Reducing Attrition in Distance Education Through Positive Self-Regulation

Cory A. Todd
Georgia Institute of Technology
corytodd@gatech.edu

Abstract

Digital classrooms are capable of providing a scalable and affordable model for the delivery and management of learning material. The constraints of these models fall into two categories. The first being technology, a field which is rapidly developing and evolving to provide massively scaled applications across the world. These are problems with absolute and tangible solutions that can be quantified and compared. The second category includes the very human factors that contribute to student achievement. Even with the best content delivery system, if students are not staying enrolled and completing their coursework through graduation, the distance learning platform can no longer grow. We will be discussing the causes of and potential solutions to student attrition in distance learning. Finally, we will be discussing the development of a tool that promotes positive self-regulated learning in this environment. This will be offered as a potential solution to the issue of student attrition in distance learning programs.

Categories and Subject Descriptors Education Technology [*Attrition in Distance Education*]:

Categories and Subject Descriptors Education Technology

Keywords Online education attrition, MOOC scalability

1. Introduction

Educational technology is delivering tremendous amounts of information to students in communities around the world. Nearly half of those surveyed in recent opinion polls agreed that the technology used in distance education is providing new opportunities for both traditional and under-served student populations(Glenn and D'Agostino, 2008). Key features of this technology include access-anywhere content, diverse resources, and the availability of near real-time communication tools. All of these features provide a foundation upon which students can grow in their respective fields. As the scale of these applications grows, so too does the importance of addressing barriers to student success. Of all the issues, one of the most important to address is student attrition because it adversely affects students and institutions alike. We will discuss causes of attrition as identified by other researchers followed by their suggested solutions. Finally, we will describe the solution we have designed based on their research and what we hope to accomplish.

Distance learning has become synonymous with low-cost, affordable, and accessible education (Street, 2010, p. 2). Students who would otherwise not have access to higher education due to socio-economic or geographic constraints can now pursue advanced degrees. As the numbers of online enrollees swell across colleges and universities both in the United States and abroad, so too does the problem of attrition (Parker, 1999). The phenomenon of student dropout is not unique to distance learning but researchers including Moody et al. have collected compelling evidence that indicates online students will, for a variety of reasons, fail to complete courses at a higher rate than in-person courses. This problem is particularly challenging because the requirements are often incomplete or changing. This is to say that non-persistent students give a number of reasons for dropping but these reasons on their own do not completely describe their individual state, and by extension, cannot be defined as a single common problem.

1.1. The Importance of Success

Technology and learning are iterative processes that compound into increasingly potent resources. In the context of online education, the delivery of content and interaction at scale is a technology problem that we continue to refine. It enables discussion and collaboration across geographic and national boundaries. It allows for such low barriers to entry that nearly any student with a will to attend can do so. The success of these students is contingent upon reliable technology and more importantly, a system that enables them to succeed.

In the context of student learning, knowledge is built on the continued success of students. When a student becomes non-persistent, that is a student that does not complete their program of study, both the student and the institution suffer. The student does not grow the skills that they set out to acquire which may limit their long-term social and economic prospects. The institution has lost time, resources, and most importantly the potential knowledge that this student could have contributed back into the program. For this reason, it is imperative that the attrition rates for online education be addressed.

The success of massively scaled classrooms is contingent on students completing the program. When the rate of student achievement does not scale at the same rate as the online classroom, this scalability becomes a liability. This underscores the very real problems identified by the research of Sorenson et. al. Their research identifies a student's perceived lack of support as a leading contributor to attrition(Sorensen and Donovan, 2017, p.218). Therefore, solving this problem is directly related to student success.

1.2. Real-world Attrition Rates

Distance learning has long-suffered from rates of attrition much higher than traditional, in-person courses. This per-

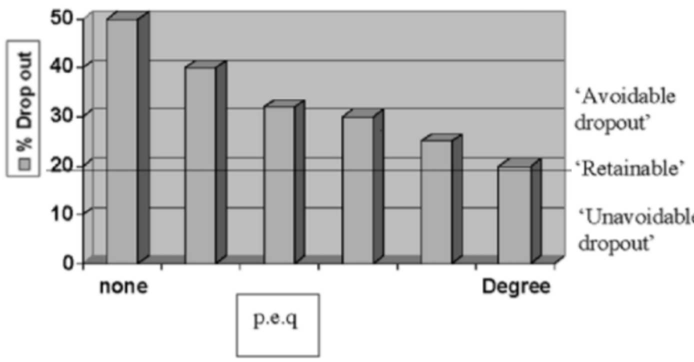

Figure 1. (Retainable and Unavoidable Dropouts [Simpson, 2004])

sistence of this issue, however, is not due to a lack of understanding or effort on the behalf of affected institutions. Research spanning more than a decade has identified the most common reasons for student non-persistence (Chyung, 2001, p. 39). Data has been collected globally from a diverse set of educational institutions ranging from community college to prestigious universities in the United States and abroad. The consensus is that online students require more individualized contact than is typically provided in an online environment. The lack of this contact can manifest itself as a variety of the symptoms but the final condition is often terminal. When a student drops a course or fails to graduate, the net effect is negative for both the student and the institution. Let's improve this outlook with an affordable, scalable, and feature-complete solution to student attrition in distance learning.

Attrition is not a problem that is unique to any particular field of study. It is ubiquitous in higher education and remains a highly troublesome issue for educational institutions (Morgan and Tam, 1999, p. 98). However, distance education appears to suffer higher rates of attrition compare to traditional in-person programs(Street, 2010, p. 4). The rate of attrition in online education tends to vary by field of study, but the general trend is that distance learning students are more likely to not complete a course of study when compared to their in-person counterparts(Wladis et al., 2017, p. 189).

This phenomenon presents a challenge because the reasons a student give for existing program may omit key certain details. In other words, a student who elects to not complete their coursework may cite a variety of reasons but it is an intensely personal decision into which the institution has little visibility. Research by Simpson et. al has shown that with more advanced degrees comes a reduced percentage of avoidable dropout rates. However, their data still shows that there is a small percentage of students that can be retained. It is in this population that we will be focusing our efforts. There is compelling evidence that suggests the program subject and level of coursework have a significant correlation with rates of attrition. Wladis infers that STEM classes tend to have lower rates of attrition because the students are more invested in the coursework than the final result (Wladis et al., 2017, p. 189). This is in contrast to programs focusing on business, philosophy, or psychology where the end result is the motivation. What this means is that students in different programs will require different forms of interaction. There is no one-size-fits-all approach that is going to work across majors and course levels. This adds significant cost and complexity to any proposed solution to the phenomenon of attrition. Moody asserts that the cost of developing distance learning is non-trivial so it critical that each institution conducts their own research to identify their unique attrition characteristics (Moody, 2004, 5).

1.3. Scope

There is a large amount of existing research into the phenomenon of student attrition. These studies have identified four distinct categories of reasons that students may cite for ceasing their educational pursuits (Morgan and Tam, 1999, p. 99).

Independent researchers have come to a consensus that personal or family reasons are the single most common cause for non-persistence in distance learning (Beer and Celeste, 2017; Street, 2010; Parker, 1999; Angelino et al., 2007; Morgan and Tam, 1999). Current research on the phenomenon of attrition in distance learning is split into two primary categories. The first is conducted through direct student surveys that attempt to discover what factors lead to the student's decision to drop a course or the student's opinion on why they received a non-passing grade. In the cases of dropping a course, either through passive or active withdrawal, there are four primary categories for student-defined barriers (Morgan and Tam, 1999, p. 99).

Situational
: The student has extenuating circumstances outside the control of the institution. This is the most commonly reported factor. Addressing factors in this category fall outside the traditional responsibilities of an educational institution.

Institutional
: The student experienced insurmountable challenges where the institution was at fault. This includes inadequate feedback and poorly formed course requirements. This category requires a productive feedback loop in which students have a say in how courses are improved.

Dispositional
: The student possesses underdeveloped study and learning skills. The subtle difference between dispositional and situational factors is that the former can be remedied by the institution through educational intervention.

Epistemological
: The student's perception of the coursework is such that that content is too difficult, too unrelated to their goal, or too theoretical.

There is additional research that exposes additional, subtle factors that are difficult to quantify. Sorensen notes that students may be apprehensive about giving the true reason for their departure (Sorensen and Donovan, 2017, p .9). Students may not want to admit that a course was too difficult for them and instead, cite a situational factor to preserve their self-esteem. There are also cases where students withdrew to avoid disciplinary action due to academic dishonesty. In these cases, the student is motivated to preserve their reputation and thus, they will cite an alternate barrier to their persistence (Sorensen and Donovan, 2017, p. 11).

We will be focused exclusively on addressing dispositional barriers that a student may experience. While potentially solvable with software, we did not feel that situational,

Positive Traits	Negative Traits
Intrinsic motivation	Avoids challenges
Self-efficacy	Unclear goals
Low test anxiety	Disengaged during lectures
Perseverance	Unclear goals
Positive outlook	Negative outlook

Figure 2. Self-Regulated Learning Characteristics

institutional, or epistemological issues could be as effectively addressed as those in the dispositional category. For this reason, all other categories will be outside the scope of this discussion.

2. Prior Work

To date, researchers have yet to discover a solution to the larger issue of attrition in distance education. Fortunately, a number of existing studies have shown that there are solutions for some of the individual causes. Furthermore, some of these solutions have been successfully implemented by a number of institutions (Moody, 2004, 5). Some of the most effective solutions are relatively simple in concept and focus on clear communication and meaningful student engagement. With these two approaches, we can effectively identify students that are at risk of attrition (Moody, 2004, p. 207).

One of the most common methods that separate researchers have found effective is having a direct path of communication between students and instructors (Street, 2010, p. 4). Given the massively scaled and remote nature of distance education and the fact that Professors and teaching assistants are not scaled at the same rate as those enrolled in the course, it understandable that this approach on its own is not practical.

Building from this idea is the concept of the learning community (Angelino et al., 2007, p. 7). Platforms such as Piazza strive to build this community by providing a platform for asynchronous discussion. This helps ease the communication burden as classes grow to enormous sizes by enabling peer-to-peer and peer-to-instructor interaction. While effective, we feel that much more can be done to promote positive learning techniques through these platforms.

2.1. Success Indicators

There is evidence that students who possess positive self-regulated learning characteristics tend to be more successful students. Data collected by Pardo et. al(Pardo et al., 2017) shows that being engaged and positive about coursework not only correlates with higher grades but also reduces the probability of a student not completing a course.

Given these traits, it stands to reason that it must be possible to identify students who are showing negative learning patterns. The online learning system is already collecting the actions, and inactions, of students so analysis of this data is a natural progression. These students can be proactively targeted for additional assistance in developing positive learning habits. Additionally, this information can be made available to the student in order to maintain transparency and empower them to make their own improvements.

2.2. Proactive Engagement

The first step in developing positive self-regulated learning habits is to make the student aware of their actions. There is evidence that engaging with these students can have positive outcomes (Michinov et al., 2011, p.249). The problem, however, still comes down to the scale at which these large classes are operating. The best way to approach problems at scale is through automation. With the presence of these digital classrooms already accessible through a web page, automating this analysis is a natural next step.

We can also leverage this scale to our advantage by using the massive amount of student content to train natural language processing algorithms. Student posts are interesting because they are typically well structured and rich with technical information. These posts may contain a high degree of technical information and will in general tightly follow the discussion topic. This provides a powerful dataset for analyzing texts from a number of STEM fields that are otherwise under-represented in the standard NLP corpus. This dataset could be used to drive advancements in computer-assisted learning of advanced sciences.

2.3. Solutions

Researchers have come to the conclusion that there is no universal solution to student attrition in distance learning. However, there are a number of proposals with significant chances of success. These solutions address institutional and epistemological factors reported by students. The common element to all solutions is active communication between all parties (Simpson, 2004, p. 93). Students need to articulate their goals and guidance counselors need to clearly state what is expected of the student. Instructors must provide clear, timely, and concise feedback on student performance. There is also strong evidence that dispositional factors can be eliminated through proper training (Simpson, 2004, p. 94). Teaching students how to learn, how to study, and how to take tests will reduce the occurrence of negative self-regulated learning behaviors, a known contributor to low achievement and non-persistence.

The learning community, as discussed earlier, has been repeatedly identified by researchers as a powerful tool in student retention (Angelino et al., 2007, p. 7). This is a place where students are actively engaged with one another and feel a sense of belonging and achievement, consistently throughout their coursework. Angelino's research suggests taking more personalized action for the student at every level. The goal is to help students feel more connected with their classmates and less invisible. Having this type of support is shown to reduce the likelihood of non-persistence.

2.4. Effectiveness

Identifying at-risk students is only the first part of the solution. The second and most critical part is the execution of tasks that eliminate barriers to distance learning. Research conducted by Simpson shows a positive outlook for solutions involving proactive student engagement (Simpson, 2004, p. 84). In this research, instructors contacted students with a pacing reminder before assignments were due. This simple act produced a significant increase in the number of submitted assignments. The submission of assignments, in this case, was found to correlate with positive overall performance. However, this research also tempers expectations with the assertion that there is a practical and fundamental level of attrition that is unavoidable, regardless of how learning is imparted as shown in Figure 1 by Simpson (Simpson, 2004, p. 81). This indicates that there is a natural minimum level of attrition in education that will not be affected by these solutions.

3. Development

3.1. Existing Solutions

Current learning platforms offer very topical information about how a student's engagement compares with the rest of the class. On Piazza, for example, the following metrics are visible to a student.

- Top 5 Student Contributors
- Post stats (total posts, total contributors, response by user type, and response time)
- Daily charts plotting unique users and posts
- Personal stats about contributions.

One obvious deficiency is a process that ranks the quality of contributions. The highest contributor could simply be posting "Me Too!" responses in an effort to keep up with participation. In fact, the quarter-term participation post by Dr. Joyner made two points very clear:

1. Student's will, invariably, attempt to catch up with participation by generating a deluge of low-quality posts. This is counterproductive.
2. Quality checks are determined manually by the teaching staff

This is not an ideal solution because there are much more productive tasks that the teaching staff can be performing.

The browser plugin space is saturated with time and productivity monitoring solutions targeting both students and professionals. This comes in the form of time trackers, focus assistants that discourage you from wandering through Reddit for too long, and some form of note-taking integration. These provide useful services but none provide the level of integration and analysis that is required to provide distance-learning focused enhancements. For these reasons, this project will make a meaningful contribution to the education technology community.

3.2. Concept

A web browser is a powerful tool for accessing and interacting with the world. Distance learning heavily depends on modern web browsers to provide responsive and immersive content to student consumers. This means that our tool will have the highest yield if it is designed to operate as a browser plugin. Cross-platform and data connectivity issues are then offloaded to the browser so the tool can focus on its core requirements. Absolute student privacy is a non-negotiable feature of this tool. Any and all means available that can anonymize and protect student data will be taken. The first step is to only take that data which we need and nothing more. The second step is to commit to mindful security practices without taking compromise.

3.3. Design

The solution architecture is that of the classic client-server pattern. The client is a browser plugin and the server provides endpoints that communicate using JSON. The browser plugin is designed to be as passive as possible, requiring no manual input from the user. Most of the difficult and interesting work is done server-side. Additionally, no passwords or personally identifiable information are required for operation.

For this project, Flask and Peewee ORM were selected in order to produce a reliable, high-quality product. Flask is solid, secure, and very easy to understand. Peewee supports a number of different database providers which will make future maintenance and expansion of this project a lot easier. The initial plan was to use Google App Engine for hosting but I opted for my own servers to save on cost. This means that there are no proprietary vendor configurations for the next maintainer to worry about. The complete server setup has been documented in the project repository.

3.3.1. Server

The server component is a RESTful JSON service that provides structured responses to authenticated clients. A request to the service will return the engagement statistics for the authenticated user. These statistics are periodically pulled from Piazza and processed into the structure response. Data is associated with users by their Piazza unique id and no actual text from the discussions are stored on the server. Natural language processing is used to measure a variety of metrics on the content and then interesting statistics are calculated from the results. A student will have a single statistical record for each course in which they are enrolled because a Piazza id is capable of identifying a user uniquely across their entire system.

The data acquisition processes use a depth-first scraping algorithm to trace a Piazza post from the topic down through each of the response trees. This is very useful for tracking the topic through the completion of the discussion. At a high level, the entire process is described as follows:

1. Authenticate with Piazza
2. Request list of top-level posts
3. For each top-level post, recursively iterate all response posts
4. For each response post, calculate and store statistics

3.3.2. NLP

NLP has effectively applied machine learning with the purpose of understanding meaning in a text. This means that we need a lot of data and a strong understanding of the problem being addressed(NLTK, 10). As an example, a common problem in NLP is sentiment analysis. This is used to determine the tone and human emotion from the written text; two components that are either exemplary or qualitative but rarely quantifiable. Vast quantities of data are analyzed to discover patterns in the phrase that are associated with happiness, anger, love, distress, or any number of sentiments that humans are capable of expressing. This data is used to train a processor task in preparation for real-world input. Naturally, there is a tradeoff between speed, performance, and time-to-develop. Training is complex and time-consuming so many libraries and NLP services provide pre-built models that are generalized to particular tasks. Real-time processing as found in chatbots requires immense computational resources to provide responses within a time period tolerated by humans. For this reason, chatbots are a commonly developed using dedicated services that handle all the heavy lifting.

Applied to Education Researchers continue to prove that individualized instruction is a powerful tool that enhances the learning experience of students (Bahçeci and Gürol, 2016, p.8). However, this individualization is costly in terms of time and human capital. A solution to this scaling issue is the automation of individualizing student instruction through intelligently curated guidance and feedback. As more curriculum moves to an online format, the potential for computer-aided learning increases because intelligent tools can automatically and instantly provide student feedback.

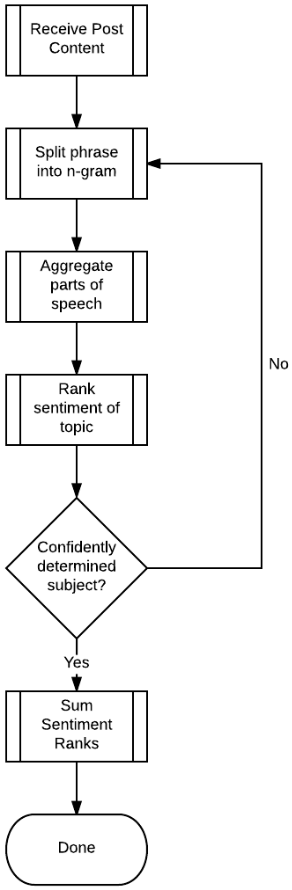

Figure 3. Sentiment Analysis

An example of this is tracking a student's writing progress throughout their K-12 attendance on a regular basis. Each state has specific requirements about what students should be able to do by certain grade levels which are enforced through high-stakes testing. By tracking writing on regular basis, perhaps standardized testing could be dissolved. NLP tools, combined with fully-online curriculum delivery and submission can provide transparent standards testing and less opaque criteria for student achievement. This puts more information in the hands of all the stakeholders in the education process.

As with anything, there are legalities to address and the likelihood that such a system will be abused. Care must be taken to protect student privacy and the integrity of the system as a whole.

Problems In the NLP space, there are a number of open problems that continue to attract innovative solutions. One such issue is that of ambiguous language. A grammar provides structure and order to phrases by providing concrete rules for categorizing the components of the text. However, these rules provide no guidance for disambiguating the syntax or semantic meaning that is extracted.

There are no general solutions to this problem but rather a variety of techniques that are specific to a domain. Some solutions attempt to address this by looking at the large context of the statement. This is not a viable option for

```
# val range is -1.0 to 1.0
# negative to positive, respectively
val = int(((val + 1) / 2) * 128)

# HSL with full saturation and half luminance
return "hsl({}, {}%, {}%)".format(val, sat, lum)
```

Figure 4. HSL Calculation

single-shot events like voice commands to your Alex but is commonly used for chatbot application. Other issues include the detection of sarcasm, the nearly infinite number of ways in which the same fact can be stated, and scalability issues across different languages.

NLP is a complex tool with the massive potential to improve education by amplifying the instructional capacity of an institution. Fortunately, most of the problems that can immediately address in education do not require solving the open NLP problems. This means that new solutions are limited only by creativity and bureaucracy, not existential limitations.

3.3.3. Chrome Plugin

The user component is a browser plugin designed to integrate with Google Chrome. Since it requires the context of Piazza in order to operate, the plugin was created using the page action patter (google, 6). This produces a plugin that is only active on pages the developer wants to interact with. This is in accordance with Google's guidelines that dictate a plugin should be additive to the user experience (google, 6).

A neat feature of the plugin is the rendering of a student's sentiment in their posts over time. This done by measuring and averaging their post polarity for each day of the course which is handled server-side. When the statistics are generated, the days are arranged sequentially to create an ordered timespan. The sentiment polarity is then mapped onto the HSL color system. In this system, we use 100% saturation and 50% for a solid color. The hue component is mapped to the polarity where red lands naturally on 0 and green lands naturally at 128. This is much more convenient than the two-variable algorithm that would be required in RGB space. As shown in Figure 4, the HSL value is encoded into a string and ejected into the user interface as a CSS style.

Authentication is a simple key-value pair that is handled automatically through the plugin through a process that is similar to JWT but without any signatures. The user key is captured from a cookie stored in the user's Piazza web session which is why we operate only in the context of Piazza. This key conveniently identifies the user uniquely across all Piazza courses. The value portion is generated by the server in response to an enrollment request. The Chrome plugin saves this key-value pair to use for future requests. The storage sync API is used to persist the key across all the user's browser so it will always be available.

4. Future Development

There are a number of features that can be added to this project to increase its effectiveness. Some of these features were omitted due to time and scope constraints, others due to difficulties in working with the Piazza platform. An unofficial Piazza API was used for the purpose of collecting the raw data so it would be a worthwhile effort to engage with the Piazza team for an official version.

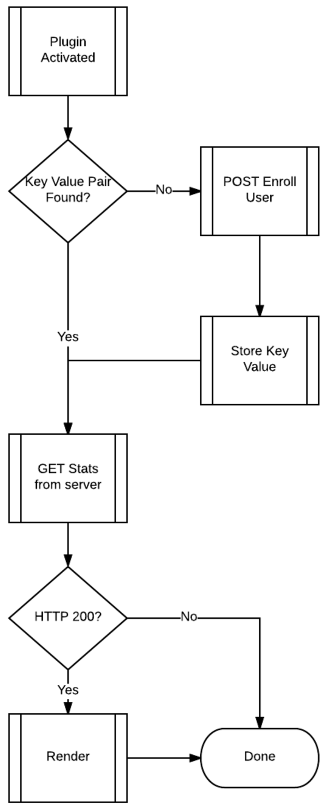

Figure 5. User Authentication

The most pertinent improvement would be to refactor how data acquisition is performed, specifically the Piazza authentication portion. Since this requires the use of my actual Piazza credentials, I perform the data processing offline and upload the database to the server on completion. This circumvents the risk of having my credentials stolen but does limit the courses that can be processed to only those in which I have been enrolled. It should also be noted that the method used to gather the data is pretty intensive so I have written random delays into the scrape algorithm so as to not cause harm to Piazza.

Another improvement would be to port the Chrome plugin to Firefox in order to reach a wider audience. Considerations were made to this end and minimal Chrome-specific APIs were used. Due to the sandboxed nature of plugins, there is a certain amount of bootstrap code required to satisfy Chrome. This is also why all of the libraries are bundled directly into the plugin so that cross-site-scripting policies enforced by Chrome can be honored.

A final idea for improvement would be the implementation of a system that proactively engages with students as they work through their coursework. Using the calculated data, autonomous bots could engage with the student to remind them to post more frequently or suggest posting topics for their courses.

5. Contribution of This Work

The primary contribution of this project is to increase student success by detecting and intervening in a student's coursework for the purpose of preventing course withdrawal. This was achieved through careful analysis of prior works and the development of a tool capable of collecting and identifying actionable events that may predict student failure. Distance learning has exponentially improved my life so I feel that any effort that increases the odds of another student becoming successful is a meaningful contribution.

6. Conclusion

Distance learning has the potential to meet, if not exceed the attrition goals of in-person instruction. Technology is inherent in online learning systems which are, by design, potent tools for the effective dissemination of learning. With a lot of data, a little time, and a commitment to breaking down student barriers, distance learning programs can take the lead as the most effective and affordable form of education. Distance education has given many students the opportunity to learn and grow but if we do not reduce the rate of attrition, it is possible that these programs will cease to exist. By taking more concrete steps to promote student engagement, attrition can be reduced and student success can be elevated. Teaching students how to practice positive self-regulated learning is an effective approach to scaling massively online courses.

Acknowledgments

I'd like to thank Dr. David Joyner for the opportunity to work on this project. With his guidance, I was able to see the value and importance of contributing to our online community. I'd like to thank my project mentor Ken Brooks for his invaluable insight into the world of higher education. Our discussions helped me strengthen my approach to this problem. I would also like to thank each of the peer reviewers from this course who helped guide this project through completion.

References

Angelino, L. M., Williams, F. K., and Natvig, D. (2007). Strategies to engage online students and reduce attrition rates. *Journal of Educators Online*, 4(2):n2.

Bahçeci, F. and Gürol, M. (2016). The effect of individualized instruction system on the academic achievement scores of students. *Education Research International*, 2016.

Beer, C. and Celeste, L. (2017). The problem of student attrition in higher education: An alternative perspective. *Journal of Further and Higher Education*, 41(6):773–784.

Chyung, S. Y. (2001). Systematic and systemic approaches to reducing attrition rates in online higher education. *American Journal of Distance Education*, 15(3):36–49.

Glenn, M. and D'Agostino, D. (2008). The future of higher education: How technology will shape learning. *New Media Consortium.*

google. chrome.pageaction.

Michinov, N., Brunot, S., Le Bohec, O., Juhel, J., and Delaval, M. (2011). Procrastination, participation, and performance in online learning environments. *Computers & Education*, 56(1):243–252.

Moody, J. (2004). Distance education: Why are the attrition rates so high? *Quarterly Review of Distance Education*, 5(3):205.

Morgan, C. K. and Tam, M. (1999). Unravelling the complexities of distance education student attrition. *Distance education*, 20(1):96–108.

NLTK. Natural language toolkit.

Pardo, A., Han, F., and Ellis, R. A. (2017). Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. *IEEE Transactions on Learning Technologies*, 10(1):82–92.

Parker, A. (1999). A study of variables that predict dropout from distance education. *International journal of educational technology*, 1(2):1–10.

Simpson, O. (2004). The impact on retention of interventions to support distance learning students. *Open Learning: The Journal of Open, Distance and e-Learning*, 19(1):79–95.

Sorensen, C. and Donovan, J. (2017). An examination of factors that impact the retention of online students at a for-profit university. *Online Learning*, 21(3).

Street, H. D. (2010). Factors influencing a learner's decision to drop-out or persist in higher education distance learning. *Online Journal of Distance Learning Administration*, 13(4).

Wladis, C., Conway, K., and Hachey, A. C. (2017). Using course-level factors as predictors of online course outcomes: a multi-level analysis at a us urban community college. *Studies in Higher Education*, 42(1):184–200.