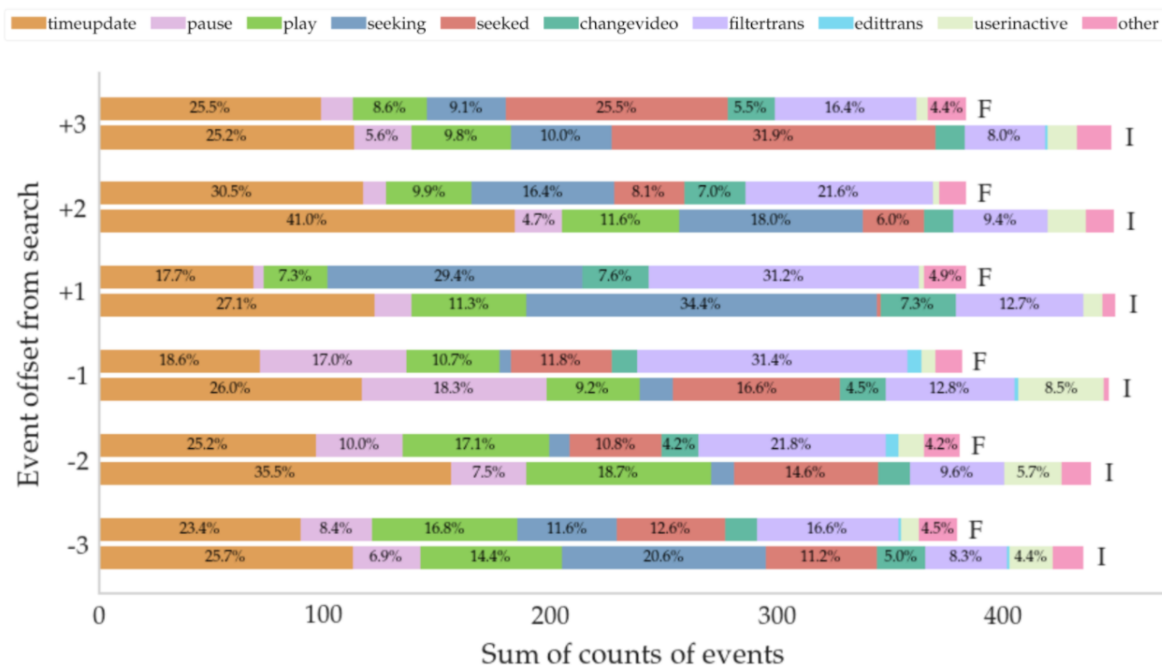


Figure 5.5: Distribution of events before and after a logged search event, aggregated for Infrequent Searchers (I) and Frequent Searchers (F).

We also note that overall, the number of activities performed collectively by Frequent Searchers around a search was only slightly fewer than that performed by Infrequent Searchers. This is expected because Frequent Searchers perform more searches, so there's overall more activity around search for Frequent Searchers. Although, at first glance, we might expect the number of activities collectively performed by all Infrequent Searchers to be much higher compared to that of all Frequent Searchers because 1) the number of Frequent Searchers is smaller than the number of Infrequent Searchers (11 vs. 155); 2) Frequent Searchers and Infrequent Searchers generally performed the same total number of all activities over their entire durations on the tool. However, the high number of Frequent Searchers searches compared to those of Infrequent Searchers seemed to mostly nullify the potential higher counts due to those two statements.

Finally, we observed that *seeking* and *changevideo* activities were more likely to occur immediately after a search event compared to before for both types of searchers. For Infrequent Searchers, the *seeking* event comprises 34.4% of all activities immediately after the search compared to under 4% immediately prior, for example, about a third of searches, Infrequent Searchers clicked on a caption presented in the search results and seeked to a location in the same video suggesting they likely found relevant results.

5.2.3 Search Sequences and Refinement Strategies

By manually coding the search queries, four major types of caption-search behaviors were identified. They were 1) *search and stop*, 2) *repeated search on the same query*, 3) *search again with minor modifications*, and 4) *search again with major modifications*. The "*earch and stop*" behavior happened frequently for Infrequent Searchers users. We speculate that they either found the information they needed or failed and lost interest in searching. Conversely, Frequent Searchers users were more likely to repeatedly search for the same queries or modify their search queries and search again.

For *repeated search on the same query*, we suggest this was employed when students were performing a focused search on a major course concept. For query modifications, we found that users generally either changed their queries a small amount or changed their queries significantly but still under the same topic. An example sequence of *minor modifications* was "eight", "problem eight", "question eight" and, "eight". We speculate that this student recalled that the instructor explained the eighth problem and tried to retrieve that part, but they were unsure what exactly the instructor said or how it was transcribed. So they started with a general query "eight", and then tried "problem eight" and "question eight". However, they ultimately went back to searching for "eight", so we suspect that the student observed too many results by searching "eight", but too few results by searching "eight problem". Another example of a sequence of searches with minor modifications was "Bresenham's" "bresen" "Bresenham's" "Bresenham's" "Brese" "Bresenham's" and finally, "Bresen".

For searches with *major modifications*, we observed that one user searched for "doubling" "3t3" "cell" "mus musculus" and, "mouse". We speculate that they were trying to retrieve relevant moments in the videos that discussed related bio-engineering concepts. On the other hand, students also optimized their search queries based on the auto-generated captions. For example, one user searched for "calloc" and then "kellog"; we speculate that they noticed that the term "calloc" was often mistranscribed as "kellog" so optimized their search queries accordingly.

Overall, we identified two types of users, Infrequent and Frequent Searchers who use the tool very differently. Infrequent Searchers tend to perform a large number of activities within a shorter period of time with a likeliness to perform other Event Types, whereas Frequent Searchers perform the same number of activities over a longer period with a likeliness to perform searches (*filtertran*). Further, Frequent Searchers tended to continue searching around a single search event. Table 5.2 summarizes their main differences.

Table 5.2: Summary of differences between Infrequent Searchers (IS) and Frequent Searchers (FS).

	IS (N=155)	**FS (N=11)**
Number of searches	Significantly Lower (mean=2.92, std.=3.03)	Significantly Higher (mean=34.91, std=15.98)
Duration (days) on too	Significantly Shorter (mean=66.3, std.=70.6)	Significantly Longer (mean=157.2, std.=106.8)
Total number of activities	No statistically significant difference found (mean=4349.8, std.=5,839.1)	No statistically significant difference found (mean=5,073.4, std.=5330.4)
Likeliness of performing search vs. other Event Type	More likely to perform other Event Types	More likely to perform search (filtertrans)
Top 3 activities immediately before search	timeupdate (26%), pause (18.3%), seeked (16.6%)	filtertrans (31.4%), timeupdate (18.6%), pause (17%)
Top 3 activities immediately after search	seeking (34.4%), timeupdate (27.1%), filtertrans (12.1%)	filtertrans (31.2%), seeking (29.4%), timeupdate (17.7%)

5.3 IS CAPTION SEARCH USED TO REVIEW PREVIOUSLY VIEWED CONTENT OR TO FIND NEW CONTENT?

We used the event log analysis to explore what video content was viewed after a student completed a search. We also asked if students were more likely to search when viewing new content or reviewing content.

To address the latter question we first defined a video review as the user performing any activity on a video at least one day after watching the same video. The one day threshold was justified as a reasonable time-period because a user may continue watching the same video over the duration of one day even if they take multiple breaks in between. Exploring alternative definitions and alternative review thresholds is out of scope for this study and a focus of future work.

Table 5.3 presents the number of searches and other activities that occurred during a review. We can see that 29.7% of the searches occur during a review compared to 19.4% of other activities. Using a Chi-square test [55], there was a significant relationship between the event type (i.e., Search or Not Search) and review, $\chi^2(1, N = 730,031) = 57.0$, $p < 0.001$. This suggests that searches were more likely to occur during a review. In terms of the number of users, 33.1% of all searchers performed at least one search during a review.

Table 5.3: Contingency table for a Chi-square test of association between searches and video review.

	Review	**Not Review**	**Total**
Search	249	588	837
Not Search	141,237	587,957	729,194
Total	141,486	588,545	730,031

We acknowledge that users may have attended the on-campus version of some of those lectures, and hence may actually be reviewing the material during their first time use of the tool for that content. However, this limitation does not diminish the above finding; that students were more likely to search for content when reviewing the same video content again.

To explore whether students generally navigated to a previously unwatched video from the search results, we computed the percentage of videos that students were watching for the first time on the tool when students navigated to the video after a search (i.e., a *changevideo* event). Out of the total 837 searches, 7.3% (61/837) led to a *changevide* event. We found that 77% (47/61) of those searches led to videos being viewed for the first time and 67% (41/61) occurred during a Review (see Table 5.4).

Table 5.4: Number of searches that led to a changevideo event. Rows represent searches that start from Review/First-time View and columns represent searches that go to Review/View

	Review	**First-time View**
View	11	6
First-time View	3	41

CHAPTER 6: STUDY III: ATTITUDES ON LEARNERSOURCED CAPTIONING

6.1 HOW DO STUDENTS USE CAPTIONS FOR LEARNING?

In this section, results on how students use captions for learning will be presented. The results consisted of two parts: (i) why and how do students use captions; and (ii) captions' impacts on students' learning experience in video-based online education.

6.1.1 Why and How Do Students Use Captions

Students mentioned multiple reasons for using captions during watching lecture videos. Captions could help them to focus on the course content, process inaudible parts, understand new concepts and their definitions, preview the upcoming content, understand equations and jargon, understand instructor's accent, and improve efficiency for multitasking. In addition, as one non-native English speaker (P-ID: 12) shared, he turned on the captions also to learn English - the process of reading captions help him understand English and learn it.

Students used captions in several ways. Caption was reported to be used either during first-time viewing or review, and for some students (N=6) it was both. Several students (N=5) reported that they used captions for note-taking purposes. They could refer to the text in the captions, and copy it in the transcripts if they needed. Some students (N=4) reported that they turned on caption while watching the video at a faster speed. In this scenario, captions could help them focus and understand the content faster. In addition, students pointed out that they found reading transcripts handy because they could skim through the text to quickly get an overview of the topic. Further, students mentioned that they used keyword search in the transcripts to find information they need. However, two native English speakers (P-ID: 2, 15) reported that they seldom use the captions. Although one of them (P-ID: 2) mentioned they would sometimes refer to captions when the lecture video was inaudible even after pausing and re-listening.

6.1.2 Caption's Impacts on Learning

Captions have positive impacts on student learning experience. Students reported that captions could help them understand new concepts and enhance learning. Some students (N=2) thought captions helped them to better arrange attention in-class and focus on important content. Some students thought captions made the note-taking process easier (N=5), and could help them with multitasking (N=2).

Several students (N=4) mentioned that captions could make the lecture videos more accessible, especially for students who are non-native speakers. Non-native English speakers (P-ID: 6, 11) reported that captions could help to reduce their cognitive burden while watching the lecture videos because they thought reading was easier than listening.

6.2 HOW DO STUDENTS EDIT CAPTIONS?

In this section, results on how students edit captions will be presented. The results consisted of four parts: (i) different kinds of errors edited & examples, (ii) editing strategies used by students, (iii) motivations for editing captions, and (iv) perceived difference between being the first and second editor.

6.2.1 Different Kinds of Errors Edited & Examples

Students fixed different types of errors. Please see the Figure 6.1 for a summary of the category of fixed errors in the captions and some concrete examples.

Category	Examples
Punctuations	Adding a comma to the end of the sentence
Capitalization	"Scores" -> "scores"
Equations	"are one" -> "R1"
Newline	Adding a newline when the sentence needs a break
STEM word	"an LP" -> "NLP"
Other	"Pasta" -> "tester"

Figure 6.1: Category of fixed errors and examples

6.2.2 Caption Editing Strategies

Participants reported various kinds of strategies they used to edit the captions. One popular strategy mentioned was referring to other course materials (textbook, lecture slides or visual content in the video, the same video on Coursera, and so on) when they were not sure about how to transcribe certain words. Other commonly reported strategies include replaying a part of the video and re-listen, slowing down the playback speed and listen carefully, and combining domain knowledge and context to guess when the speech was inaudible. Another popular strategy mentioned by students was to check for consistency between audio, visual and text information - they hoped to make the captions consistent with other course materials presented in the class. Nevertheless, students indicated that they were often not sure what the standards should be. This result suggested that it could be useful to provide a guideline with standardized editing glossary and procedures.

One student (P-ID: 6) pointed out that sometimes she would go through the video twice: first time to get the general meaning of the video, second time to make close edits, especially for equations. This student also noted that sometimes they needed to move words from one line to another to align with the timeline. Another student (P-ID: 13) reported that he made the font size bigger to clearly see the text, paid attention to timestamp - whether audio was synced with the caption, used browser-search to find and fix all instances of the same error together, and played video with caption again after all edits to double check. Interestingly, one student (P-ID: 15) mentioned that initially she followed the speech and tried to find errors in each line. Later she was able to identify common cases where the algorithm would make mistakes and was able to correct them faster, though she still went line by line to make sure the edits were correct.

6.2.3 Motivations for Editing

Since a 1% extra credit was offered for each set of videos edited and reviewed, most students recognized that one of their motivations for editing was the extra credit. Other than the extra credit, a common motivation reported by students was the contribution to the class - their edits to the captions could improve the course and benefit their peers who are or will be taking this course. Students took it as a way to give back to the class. In addition, some students reported that the caption editing and reviewing activity could help them understand the course materials better. For example, one participant (P-ID: 6) mentioned that she had an exam 1-2 weeks after the caption editing and reviewing activity, and she thought this activity helped her prepare for the exams because normally she would not be able to pay this much attention to every single word, and she scored higher in their final exam after the activity than midterms before the activity. Variously, a student (P-ID: 7) reported that she participated in this activity for fun, and suggested that the activity could have more gamification elements in the future. Another motivation mentioned by a student (P-ID: 4) was that they would like to see transcripts in their preferred way of writing (e.g.

grammar, capitalization etc.)

6.2.4 Perceived Differences between Being the First and Second Editor

This subsection presents students' perceptions of the differences between being the first and second editor and how they collaborate with each other. The results are polarized - students either felt no difference or very different between the two roles.

For students that thought there was not much difference between being the first and second editor, they generally applied the same criteria when doing the editing and reviewing, and the knowledge of being first vs. second editor did not affect their motivations or sense of responsibility. However, for those that felt there was indeed a difference, their attitudes and behaviors were different between the two roles. Some students thought the first editor could focus on major errors while second editor could check minor errors. Some students thought first author should edit more than the second editor. Therefore when they were the second editor, they would not be as active in editing as when they were the first editor. These students also reported that they spent less time as the second editor. They felt more relaxed and paid less attention as they thought the first editor should take the major responsibilities. For instance, one participant (P-ID: 11) shared that she felt less responsible and watched the video at twice the speed when serving as the second editor. However, some people had different opinions on this. For example, one participant (P-ID: 2) noted that as the second editor, he would be more careful in making any major edits because he did not want to step on anyone's toes or question their work. From his point of view, he was willing to take more risks when being the first editor.

6.3 STUDENTS' SUGGESTIONS FOR IMPROVEMENTS

In the interviews, students shared some suggestions for improvements, on the system for editing captions and the mechanism of the caption editing activity. A Summary of students' explicit suggestions is presented in this section.

6.3.1 Suggestion 1: Provide a Clear Guideline for Caption Editing

Students suggested that guidelines should be developed to improve caption editing efficiency. For example, it would be helpful if they could know what kind of errors to focus on that can help the most. Ideally, there could be a tutorial highlighting different kinds of errors and how to fix them (e.g., filler words, mid-sentence wording changes, etc.) When asked about guidelines for STEM-related symbols, student (P-ID: 2) said it should be useful for those without enough background knowledge. Another participant (P-ID: 11) suggested that it would be useful to provide students with a list of key concepts with their explanations per lecture (i.e. a glossary) to help with transcribing.

6.3.2 Suggestion 2: Visualize Contribution as a Motivation

Students suggested that if their contributions could be visualized, then they would be more motivated to edit the captions. For example, there could be a healthy leaderboard competition based on the number of edits or simply publish top contributors' names in a class-wide message. This kind of recognition would make students feel more motivated to contribute to the captions. One student (P-ID: 7) suggested that there could be some gamification features added to the user interface of the caption editing panel, such as levels of editors based on the number of edits, and animation for editing, for example, an animation of fireworks every 100 edits.

6.3.3 Suggestion 3: Announce the Activity Early

In this study, the caption editing activity was announced late when the semester was about to end. Students suggested that it would be better if the activity could be announced earlier when the semester just began. In this way, students could decide when they want to do the edits. In addition, releasing the activity at the beginning of the course would enable students to edit the captions during the first time of watching the videos and do not have to re-watch them for this activity.

6.3.4 Suggestion 4: Automate the Repetitive Tasks

Some students complained that they were editing some repetitive errors a couple of times during their caption editing experience. Students thought that it would be helpful if the system could automatically detect similar errors based on their previous edits to automatically correct some errors or make suggestions for new edits. This automation could save the editor's time and make the process more efficient.

6.3.5 Suggestion 5: Add a Moderator to Check the Edited Captions

Students suggested that it could be beneficial to have a moderator (for example, a teaching assistant) that has adequate knowledge about the course to check the edited captions. Knowing that their work would be checked by a moderator could make the editors be more careful and motivated to make better quality edits with fewer errors. In addition, as one student (P-ID: 18) suggested, extra credit could be based on the quality of edits made. Although quality might be subjective, standards could be made by the instructor. For example, giving more weight to fixing equations than grammatical errors. Adding a moderator could better assure the quality of the edited captions.

CHAPTER 7: DISCUSSION

7.1 A TAXONOMY OF SEARCH QUERIES

Taxonomies of search queries have been developed for other domains such as general web search [71], e-commerce [72], and educational queries on the web [73] but not for caption-based search for educational videos. In Section 5.1.2, we categorized the search queries that students used for retrieving content for video-based learning in an authentic classroom setting. This provided a better understanding of the students' informational needs and a useful classification for additional research in this area.

7.2 SEARCHING FOR IMPERFECT AND INACCURATE CAPTIONS

In Section 5.2.3, we observed that errors existed in automatically generated captions. This is unsurprising as generating captions especially for STEM videos is challenging [6]. We found that students used strategies to modify their queries to match the erroneous words (e.g., search for "kellog" in place of "calloc") to overcome those errors. We suggest phoneme (audio) indexing (e.g. Soundex [69]) would facilitate successful matching for similarly sounding words and phrases that are misspelled or mistranscribed. Further it might be possible to train machine learning models to learn from student search strategies (e.g., based on phonetic similarity of consecutive search terms) and utilize them to correct errors in the lecture video captions.

Some users also corrected the transcripts using the *edittrans* function supported by the platform (refer to Figure 5.5), though the amount was only a small fraction of their activities ($< 4\%$). In other words, occasionally, when users were searching, they might observe a caption error and be motivated to correct it. Future education tools could use this opportunity to nudge students to correct those errors because this would immediately help them find better results. Learnersourcing [74, 75] uses a similar approach to motivate learners to improve educational content such that it benefits both individuals and their peers. For example, learnersourcing has been used for editing foreign language captions [51].

7.3 SEARCH ALGORITHMS

Firstly, from sections 5.1.2 and 5.1.3, we found that the majority of the frequently searched terms were STEM-related. Thus, it would be productive for future work to focus efforts on improving the accuracy and utility of the tools for searching domain-specific keywords. However, there is also a long-tail of search queries. Similar to other search domains like general web [52], improving the accuracy for the long tail of all "rare" and unique queries may be challenging.

Secondly, to further study and improve accuracy of search algorithms, it is important to define what constitutes a successful search. The event sequences described in this study could be used to build a naive indicator of a successful within-video and within-course search activity. Currently, success is based on an increase in the number of seeking activities after search (Section 5.2.3) and we found that keyword-based search seems reasonably sufficient perhaps because students search for terms that are present in the lecture captions (e.g. STEM terms discussed in Section 5.1).

However identifying successful searches based on event logs will require additional studies. For example, the full caption text of the search results might have satisfied the student's question, or the subsequent duration of video viewed, i.e. dwell-time, (vs. the user performing additional searches or selecting alternative video moments) may be a useful measure of the value of the video clip(s) proposed by the system or selected by the user. Beyond clicks and dwell-time, researchers could also measure the utility of the search results as it is a better indicator of success [76]. For example, researchers have used questionnaires to manually measure information learned during educational video search in a lab-setting [54]. Further research is required for automatically measuring utility of search results in authentic settings.

Thirdly, we found that students used multiple consecutive search queries to potentially find a better and complete set of search results (Section 5.2.3). Previous work [54] showed that students who used fewer queries (keywords) during video search had better learning outcomes. Thus, a future research question is, "Is it possible to automatically suggest precise queries by training machine learning models on search log data and caption data?" For example, automated Query Suggestion and Reformulation techniques for general web search [53] or K-12 educational search [77] could be adapted for caption-based educational video search. However, care must be taken as to not distract students or increase their cognitive burden [54].

7.4 STUDENT-DRIVEN CONTENT

Our findings in Section 5.3 showed the students used search to navigate to previously unwatched videos suggesting that search could be used as a means to discover new (unwatched) educational videos in a targeted way. Such non-linear navigation patterns (i.e., not strictly following the lecture video order defined in the course) using caption search suggests that the search log activity could potentially be used in two pedagogically interesting ways. Firstly, to automatically recommend relevant video segments to students. Secondly, to structure the course content in a way that students tend to find more useful compared to the traditional fixed course structure that is defined by the instructor. Using log data to augment educational videos has been shown to be useful to students [48]. Our exploration provides more insights on leveraging log data for video-based learning. For example, we found that students search when reviewing a video a second time (Section 5.3). Future studies could manually or automatically [78] identify student intent behind search queries, e.g., preparing for exams vs. working on assignments. Combining intent with the log data, it may be possible to automatically create exam helper modules that map difficult exam concepts (i.e., user search terms) to the corresponding video time segments that students found most useful.

7.5 IMPLICATIONS FOR UDL

Universal Design for Learning provides a framework and context to construct and deliver accessible and inclusive courses. Moreover, when implemented, UDL provides each student with multiple learning pathways that can span and intersect with multiple content modalities (e.g., video, captions, figures, audio and descriptive text). The caption-based video search function could potentially contribute to all three principles of UDL: "provide multiple means of engagement", "provide multiple means of representation", and "provide multiple means of action & expression." We suggest that the caption-based video search function should be mapped to the "multiple representation" category of the UDL principles since it provides learners with a new learning pathway to retrieve content they need and customize the display of information. The results of Study I (reported in [8]) suggested that search behaviors were predictive of improved exam scores. From the perspective of UDL - it is unsurprising that student search behaviors cannot be characterized by a single prototypical search behavior. Instead engineering students used "Infrequent Searchers" and "Frequent Searchers" behavioral patterns to find and review relevant content (Section 5.2). As the UDL framework has been discussed and adopted to support the needs of all students [79], our study showed that caption-based video search could still be improved to support varied navigational learning behaviors and needs in video-based educational platforms. For future work, we will further investigate types of queries by frequent and infrequent searchers and provide design guidelines for caption-based video search. We encourage all instructors to adopt a text-searchable and accessible video platform, as part of a larger UDL-based approach to effective, accessible, and inclusive education.

7.6 IMPLICATIONS FOR LEARNERSOURCED CAPTIONING

Study II showed that captions are helpful for indexing and searching videos. In Study III, we further investigated why and how students use captions in video-based online learning. We found that captions helped students to focus on the course content, process inaudible parts, understand new concepts and their definitions, preview the upcoming content, understand equations and jargon, understand instructor's accent, and improve efficiency for multitasking. Students used captions to take notes, understand inaudible speech, and search across the video. Results suggested that captions help to improve the accessibility of the lecture videos, especially for students who are non-native speakers. Therefore, captions could be used to enhance video-based online learning environment.

Students helped to fix errors in automatically generated caption for lecture videos. Interview results suggested that students could fix all the six types of errors. We investigated students' editing strategies, motivations, and their perceptions of collaborating with a co-editor, either as a first-time editor or a reviewer. The results helped us to understand how students individually or collaboratively edit the captions. Combining this understanding with students' explicit suggestions for improvements, we interpreted potential new features or interface design that support students' caption editing efforts.

To support students' editing strategies, for example, referring to other course materials, we should provide them with guidelines of editing standards and procedures, with a glossary of common terms and equations used in the class. In this way, the editor could find resources they need and be more confident when they are faced with complicated terms or equations. On the other hand, students hope to make the captions consistent with other course materials presented in the class. In addition to providing editing guidelines, we could also work on the system side to add some automatic features such as auto corrections or suggestions based on editors' previous edits.

Other than the extra credit, students had multiple motivations for editing the captions. They are willing to give back to the class, and would be happy to be recognized for their contributions. Therefore, giving students credits for their contribution could help to motivate them for editing captions. For example a leader board competition of top editors could be implemented. Students also mentioned that adding animation features could engage them in the caption editing task. However, there is a possibility that native English speakers are more likely and feel easier to edit captions than non-native English speakers. To provide equal opportunities for all students, we offered other extra-credit activities simultaneously.

Finally, we explored students' attitudes, motivations, and strategies in collaborating with other learners to fix captions. Students shared their thoughts on the difference between being the first and second editor. Some students felt that the first editor should be more responsible, needed to do more work and pay more attention. They reported that they found themselves feeling more relaxed, spending less time, and paying less attention when being the second editor as they thought someone else had already fixed major errors. Students suggested that adding a moderator (e.g. a course staff) to check their edits might ensure better caption quality. In future research, we would further explore how to better support the collaboration between student editors.

CHAPTER 8: CONCLUSION

This thesis consisted of a series of three studies on students' attitudes, behaviors, and learning outcomes from using ClassTranscribe, a Universal Design for Learning (UDL) featured video-based online learning platform.

The first study examined learning outcomes with respect to exam scores and found statistically significant effects with video minutes viewed. It is the first study to report and evaluate learning outcomes of ClassTranscribe. The potential for future research and education insights using these (and similar future data) is broad and significant, including for example, machine learning models and multivariate statistical models to predict student outcomes and behaviors, and Hidden Markov Models to model event sequences. The results presented were based on a single measure - video minutes viewed. Future work will explore other learner behaviors that are possible with the data presented here (for example, engagement effects due to full-screen viewing, effects of search and transcription use, and comparing video minutes to wall-clock minutes to determine which is a better predictor of learning in technically dense computer science content). The results and experience of this study were also used to identify limitations and prioritize future improvements to the tool (for example, encouraging greater use of search and share functionality and improving accessibility).

The second study extended the first study. We researched and reported in detail the search-related activities of students in engineering courses using ClassTranscribe, an online web application that supports principles of Universal Design for Learning. By analyzing the system logs of student interactions with the tool, we studied the student behaviors during caption-based search. Our findings have both practical implications and implications for UDL for improving video-based learning, specifically by using caption-based search. We identified and fitted a Zipfian power law in search query terms ($k = 32.1$, $s = 0.58$), created a taxonomy and categorization of search queries and examined video and search-related actions prior to- and post- search events that varied between students categorized as Frequent Searchers and Infrequent Searchers. We examined differences in search and video-choice behaviors when students were watching new content versus reviewing previously-viewed content. A detailed understanding of search not only provided insights into student search-based interactions to find video content but also suggested how students are using searchable video-based content to learn in undergraduate engineering courses, and raised new research questions and ideas to improve the pedagogical utility of caption-search, video presentation and video-based learning. These results also demonstrated that students choose different learning pathways (infrequent vs. frequent vs. no searching) which provided empirical support for a UDL approach to course content design and delivery.

The third study presented findings on learnersourced caption editing for online lecture videos. We deployed ClassTranscribe in a large (N=387) text retrieval and mining course where 58 learners participated in editing captions of 89 lecture videos. Each lecture video was edited by two editors sequentially and 18 editors participated in follow-up interviews. From the interview data, we summarized the following results: (i) Students used captions for various reasons, in multiple ways, and evaluated captions to be helpful for learning; (ii) Students adopted various kinds of strategies to fix different types of caption errors; (iii) In addition to the extra credit, students had other motivations for fixing captions; (iv) Students' perceived differences between being the first and second editor; (v) Explicit suggestions from students on improving the caption-editing activity. Our study provided both system design suggestions and practical implications for advancing learnersourced captioning for video-based online education.

REFERENCES

[1] L. McCarron, "Creating accessible videos: Captions and transcripts," *Communications of the Association for Information Systems*, vol. 48, no. 1, p. 19, 2021.

[2] J. S. Cross, N. Keerativoranan, M. K. J. Carlon, Y. H. Tan, Z. Rakhimberdina, and H. Mori, "Improving mooc quality using learning analytics and tools," in *2019 IEEE Learning With MOOCS (LWMOOCS)*. IEEE, 2019, pp. 174–179.

[3] A. S. Clossen, "Beyond the letter of the law: Accessibility, universal design, and human-centered design in video tutorials," *Pennsylvania Libraries: Research & Practice*, vol. 2, no. 1, pp. 27–37, 2014.

[4] M. Wald, "Concurrent collaborative captioning," 2013.

[5] B. Parton, "Video captions for online courses: Do youtube's auto-generated captions meet deaf students' needs?" *Journal of Open, Flexible, and Distance Learning*, vol. 20, no. 1, pp. 8–18, 2016.

[6] R. Ranchal, T. Taber-Doughty, Y. Guo, K. Bain, H. Martin, J. P. Robinson, and B. S. Duerstock, "Using speech recognition for real-time captioning and lecture transcription in the classroom," *IEEE Transactions on Learning Technologies*, vol. 6, no. 4, pp. 299–311, 2013.

[7] "Captioning and transcription for stem content," Jun 2019. [Online]. Available: https://www.3playmedia.com/blog/captioning-and-transcription-for-stem-content/

[8] L. Angrave, K. Jensen, Z. Zhang, C. Mahipal, D. Mussulman, C. Schmitz, R. Baird, H. Liu, M. Wu, and R. Kooper, "Improving student accessibility, equity, course performance, and lab skills: How introduction of classtranscribe is changing engineering education at the university of illinois," in *ASEE annual conference & exposition*, 2020.

[9] L. Angrave, Z. Zhang, G. Henricks, and C. Mahipal, "Who benefits? positive learner outcomes from behavioral analytics of online lecture video viewing using classtranscribe," in *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 2020, pp. 1193–1199.

[10] C. Mahipal, L. Angrave, Y. Xie, B. Chatterjee, H. Wang, and Z. Qian, "“what did i just miss?!” presenting classtranscribe, an automated live-captioning and text-searchable lecture video system, and related pedagogical best-practices," in *ASEE Annual Conference and Exposition, Conference Proceedings*, 2019.

[11] T. Tuna, J. Subhlok, L. Barker, S. Shah, O. Johnson, and C. Hovey, "Indexed captioned searchable videos: A learning companion for stem coursework," *Journal of Science Education and Technology*, vol. 26, no. 1, pp. 82–99, 2017.

[12] M. Moridani, "Asynchronous video streaming vs. synchronous videoconferencing for teaching a pharmacogenetic pharmacotherapy course," *American journal of pharmaceutical education*, vol. 71, no. 1, p. 16, 2007.

[13] S. Nikzad, A. Azari, H. Mahgoli, and N. Akhoundi, "Effect of a procedural video cd and study guide on the practical fixed prosthodontic performance of iranian dental students," *Journal of Dental Education*, vol. 76, no. 3, pp. 354–359, 2012.

[14] S. Ramlogan, V. Raman, and J. Sweet, "A comparison of two forms of teaching instruction: video vs. live lecture for education in clinical periodontology," *European Journal of Dental Education*, vol. 18, no. 1, pp. 31–38, 2014.

[15] L. Breslow, D. E. Pritchard, J. DeBoer, G. S. Stump, A. D. Ho, and D. T. Seaton, "Studying learning in the worldwide classroom research into edx's first mooc." *Research & Practice in Assessment*, vol. 8, pp. 13–25, 2013.

[16] D. T. Seaton, Y. Bergner, I. Chuang, P. Mitros, and D. E. Pritchard, "Who does what in a massive open online course?" *Commun. ACM*, vol. 57, no. 4, pp. 58–65, Apr. 2014. [Online]. Available: http://doi.acm.org/10.1145/2500876

[17] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Engaging with massive online courses," in *Proceedings of the 23rd international conference on World wide web*, ACM. New York, NY, USA: ACM, 2014, pp. 687–698.

[18] R. F. Kizilcec, C. Piech, and E. Schneider, "Deconstructing disengagement: analyzing learner subpopulations in massive open online courses," in *Proceedings of the third international conference on learning analytics and knowledge*, D. Suthers, K. Verbert, E. Duval, and X. Ochoa, Eds., ACM. New York, NY, USA: ACM, 2013, pp. 170–179.

[19] H. K. Evans and V. Cordova, "Lecture videos in online courses: A follow-up," *Journal of Political Science Education*, vol. 11, no. 4, pp. 472–482, 2015.

[20] P. Jones, K. Naugle, and M. Kolloff, "Teacher presence: Using introductory videos in online and hybrid courses," *Learning Solutions Magazine*, pp. 1–10, 2008.

[21] N. Scagnoli and C. Packard, "Improving a blended delivery model in a large enrollment business course," *Proceedings of the Academy of Human Resource Development*, 2011.

[22] N. I. Scagnoli, J. Choo, and J. Tian, "Students' insights on the use of video lectures in online classes," *British Journal of Educational Technology*, vol. 50, no. 1, pp. 399–414, 2019.

[23] J. C. Ren, M. Hasegawa-Johnson, and L. Angrave, "Classtranscribe: a new tool with new educational opportunities for student crowdsourced college lecture transcription," in *ISCA International Workshop on Speech and Language Technology in Education, SLaTE 2015, Leipzig, Germany, September 4-5, 2015*, S. Steidl, A. Batliner, and O. Jokisch, Eds. Baixas, FRANCE: ISCA, 2015, pp. 179–180.

[24] T. Tuna, J. Subhlok, L. Barker, V. Varghese, O. Johnson, and S. Shah, "Development and evaluation of indexed captioned searchable videos for stem coursework," in *Proceedings of the 43rd ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '12. New York, NY, USA: ACM, 2012. [Online]. Available: http://doi.acm.org/10.1145/2157136.2157177 pp. 129–134.

[25] B. N. Shiver and R. J. Wolfe, "Evaluating alternatives for better deaf accessibility to selected web-based multimedia," in *Proceedings of the 17th international ACM SIGACCESS conference on computers & accessibility*, 2015, pp. 231–238.

[26] S. Burgstahler, "Universal design: Implications for computing education," *ACM Transactions on Computing Education (TOCE)*, vol. 11, no. 3, p. 19, 2011.

[27] R. H. Rutherfoord and J. K. Rutherfoord, "Exploring teaching methods for on-line course delivery-using universal instructional design," in *Proceedings of the 9th ACM SIGITE conference on Information technology education*, ACM. New York, NY, USA: ACM, 2008, pp. 45–50.

[28] D. R. Parker, L. E. Robinson, and R. D. Hannafin, ""blending" technology and effective pedagogy in a core course for preservice teachers," *Journal of Computing in Teacher Education*, vol. 24, no. 2, pp. 49–54, 2007.

[29] P. L. Davies, C. L. Schelly, and C. L. Spooner, "Measuring the effectiveness of universal design for learning intervention in postsecondary education." *Journal of Postsecondary Education and Disability*, vol. 26, no. 3, pp. 195–220, 2013.

[30] M. E. King-Sears, T. M. Johnson, S. Berkeley, M. P. Weiss, E. E. Peters-Burton, A. S. Evmenova, A. Menditto, and J. C. Hursh, "An exploratory study of universal design for teaching chemistry to students with and without disabilities," *Learning Disability Quarterly*, vol. 38, no. 2, pp. 84–96, 2015.

[31] T. Dean, A. Lee-Post, and H. Hapke, "Universal design for learning in teaching large lecture classes," *Journal of Marketing Education*, vol. 39, no. 1, pp. 5–16, 2017.

[32] J. H. Flavell, "Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry." *American psychologist*, vol. 34, no. 10, p. 906, 1979.

[33] J. D. Bransford, A. L. Brown, R. R. Cocking et al., *How people learn.* Washington, DC: The National Academies Press, 2000, vol. 11.

[34] J. L. Nietfeld, L. Cao, and J. W. Osborne, "Metacognitive monitoring accuracy and student performance in the postsecondary classroom," *The Journal of Experimental Education*, vol. 74, no. 1, pp. 7–28, 2005. [Online]. Available: http://www.jstor.org/stable/20157410

[35] A. Young and J. D. Fry, "Metacognitive awareness and academic achievement in college students." *Journal of the Scholarship of Teaching and Learning*, vol. 8, no. 2, pp. 1–10, 2008.

[36] J. Parham, L. Gugerty, and D. Stevenson, "Empirical evidence for the existence and uses of metacognition in computer science problem solving," in *Proceedings of the 41st ACM technical symposium on Computer science education*, ACM. New York, NY, USA: ACM, 2010, pp. 416–420.

[37] J. Broadbent and W. Poon, "Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review," *The Internet and Higher Education*, vol. 27, pp. 1–13, 2015.

[38] M. Cardelle-Elawar, "Effects of metacognitive instruction on low achievers in mathematics problems," *Teaching and Teacher Education*, vol. 11, no. 1, pp. 81–95, 1995.

[39] S. W. Vanderstoep, P. R. Pintrich, and A. Fagerlin, "Disciplinary differences in self-regulated learning in college students," *Contemporary educational psychology*, vol. 21, no. 4, pp. 345–362, 1996.

[40] M. Mani and Q. Mazumder, "Incorporating metacognition into learning," in *Proceeding of the 44th ACM technical symposium on Computer science education*. New York, NY, USA: ACM, 2013, pp. 53–58.

[41] G. Schraw, "Promoting general metacognitive awareness," *Instructional science*, vol. 26, no. 1-2, pp. 113–125, 1998.

[42] M. V. Izzo and W. M. Bauer, "Universal design for learning: enhancing achievement and employment of stem students with disabilities," *Universal Access in the Information Society*, vol. 14, no. 1, pp. 17–27, 2015.

[43] R. D. Black, L. A. Weinberg, and M. G. Brodwin, "Universal design for learning and instruction: Perspectives of students with disabilities in higher education," *Exceptionality Education International*, vol. 25, no. 2, 2015.

[44] J. Griful-Freixenet, K. Struyven, M. Verstichele, and C. Andries, "Higher education students with disabilities speaking out: perceived barriers and opportunities of the universal design for learning framework," *Disability & Society*, vol. 32, no. 10, pp. 1627–1649, 2017.

[45] E. Delen, J. Liew, and V. Willson, "Effects of interactivity and instructional scaffolding on learning: Self-regulation in online video-based environments," *Computers & Education*, vol. 78, pp. 312–320, 2014.

[46] A. Defranceschi and M. Ronchetti, "Video-lectures in a traditional mathematics course on itunes u: Students' feedback," in *2011 14th International Conference on Interactive Collaborative Learning*. IEEE, 2011, pp. 219–224.

[47] T. Tuna, T. Dey, J. Subhlok, and L. Leasure, "Video supported flipped classroom," in *Proceedings of EdMedia + Innovate Learning 2017*, J. P. Johnston, Ed. Washington, DC: Association for the Advancement of Computing in Education (AACE), June 2017. [Online]. Available: https://www.learntechlib.org/p/178432 pp. 1159–1171.

[48] J. Kim, P. J. Guo, C. J. Cai, S.-W. Li, K. Z. Gajos, and R. C. Miller, "Data-driven interaction techniques for improving navigation of educational videos," in *Proceedings of the 27th annual ACM symposium on User interface software and technology*, 2014, pp. 563–572.

[49] R. Deshpande, T. Tuna, J. Subhlok, and L. Barker, "A crowdsourcing caption editor for educational videos," in *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*. IEEE, 2014, pp. 1–8.

[50] A. Cross, M. Bayyapunedi, D. Ravindran, E. Cutrell, and W. Thies, "Vidwiki: Enabling the crowd to improve the legibility of online educational videos," in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 2014, pp. 1167–1175.

[51] G. Culbertson, S. Shen, E. Andersen, and M. Jung, "Have your cake and eat it too: Foreign language learning with a crowdsourced video captioning system," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2017, pp. 286–296.

[52] D. Jiang, J. Pei, and H. Li, "Mining search and browse logs for web search: A survey," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 4, no. 4, pp. 1–37, 2013.

[53] H. K. Azad and A. Deepak, "Query expansion techniques for information retrieval: a survey," *Information Processing & Management*, vol. 56, no. 5, pp. 1698–1735, 2019.

[54] P.-L. Lei, C.-T. Sun, S. S. Lin, and T.-K. Huang, "Effect of metacognitive strategies and verbal-imagery cognitive style on biology-based video search and learning performance," *Computers & Education*, vol. 87, pp. 326–339, 2015.

[55] K. Pearson, "X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175, 1900.

[56] J. Han, M. Kamber, and J. Pei, "Data mining concepts and techniques third edition," *The Morgan Kaufmann Series in Data Management Systems*, vol. 5, no. 4, pp. 83–124, 2011.

[57] J. Corbin and A. Strauss, *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications, 2014.

[58] A. Straus and J. Corbin, "Basics of qualitative research: Grounded theory procedures and techniques," 1990.

[59] K. Charmaz, *Constructing grounded theory: A practical guide through qualitative analysis*. sage, 2006.

[60] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[61] G. W. Corder and D. I. Foreman, *Nonparametric statistics: a step-by-step approach*, second edition ed. Hoboken, New Jersey: Wiley, 2014.

[62] G. Zipf, *The Psycho-Biology of Language: An Introduction to Dynamic Philology*, ser. Cognitive psychology]. Routledge, 1999. [Online]. Available: https://books.google.com/books?id=w1Z4Aq-5sWMC

[63] G. K. Zipf, *Human behavior and the principle of least effort: An introduction to human ecology.* Ravenio Books, 2016.

[64] F. Lillo and S. Ruggieri, "Estimating the total volume of queries to google," in *The World Wide Web Conference*, 2019, pp. 1051–1060.

[65] F. J. Massey Jr, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[66] M. Cristelli, M. Batty, and L. Pietronero, "There is more than a power law in zipf," *Scientific reports*, vol. 2, no. 1, pp. 1–7, 2012.

[67] R. L. Thorndike, "Who belongs in the family?" *Psychometrika*, vol. 18, no. 4, pp. 267–276, 1953.

[68] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The annals of mathematical statistics*, pp. 50–60, 1947.

[69] N. Archives and R. Administration, "The soundex indexing system," 2007.

[70] D. J. Sheskin, *Handbook of parametric and nonparametric statistical procedures.* crc Press, 2020.

[71] A. Broder, "A taxonomy of web search," in *ACM Sigir forum*, vol. 36, no. 2. ACM New York, NY, USA, 2002, pp. 3–10.

[72] P. Sondhi, M. Sharma, P. Kolari, and C. Zhai, "A taxonomy of queries for e-commerce search," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018, pp. 1245–1248.

[73] M. A. Maabreh, M. N. Al-Kabi, and I. M. Alsmadi, "Query classification and study of university students' search trends," *Program*, 2012.

[74] J. Kim, "Organic crowdsourcing systems." in *AAAI Spring Symposia*, 2016.

[75] J. Kim et al., "Learnersourcing: improving learning with collective learner activity," Ph.D. dissertation, Massachusetts Institute of Technology, 2015.

[76] L. Azzopardi, P. Thomas, and N. Craswell, "Measuring the utility of search engine result pages: an information foraging based measure," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018, pp. 605–614.

[77] I. B. Vidinli and R. Ozcan, "New query suggestion framework and algorithms: A case study for an educational search engine," *Information processing & management*, vol. 52, no. 5, pp. 733–752, 2016.

[78] J. Singh and A. Anand, "Model agnostic interpretability of rankers via intent modelling," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, pp. 618–628.

[79] P. K. Ralabate, "Universal design for learning: Meeting the needs of all students," *The ASHA Leader*, vol. 16, no. 10, pp. 14–17, 2011.