# EVALUATING THE POTENTIAL OF TOPIC MODELING AS A SCREENING TOOL FOR ANALYZING STUDENTS' CONSTRUCTED RESPONSE ESSAYS FOR POTENTIAL INDICATORS OF NEGATIVE SENTIMENT OR RISKY BEHAVIOR

by

ELAINE VU DUONG

(Under the Direction of Allan Cohen)

## ABSTRACT

The increased use of computer technology in education has led to a corresponding increase in computerized testing in schools (Dooey, 2008). The COVID-19 pandemic has sped up the transition to online testing (Adedoyin & Soykan, 2020). As a result of the pandemic, students are displaying more signs of anxiety and depression, due in part to the lack of social interaction, a potentially strained home environment, feeling on their own to learn, and losing track of assignments (Cao, et al., 2020; Noble, Hurley, & Macklin, 2020). Teachers may not be able to see their students due to social distancing and students may be reluctant to reveal issues through a virtual format (Son et al., 2020). Identifying indicators of mental disorders can be difficult as indicators manifest in different ways (Muris et al., 2001). Identifying these signs early can have a significant impact on a child's life (McLoone, Hudson & Rapee, 2006). A method for detecting potential indicators in students' writing may help educators identify early signs of risky behavior. There has not yet been reported a method of computerized analysis for detecting indicators of risky behavior or negative sentiment through students' writing.

In this study, I investigate using topic modeling (Blei, 2012) as a method of detecting potential indictors of negative sentiments. The objective of this study is not to diagnose students of disorders through their essays. Rather, it is to explore the development of an internal tool that could be used to assist educators in flagging potentially at-risk students. Topic modeling was used to analyze students' written responses to standardized test questions for detection of negative sentiment as potential indicators of future risky behavior. Results indicate topic models can identify negative sentiment and risky behavior. Additional validation measures such as sentiment analysis and human raters were used to confirm topic modeling results.

INDEX WORDS: Topic modeling, Sentiment analysis, Text analysis, Negative sentiment

# EVALUATING THE POTENTIAL OF TOPIC MODELING AS A SCREENING TOOL FOR ANALYZING STUDENTS' CONSTRUCTED RESPONSE ESSAYS FOR POTENTIAL INDICATORS OF NEGATIVE SENTIMENT OR RISKY BEHAVIOR

by

ELAINE VU DUONG

B.A., The University of Georgia, 2017

M.Ed., The University of Georgia, 2018

M.S., The University of Georgia, 2021

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

# EVALUATING THE POTENTIAL OF TOPIC MODELING AS A SCREENING TOOL FOR ANALYZING STUDENTS' CONSTRUCTED RESPONSE ESSAYS FOR POTENTIAL INDICATORS OF NEGATIVE SENTIMENT OR RISKY BEHAVIOR

by

ELAINE VU DUONG

| | |
|---|---|
| Major Professor: | Allan Cohen |
| Committee: | Michele Lease |
| | Laura Lu |
| | Paula Mellom |

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2023

# DEDICATION

Dedicated to my mom, Kim Phuong Vu, whose sacrifice, unwavering support, and endless love, made this moment possible. I love you always.

# ACKNOWLEDGEMENTS

First, I want to thank the amazing educators and my cohort from the QM program. I was so lucky to have received my education from professors that care deeply and to be surrounded by a cohort that lifted me up. Specifically, I want to thank my advisor, Allan Cohen, who believed in me from the very beginning. It was the privilege of a lifetime to be your student. The world would be a better place if all mentors cared as deeply as you.

Secondly, I want to thank my CLASE family for being my haven. Specially, thank you, Jodi Weber and Paula Mellom, the best bosses ever, who loved, encouraged, and advocated for me. I am indebted to you both for the priceless gift of a sponsored education. I am forever changed by all the opportunities that you've given me and all that you've taught me.

Thirdly, I want to thank my friends and family. Specifically, my grandma and aunts who raised me, Mr. Johnny, my second parent, and my best friends, Kim, Krisna, Gracie, Anna, and Kate, who listened to and loved me through all the highs and lows.

Next, I want to thank the love of my life, Sam Trent, for being my biggest cheerleader and confidante. You believed me in when I didn't believe in myself. You, more than anyone, saw what this journey took. You always held space for me, lifted me up, made me laugh and loved me. I couldn't have done this without you and I can't wait to spend the rest of my life with you.

Finally, I want to thank my mom. All that I am, I am because of you. You put the world at my feet and made me believe that I was capable of anything. You gave so much for my dreams to become a reality. This achievement belongs to us. There are not enough words to thank you for everything that you've done. I'm so lucky to call you, my mom.

TABLE OF CONTENTS

Page

# LIST OF TABLES

Page

# LIST OF FIGURES

Page

# CHAPTER 1

## INTRODUCTION

The increased use of computer technology in education has led to a corresponding increase in the shift towards some form of online or computerized testing for high- and low-stakes assessments in schools (Dooey, 2008). The COVID-19 pandemic has only sped up this transition to increasing use of online testing (Adedoyin & Soykan, 2020). The combination of the shift towards online education and assessment during a pandemic has led to a unique circumstance that was neither possible nor anticipated before. This circumstance is the unprecedented social isolation of students many of whom have been able to document their experience through technology (Ellis, Dumas & Forbes, 2020).

There is a concurrent shift in use of educational technology towards the increased use of online scoring of students' answers. This, in turn, has increasingly led to the use of computerized scoring of assessments (e.g., Shermis & Burstein, 2013). This shift has resulted in both positive and negative outcomes. With respect to positive outcomes is the capability of easily grading and analyzing students' essays in a way that wasn't possible before assessments in machine readable form. However, this also resulted in the decreased likelihood of students' answers being read by a human reader. A potentially problematic result of this lack of human grading is that possible indicators of negative sentiments may be less likely to be detected unless computerized means are developed to detect them (Hara & Kling, 2003). The nature of online learning, while useful for delivering instruction over distance to students, also creates another potential barrier to getting help for potentially at-risk students.

Research has suggested that, as a result of the pandemic, students are displaying more signs of anxiety and depression, due in part to the lack of social interaction, not being able to physically be in school, feeling on their own to learn, and losing track of assignments (Cao, et al., 2020). Another possible contributor to students' anxiety and depression is a strained home environment as a result of one or both parents having to work from home or suffering some amount of financial stress (Noble, Hurley, & Macklin, 2020). The possible resulting negative sentiments and emotions due to the environment, such as anxiety or depression, can manifest themselves in different ways. Teachers may not be able to see their students due to social distancing and students may be reluctant to reveal issues to teacher through a virtual format where their parents or others may overhear (Son, Hegde, Smith, Wang, & Sasangohar, 2020). In this study, I investigate developing and testing of an alternative method of analysis, called topic modeling (Blei, 2012), in order to detect potential indictors of negative sentiments. One way of detecting these negative sentiments can be through the analysis of the words students use in their answers to online tests. Research has shown that sentiment analysis of these answers is possible through text analysis (Mouthami, Devi, & Bhaskaran, 2013). In this study, I use topic modeling to analyze the text of students' written answers to online tests to detect possible indicators of negative sentiments. Specifically, students' written answers to standardized formative assessments will be analyzed using topic models with an eye to detecting negative sentiments. The results of the study will help inform whether this kind of analysis can aid educators in identifying at-risk students early on and getting students the help that they need as soon as possible.

The objective of this study is not to diagnose students of disorders through their essays. Rather, it is to explore the development of an internal tool, specifically, one that could be

deployed at the same time as automated scoring of examinations, that could be used to assist educators in flagging potentially at-risk students. This tool is designed to provide another safeguarding method to screen students for potentially at-risk behaviors.

Statement of the Problem

Identifying potential indicators of risky behavior or negative sentiment in children of all ages can be difficult as signs of abuse, anxiety, depression, etc., can manifest themselves in different ways or may even not be displayed outwardly at all (Muris, Meesters, van Melick, & Zwambag, 2001). Signs are often used to indicate symptoms, which a person can have even if not observable or conveyed to others (Nock, 2008). Identifying these signs early can have a significant impact on a child's life. Studies have shown that the longer a child suffers from abuse, depression, etc., the harder it is to recover (Walker, Chang, Vera-Hernandez, & Grantham-McGregor, 2011). For many such children, it is often difficult to articulate what they are feeling to their teachers and their peers. Detecting this behavior early on is important as risky behavior due to anxiety, depression, etc. can develop rapidly and can become harder to address if left untreated (McLoone, Hudson & Rapee, 2006).

Research suggests students are able to and have expressed negative thoughts and sentiments of themselves and their surroundings in their course-related writing (Bonifacci, Candria, & Contento, 2008; Frayne & Wade, 2006). Thus, a method for detecting these potential indicators in students' writing may help educators and parents identify early signs of possibly risky behavior. Studies have shown that text analysis can detect the emotions of the writer (Bae & Lee, 2012; Mohammed & Yang, 2013; Rude, Gortner & Pennebaker, 2010). However, there has not yet been reported a method of computerized analysis for detecting indicators of risky behavior or negative sentiment through students' writing, even though writing is one of the most

common activities for students. The utility of topic modeling for this purpose is that it can be deployed simultaneously as a regular part of the test scoring process.

Based on research described below, and on the need to address the problem, we use topic modeling to analyze students' written responses to test questions for presence of detection of negative sentiment as potential indicators of future risky behavior. Heretofore, this kind of detection has been done as a result of hand scoring during the process of scoring of students' answers to test questions. The conjecture in this study is that the methods proposed can implemented to provide for rapid detection of negative sentiments indicative of potentially risky behavior. Further, this approach can be used across a variety of different kinds of student writing. In this way, one of the results of this study will be the development and testing of a methodology that can be deployed to detect potentially at-risk students in order to get them help quickly.

# CHAPTER 2

# LITERATURE REVIEW

## Transition to online learning and assessment

The use of online learning is growing as technology continues to become more and more prominent in everyday life. Even when students are learning in person in school, online learning supplements are utilized (Drain, Grier, & Sun, 2012). Studies have shown, even when online learning is successful at improving students' academic performance, students still desire the social aspect of in-person learning (Drain et al., 2012; Tucker, 2012). Facilitating socialization through online learning is difficult and teachers may not have the resources or ability to do so. When students feel that they are a part of a community, even when it is strictly online, they show improvements in their mental health (Tucker, 2012). Loades et al., (2021) has noted that the nature of the current environment, brought on by the COVID 19 pandemic and the continued shift towards more use of technology in education, has resulted in an increased sense of isolation for many students and increasing mental health problems. In this kind of context, there is a need to develop the resources for detecting symptoms of mental health problems as early as possible and using methods that can be deployed in an online environment.

## Text analysis and opportunities

The digitization of essays brought on by online education and assessment has provided for opportunities for computerized analyses of student-created texts possible in an unprecedented way. Educators get rapid and direct access to what students have written without having to read through essays and transcribe them exactly as written. This is essentially an instant process with

relatively low possibilities of transcription errors. Online assessment, in particular, has made comprehensive text analysis of students' writing possible.

There are different methods of text analysis depending on the content area being studied and the purpose of the text analysis. One of the more popular methods for general analysis of text is natural language processing (NLP). A prominent use of NLP is looking for semantic meaning in a corpus of text (Dumais, 2004). This approach has been used in latent semantic analysis (LSA; Foltz, Kintsch, & Landauer, 1998; Landauer, Foltz, & Laham, 1998). The method being proposed in this research uses latent Dirichlet allocation (LDA; Blei, Ng, & Jordan, 2003) which looks for latent themes in a corpus of text. If a student were to write something indicating risky behavior (e.g., "I want to kill something"), scoring using LSA might not trigger any sort of review as educators would just be presented the score on the student's writing. This makes it harder to detect risky behavior through writing as the LSA algorithm is not likely to be used to detect underlying themes in the writing. Latent clusters including words such as "kill", however, would be detectable in an LDA analysis. I am proposing using a combination of sentiment analysis and topic modeling as the desired method of analysis because this type of topic modeling can be used to detect negative sentiment indicators. These are explained below.

Sentiment analysis

Sentiment analysis is more formally defined as "the process of computationally identifying and categorizing opinions expressed in a piece of text to determine the writer's attitude towards a topic" (Sentiment Analysis, n.d.). This method of analysis automatically defines the valence, emotions, and other affectual states from text (Mohammed, 2016). Sentiment analysis considers the three prominent dimensions of meaning, valence (good-bad), potency (strong-weak), and activity (active-passive) (Osgood, Suci, & Tannenbaum, 1957).

Valence, in the context of sentiment analysis, can be defined as the tone of the text on a spectrum of positive, negative, and neutral (Mohammed, 2016). Words such as anger or fear have a negative valence and words such as joy have a positive valence. In sentiment analysis, valence can also be described as the polarity, which denotes how positive or negative a word or statement may be. Potency defines the strength of the polarity which is how powerful or weak the statement may be. For example, something could bring a "little joy" which has weak potency or "much anger" which is strong potency. Lastly, there is activity which defines the action of the word. An example of an active word is "exciting" while a passive word is "calm" (Xiang et al., 2020).

The majority of sentiment analyses use statistical machine learning techniques (Mohammad, 2016). However, there are still a few approaches that utilize rule-based approaches (Neviarouskaya, Prendinger, & Ishizuka, 2009). Sentiment analyses that use statistical modeling software for automated detection commonly use supervised approaches (Mohammed, 2016). In supervised learning approaches, the algorithm first learns the model through a pre-designated set of correctly identified sentiment, which is a process called training. The algorithm is then able to predict sentiment based on what it has learned through the training process.

A part of training the model is determining sentiment association with words. Some words are naturally associated with positive valence such as terrible or excellent. Words that are more neutral or can be seen as either positive or negative require more interrater agreement to assign a sentiment association. Sentiment association can also be done at the sentence level. These sentiment associations can be generated manually or automatically and added into a lexicon for the algorithm. Automated sentiment association lexicons also add weights or scores to each word that indicate the strength of association between the word and the affect category.

For example, the word, terrible, may be given a score of -1 while the word, terrific, may be given a score of +1. These scores can be combined to calculate sentiment. These predetermined scores represent a prior estimate in sentiment association. Priors are predetermined parameters for statistical machine learning analyses. There are several sentiment association lexicons available online that can be freely used (Mohammad & Turney, 2012; Warriner, Kuperman, & Brysbaert, 2013) as manually assigning scores for each word and sentence can be extremely time consuming.

Sentiment analysis is commonly used to detect opinions towards products or people (Mohammed, 2016) although it has been implemented in a variety of different contexts. The results from sentiment analyses are generally used to improve relations and processes and to predict future actions (Bougie, Pieters, & Zeelenberg, 2003). Some examples of detecting sentiment through text include predicting heart attack rates through sentiments in tweets (Eichstaedt, et al., 2015), confirming theories in literary analysis (Hassan, Abu-Jbara, & Radev, 2012), detecting satisfaction through product reviews (O'Connor, et al., 2010) and automatically detecting cyber-bullying (Nahar, Unankard, Li, & Pang, 2012). Sentiment analysis, employing automatic methods, have also been able to detect depression in adults (Cherry, Mohammad, & De Bruijn, 2012; Pennebaker, Mehl, & Niederhoffer, 2003) and college students Rude, Gortner, & Pennebaker, 2004). In the field of education, sentiment analysis has been used to detect emotions in students' evaluation responses towards classroom activities (Nunez, 2020). Research suggests that it is possible to detect emotions, feelings and attitudes through text using sentiment analysis. Research has shown certain mediums may be better more likely to elicit negative sentiment than others. Havigerova et al. (2019) found in their study that the most useful genre for detecting the signs of depression was informal text such as holiday letter

correspondence. Another common form of informal text that is analyzed for negative sentiment is workplace emails. Mohammed and Yang (2013) were able to track emotions and negative sentiment through workplace emails. Das, Kim & Kothari (2017) analyzed corporate email content to assess factors predicting escalating risk or the gradual shifting of other critical characteristics within the firm before they are eventually manifested in observable data such as financial outcomes. Das et al.'s (2017) results suggested that email content meaningfully predicted increased risk and potential malaise. Other structural characteristics, such as the average email length, were also found to be strong predictors of risk and subsequent performance. Das et al. (2017) presented three spatial analyses of internal corporate email communication, vocabulary trends, and topic analysis. Results indicated that sentiments extracted from emails accurately predicted increased risk and declining health to the company. These results demonstrate the potential of using sentiment analysis to predict and flag potentially problematic behavior.

Another non-formal medium that has been popular for sentiment analysis is tweeting. Sentiment analysis has been used to determine people's sensitivity and behavior, as well as the intensity (i.e., positive, negative, and neutral) of these sentiments and behavior (Kirelli & Arslankaya, 2020). Bae and Lee (2012) were able to distinguish between the positive and negative tweets based the interactions between followers and popular public figures. These results indicate the ability of sentiment analysis to detect positive and negative sentiment through short texts.

There is also a difference in how negative sentiment may be expressed depending on the person. Havigerova, Haviger, Kucera, & Hoffmannova, (2019) found differences in writing between males and females in how they expressed negative sentiment, both in-person and

through text. In addition, Havigerova et al. (2019) also found men and women expressed signs of depression differently through text. Mohammed and Yang (2013) detected differences between genders in the use of emotion-laden words in work-place emails. Mohammed and Yang (2013) found that women tend to use more words from the joy to sadness lexicon while men were more likely to use words from the trust to fear lexicon.

Using sentiment analysis of students' writing can potentially provide educators with a similar type of early warning system to help teachers detect the potential signs of students' negative sentiments before they are manifested in risky behavior or worsened mental health. As suggested earlier, this is especially important in light of the recent increase in online instruction and subsequent social isolation due to the pandemic. Students are online some or most of the time due the pandemic. Isolation and lack of social interaction with peers has been shown to have negative impacts on students' mental health (Cao et al., 2020; Elmer, Mepham & Stadfeld, 2020; Sit, Chung, Chow, & Wong, 2005).

Utilizing topic modeling as a method of sentiment analysis would provide educators another tool for detecting the potential for risky behavior.

Potential presentation of disorder indicators

Externalizing disorders such as behavioral problems, are relatively easy for school personnel to identify in in-person settings in the school (Weist, Rubin, Moore, Adelsheim, & Wrobel, 2007). These disorders are often indicative of early or actual signs of mental health problems in students (DeSocio & Hootman, 2004). The internalizing problems, however, such as depression, anxiety, and suicidal ideation are more subtle and more difficult for school personnel to identify (DeSocio & Hootman, 2004; Weist et al., 2007; Allison, Nativio, & Mitchell, 2013). Ruchkin, Sukhodolsky, Vermeiren, Koposov, & Schwab-Stone (2006) found in a cross-cultural

study investigating depressive symptoms in students that depressive symptoms were associated with increasing levels of internalizing as well as externalizing. Depression can also be comorbid with other internalizing disorders (such as anxiety disorders) or can have an external presentation among children (Hendron & Kearney, 2016). Mental health problems can lead to a cycle of poor performance in school. For example, children who have mental health problems are more at risk of poor school performance and children who have difficulty with learning are more at risk of developing mental health problems (DeSocio & Hootman, 2004). So, it is important to detect the signs of mental health problems early.

Rude, Gortner, & Pennebaker (2010) examined essays using sentiment analysis written by currently depressed, formerly depressed, and never depressed college students for differences in language that reflected the cognitive operations associated with depression and depression-vulnerability. A text and sentiment analysis program called Linguistic Inquiry and Word Count (LIWC) computed the incidence of words in predesignated categories. Consistent with Beck's (1967) cognitive model and with Pyczsinski and Greenberg's self-focus model of depression (Greenberg & Watson, 2006), depressed participants used more negatively valanced words and used the word, "I", more than never-depressed participants. Formerly depressed (presumably depression-vulnerable) participants did not differ from never depressed participants on these indices of depressive processing. However, consistent with the models and prediction, formerly depressed participants' use of the word "I" increased across the essays and was significantly greater than use by never depressed writers. Pennebaker et al. (2003) found that pronoun usage provided insight into a person’s emotions as pronouns differentiate the markers of self and group identity. These results indicate the potential to detect indicators and presentations of negative sentiment or disorders, such as major depression, through text.

Allison, Nativio, & Mitchell, (2013) identified symptoms of depression and anxiety in students in the school setting for depression and anxiety. These symptoms include depressed mood, loss of interest or pleasure, loss of concentration, weight loss or gain, insomnia, hypersomnia, fatigue, decreased concentration, thoughts of death or suicide, restlessness, irritability, etc. Some of these symptoms such as loss of pleasure, thoughts of death or suicide can manifest themselves in text (Nobles, Glenn, Kowsari, Teachman, & Barnes, 2018; Ophir, Tikochinski, Asterhan, Sisso, & Reichart, 2020).

Two tools can be used for identifying episodes of psychopathology: (1) DSM-5 and (2) Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS). The Diagnostic and Statistical Manual of Mental Disorders (DSM) was developed by the American Psychiatric Association to create an American system to resemble the International Classification of Diseases series created by the World Health Organization (American Psychiatric Association, 2013). Since its inception, the DSM has been updated five times and the most updated version is referred to as the DSM-5 (Widiger & Clark, 2000). The purpose of the DSM is to provide a scientifically based nomenclature for the classification of psychiatric. The K-SADS uses DSM-5 criteria to diagnose episodes of psychopathology in children through semi-structured interviews (Kaufman, & Schweder, 2004).

The DSM-5 and KSADS, two tools for identifying episodes of psychopathology, look for phrases such as “not interested, annoyed, hurting myself, don’t feel like eating, etc.,” as potential indicator flags for major depression. These same phrases or words can show up in students’ writing and could be flagged as indicators of potential depression or anxiety.

Disorders, such as generalized anxiety disorder, social anxiety disorder, disruptive mood dysregulation disorder (DMDD), major depression, panic disorder and PTSD, all present

differently at different ages and external/internal signs vary from person to person. In person, school personnel can sometimes catch these signs early as they can observe students in person and over time in their day-to-day activities. Even when schools are back to teaching full-time in person, having another method to detect for presentations of disorders and negative sentiments can be helpful to educators and students.

## Topic Modeling

Topic models (Blei, 2012) provide a potentially useful method for detecting differences in the patterns of students' use of latent themes in their constructed response (CR) answers. As suggested above, topic models are statistical algorithms for detecting the latent themes in a corpus of documents. These have been applied to detect latent themes in large corpora of documents as well as patterns in genetic data, images and social networks (Blei, 2012). Topic models have also been used for detecting latent patterns in use of topics in twitter messages (Paul & Dredze, 2011), in pre-service teachers' reflective texts (Chen, Yu, Zhang & Yu, 2016; Gibson & Kitto, 2015), and in MOOC discussion forums (Ramesh et al., 2014; Reich et al., 2016). In the context of educational testing, they have also been used to detect students' growth and change in use of academic vocabulary on middle grades tests of science inquiry knowledge (Kim et al., 2017).

LDA is one of the simplest topic modeling algorithms. The purpose of LDA is to detect the latent thematic structures from a corpus of documents. LDA assumes that the documents are random mixtures of words over latent topics. That is, each document can have a different probability of using each of the topics in the model. Further, each topic is assumed to be composed of a random mixture of words. This means that the words in the corpus appear in each topic with different probabilities. The latent topics are composed of clusters of words that co-

occur in a collection of documents. The words all have different probabilities of co-occurring in each topic and the topics all have different probabilities of being used in each document. The words that co-occur with the greatest frequency have the highest probability for that topic and are typically used to help interpret the meaning of the topic. The assumptions and mechanisms of the LDA methodology will be further discussed in the methodology section.

Topic modeling has been used as a tool for detecting aggressive sentiments in short texts, such as comments, reviews, tweets etc. Ventirozos, Varlamis, & Tsataronis (2017) presented an approach in which sentiment analysis was used at the message level to analyze sentiment while also taking into account the whole communication thread. In this context, Ventirozos et al. analyzed users' discussions boards on the internet and were able to use topic modeling to identify aggressive behavior on certain boards. Results indicated it was possible to detect aggressive, inappropriate or antisocial behavior such as cyberbullying through the use of this type of text mining. In this dissertation, I propose to investigate the effectiveness of LDA for detection of negative sentiments and negative topics in answers to constructed response tests.

# CHAPTER 3

# METHODOLOGY

## Pre-processing of data

The first steps in the use of topic models is to pre-process the textual data to prepare it for analysis (Boyd-Grabber, Mimno & Newman, 2014). Pre-processing helps to ensure that the results are more usefully interpretable. First, punctuation is removed from the responses and all responses are converted to lower case (Wheeler et al., 2020). Stop words, which are high frequency words that don't often carry important semantic information are removed. Examples of stop words are "the", "what", "is", etc. Pre-processing the data eliminates such things as misspellings and typos, which could be detected as separate unique words. Words that are shorter than a minimum length may be removed as well. The pre-process procedures also include stemming the words. Stemming the vocabulary converts all different tenses and forms of a word to a base word so that the same base words of different words are not considered as separate unique words. For example, buy is the base word. Buying and buys would be converted to buy during preprocessing. Finally, blank responses are removed.

Second, an expert with specializations in clinical psychology will be consulted to help understand and create criteria and methods for identifying potential indicators of these conditions in students' writing. Based on the expert's input, criteria will be constructed to assist in detecting the potential negative sentiment and risky behavior indicators in students' written text. The next paragraphs will depict how the criteria for detecting risky behavior and negative sentiment was constructed using the DSM-5 and K-SADS tools.

DSM-5 indicators: criteria selection

The DSM-5 manual will be used in conjunction with the semi-structured interviews provided in the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS). The K-SADS was used to create criteria for potential indicators of the following disorders that may present themselves in children's and adolescents' text: generalized anxiety disorder, social anxiety disorder, disruptive mood dysregulation disorder (DMDD), major depression, panic disorder and PTSD. The disorders listed above were selected after consulting with a clinical psychologist, who holds a PhD in Clinical Psychology, about which disorders might be most likely to present themselves through text and for the proposed study students' age group. The clinical psychologist also advised on potential indicator criteria as well as the selection process.

The K-SADS diagnostic tool contains add-ons such as written fact sheets in which children and adolescents can write down their feelings (e.g., about anxiety, depression, etc.) and can assist in providing structures to identify patterns in students' writing. The K-SADS diagnostic tool provides words to look for in their semi-structured interview documents as well as in the written fact sheets, that can help clinicians in diagnosis (Abeles & Verduyn, 2009).

The six disorders listed above are described in the DSM-5 and found in the corresponding K-SADS interview diagnosis tool. The K-SADs has structured, semi-structured and non-structured interviews to diagnosis disorders from the DSM-5. The semi-structured interviews were used as a reference to identify potential indicators due to its format. That is, the format of the semi-structured interview proposes questions for patients and allows them to answer how they see fit. These questions do not seek a yes or no answer. The K-SADs then has suggested phrases that someone with a certain disorder might present with. For example, based on the K-SADS and DSV-5 criteria, someone presenting with PTSD might say "I can't trust anyone ever

again". For this particular example, the phrase, "I can't trust anyone ever again", would be documented, added to an indicators database, and the individual words as well as the sentence as a whole, would be set as potential indicator flags in students' essays. If the individual indicator words, such as "trust", were to appear among the most used words in a topic structure, this could help characterize and interpret the topic. Topics that contain the most indicator words could potentially be identified as a flagging topic and essays with high probabilities of occurrence in this topic may be more likely to display the indicators that are being looked for.

Phrases listed by the K-SADS and DSM-5 for each of the six specified disorders, generalized anxiety disorder, social anxiety disorder, disruptive mood dysregulation disorder (DMDD), major depression, panic disorder and PTSD, were identified and used as indicators for the topic modeling analysis. Key words that were designated by the KSADS and DSM-5 were also recorded and added to the potential indicator database of phrases and words.

Creating the sentiment lexicon

The sentiment lexicon is the set of terms that were used as indicators of negative sentiment. The lexicon was created for this study by looking at the semi-structured interviews from the K-SADS manual, for each of the individual disorders, and identifying key words from the structured interview that were related to each of the disorders. The words were then compiled into a list that formed the sentiment lexicon. The method for adding words to the sentiment lexicon follows the example earlier. If the K-SADS and DSM-5 manual identify the lack of trust as a potential indicator of anxiety and as something that a person with this disorder might say, the indicator word, "trust", is added to the sentiment lexicon. Indicator words from the six disorders were all added to the sentiment lexicon (see Appendix D).

For purposes of this study, any words identified from the K-SADS and DSM-5 comparison that were added to the lexicon, were labeled as negative sentiment. The presentation of any of the words in students' writing indicated the potential for negative sentiment. In theory, the more words from the lexicon that appeared in students' writing, the more negative sentiment the essay would display. Essays with large occurrences of the lexicon may be indicative of the potential signs of negative sentiment and/or disorders.

Once the criteria were identified, exploratory LDA analysis was used on the CR answers from the English and Language Arts test for 9$^{th}$ grade students. The objectives of this analysis were to determine whether it was possible to identify signs of depression, anxiety and negative sentiments in students' written essays using topic modeling and the created sentiment lexicon.

LDA was then used to extract topics from the corpus. This was done through using a collapsed Gibbs sampling algorithm implemented in the R code package, *topicmodels* (Hornik & Grun, 2011). The package is used to estimate the parameters for the model through a Monte Carlo Markov chain (MCMC) algorithm (Chang, 2010; Chang, 2015; Heinrich, 2009). The MCMC algorithm has two phases, a burn-in phase to allow the MCMC chain to converge, and a post-burn-in phase to estimate the parameters of interest. Convergence will be determined through the use of the R package, *topicmodels* (Grun & Hornik, 2011). Through the *topicmodels* package, convergence was assessed by setting a maximum number of iterations and a tolerance for the relative change in the likelihood for both the variational inference and EM algorithm steps. Following convergence, the post-burn-in phase is used to obtain the posterior estimates of the parameters.

LDA

LDA is an unsupervised, probabilistic model that can be used for analyzing the thematic structure of a corpus of textual documents (Jelodar et al., 2019). That is, it is a statistical model designed to detect the latent topics in a corpus of documents (Blei et al., 2003, 2013; Jelodar et al., 2019; Mardones-Segovia et al., 2021).

LDA assumes that each document can be represented as a probabilistic distribution over latent topics. In the LDA model, the topic distributions in all documents in the corpus share a common Dirichlet prior (Jelodar et al., 2019). In LDA, a collection of *D* documents, is referred to as a corpus, and each corpus is a mixture over *K* topics. Each topic is a probability distribution over the vocabulary of *V* unique words (Blei et al., 2003).

LDA estimates two component parameters: topics, which is a corpus-wide parameter, and topic proportions, which is a document-wide parameter. Topics, represented by their different probabilistic distributions are a set of probabilities over the vocabulary in the corpus and are denoted by $\phi_k$, where $k \in 1,2, \dots, K$. Topic proportions, $\theta_d$ where $d = 1, \dots, D$, also referred to as document-topic probabilities, are a mixture of proportions over the *K* topics for each document. Topic proportion, $\theta_d$, is a document-level variable that is sampled once per document. LDA also estimates the topic assignment parameter. The topic assignment parameter, $z_{d,n}$ estimates the topic membership for every word in each document and is denoted where $d = 1, \dots, D$ and $n = 1, \dots, N^{(d)}$. $z_{d,n}$ are word-level variables that are sampled once for each word in each document and $N^{(d)}$ indicates the number of words in document *d* (Blei et al., 2003).

LDA assumes the topics ($\phi_k$) follow a Dirichlet distribution with a prior hyperparameter $\beta$, topic proportions ($\theta_d$) follow a Dirichlet distribution with a prior hyperparameter $\alpha$, and topic assignments ($z_{d,n}$) follow a multinomial distribution with parameters $\theta_d$ and $N^{(d)}$ (Blei et al., 2003; Mardones-Segovia et al., 2021). LDA utilizes the topic assignment from each word to

estimate the topics and topic proportions by assuming the words for each document ($w_{d,n}$) follow a multinomial distribution with parameters $\phi_k$ and $N^{(d)}$ (Ponweiser, 2012). The joint probability distribution of all latent and observed variables for LDA is expressed as

$$p(w, z, \theta, \phi|\alpha, \beta) = \prod_{k=1}^{K} p(\phi_k|\beta) \prod_{d=1}^{D} p(\theta_d|\alpha) \prod_{n=1}^{N^{(d)}} p(z_{d,n}|\theta_d) p(w_{d,n}|\phi, z_{d,n}),$$

where $p(\phi, \beta)$ is the probability distribution of all topics over the vocabulary in the corpus, $p(\theta|\alpha)$ is the probability distribution of the topic proportions of a document, $p(z_n|\theta)$ is the probability distribution of the topic assignments of a document, and $p(w_n|\phi, z_n)$ is the probability distribution of the observed words of a document (Mardones-Segovia et al., 2021; Ponweiser, 2012).

Integrating over $\theta$ and $\phi$, and summing over $z$, the marginal likelihood of a document is

$$p(w|\alpha, \beta) = \int_{\phi} \int_{\theta} p(\theta|\alpha) p(\phi|\beta) \prod_{n=1}^{N} \sum_{z_n} p(z_n|\theta) p(w_n|z_n, \phi) d\theta d\phi$$

The marginal likelihood of each document is used to calculate the likelihood of a corpus by taking the product of each document's marginal likelihood. It can be expressed as:

$$p(W|\alpha, \beta) = \prod_{d=1}^{D} p(\boldsymbol{w}_d|\alpha, \beta)$$

It is not possible to obtain exact estimates of the parameters (Blei et al., 2003) so different methods have been used to estimate parameters. Markov chain Monte Carlo estimation (Blei et al., 2003; Griffiths & Steyvers, 2004) and collapsed Gibbs sampling (Griffiths & Steyvers, 2004) are two of the methods that have been used to estimate LDA parameters (Mardones-Segovia et al., 2022.

Prior hyperparameters play an important role in the estimation of the LDA model and are specified before running LDA (Syed & Spruit, 2018). The Dirichlet distribution assumes the parameters are mutually independent and the proportions of documents and words within topics are independent (Geiger & Heckerman, 1997). α and β are the symmetrical Dirichlet hyperparameters whose values influence the document-topic and topic-term probabilities, respectively (Mardones-Segovia et al., 2021; Vayansky & Kumar, 2020). Symmetrical priors assume *a priori* that all topics have equal probability to be assigned to a document and all words have an equal chance to be assigned to a topic (Syed & Spruit, 2018). Symmetrical priors are often used when prior values are equal.

α represents the prior count of the times an individual topic is observed in a document and is represented as $\alpha_i > 0$ for *k*-vectors of elements (Vayansky & Kumar, 2020). The α parameter influences the combination of topics with large α causing the distribution of topics to move from extremities, referred to as smoothing, and small α values spreading distributions (Laudauer, et al., 2013; Vayansky & Kumar, 2020). The β parameter is defined as the number of occurrences of words sampled from a topic prior to the observation of any word within the corpus (Laudauer, et al., 2013; Vayansky & Kumar, 2020). β determines the degree of smoothing of word distributions over all the topics. α and β values larger than one indicate documents and words are equally likely to belong to all topics while α and β values less than one indicate the distributions of documents and words are sparser. When α and β have low values, LDA yields document and topics that are mutually exclusive (Syed & Spruit, 2018).

The LDA utilizes an algorithm that consist of two main components, the Markov chain Monte Carlo (MCMC) and Gibbs sampling (Porteous, Newman, Ihler, Asuncion, Smyth & Welling, 2008) to train and select models.

MCMC is an approximation and inference algorithm that uses Monte Carlo integration. Monte Carlo integration is used to form a Markov chain, which is a "string of arbitrary variables whose transition probabilities among distinct values in the state space are solely contingent on the existing state of the random variable" (Vayansky & Kumar, 2020, p. 11). The MCMC method uses the previous sample's value to produce the value of the next sample at random. This method simulates directly drawing from a complicated probability distribution such as high-dimensionality distribution and is done iteratively (Vayansky & Kumar, 2020; Walsh, 2004).

Gibbs sampling. The purpose of Gibbs sampling is to approximate posterior inference. The Gibbs sampling method was derived from the MCMC algorithm and seeks to generate a Markov chain with the desired posterior distribution as its stationary distribution. Gibbs sampling draws from a conditional distribution of the posterior variable (Vayansky & Kumar, 2020; Walsh, 2004). Gibbs sampling generates posterior samples by sweeping through each variable to sample from the variables' conditional distribution while the remaining values are fixed to their current values (Lynch, 2007; Yildirim, 2012). This process continues until convergence is achieved. This means the stationary distribution should reflect the posterior distribution after sufficient post-burn-in iterations of drawing samples (Vayansky & Kumar, 2020; Walsh, 2004). In LDA, Gibbs sampling takes into account two parts: (1) how much each topic is present in a document and (2) how much each topic is associated with a word. For each word, Gibbs sampling assigns a vector of probabilities that denotes how likely a word will be assigned to a certain topic. Gibbs sampling in the LDA model (Griffiths & Steyvers, 2004; Grun & Hornik, 2011) draws from the posterior distribution $p(z|w)$ obtained by the sampling from

$$p(z_i = K)|w, z_{-i}) \propto \frac{n_{-i,K}^{(j)} + \delta}{n_{-i,K}^{(.)} + V\delta} \frac{n_{-i,K}^{(d_i)} + \alpha}{n_{-i,.}^{(d_i)} + k\alpha}$$

$z_{-i}$ is the vector of current topic memberships of all words without the *i*th word $w_i$. $j$ indicates that $w_i$ is equal to the *j*th term in the vocabulary. $n_{-i,K}^{(j)}$ gives how often the *j*th term of the vocabulary is currently assigned to topic K without the *i*th word. The dot . implies that summation over this index is performed. $d_i$ indicates the document in the corpus to which word $w_i$ belongs. In the Bayesian model, β and α are the parameters of the prior distributions for the term distribution of the topics and the topic distribution of documents, respectively (Grun & Hornik, 2011).

In theory, Gibbs sampling methods will eventually result in the convergence of the samples that reflect the true posterior distribution. However, convergence is not always guaranteed and may not be necessary. Once the model has converged, this indicates an equilibrium has been achieved; this is called stationarity (Vayansky & Kumar, 2020). Variation between the data points from the stationarity point onwards is defined as standard error (Walsh, 2004; Vayansky & Kumar, 2020). There is also a method of addressing samples from the beginning that do not accurately reflect the desired distribution. This is called the "burn-in" phase. The burn-in phase refers to the phase in which the algorithm has yet to converge such that the posterior does not represent the correct distribution. The iterations in this phase are discarded. This phase assists the algorithm in reaching convergence.

A concern with the LDA algorithm and approximation method is the final results (the topics) for the same data can be different due to the settings of the initial conditions which results in difficulty with model replication and significant noise issue (Steyvers & Griffiths, 2011). The differences in topic structure from the algorithm could be due to "order effect". The order effect causes the variations in topics produced. This is due to the different training data sets it utilizes and the order in which they are utilized by the algorithm (Agrawal, Fu & Menzies, 2018). Each

iteration the model creates could result in different topics due to the variation in training datasets. Due to the order effect, iterative testing of different *k*-values has been found to not be effective in optimizing the number of topics. In this context, *k* represents the dimension of the distribution, the dimension of the topic variable, z, and specifies the total number of topics for the model (Vayansky & Kumar, 2020).

The exploratory use of LDA will then yield topic structures based on the different candidate models. The topic model analysis provides a probability of membership of each word on each topic. The best fitting topic model will be analyzed for the 30 words with the highest probability for each topic. Analysis of these words provided some idea of what the interpretation was of each topic.

The exploratory LDA analysis will use the R package of "topicmodels". The *topicmodels* package will be used to estimate the LDA model by applying a collapsed Gibbs sampling algorithm to the document-by-word matrix (Hornik & Grun, 2011).

*topicmodels* R Package

The R package, *topicmodels*, provides an interface for fitting an LDA model using the variational expectation maximization (VEM) algorithm and Gibbs sampling as implemented by Blei et al. (2003), (Grun & Hornik, 2011). The package requires the number of topics to be fixed a-priori. In *topicmodels*, the LDA model assumes the following generative process for a document $w = (w_1, \ldots, w_n)$ of a corpus, *D*, containing *N* words from a vocabulary consisting of *V* different terms, $w_i \in \{1, \ldots, V\}$ for all $i = 1, \ldots, N$. For the LDA model, *topicmodels* implements the generative model by the following steps:

(1) Determine the distribution of β by β ~ Dirichlet(δ) where β is the term distribution of topics and contains the probability of a word occurring in a given topic

(2) Determine the proportion θ of the topic distribution of the document by θ ~ Dirichlet(α)

(3) Choose a topic, $z_i$ ~ multinomial (θ) for each of the *N* words $w_i$

(4) Choose a word, $w_i$, from a multinomial probability distribution conditioned on the topic $z_i$ for each of the *N* words $w_i$

For the maximum likelihood (ML) estimation of LDA, *topicmodels* maximizes the log-likelihood of the data with respect to the model's hyper parameters, α and β. The estimation of the log-likelihood for one document for LDA is denoted as:

$$\ell(\alpha, \beta) = \log\left(p(w|\alpha, \beta)\right)$$

Gibbs sampling estimation requires the values for the parameters of the prior distributions, α and β, to specified a-priori (Grun & Hornik, 2011). For Gibbs sampling, the log-likelihood is given by:

$$\log\big(p(w|z)\big) = k\log\left(\frac{\Gamma(V\delta)}{\Gamma(\delta)^V}\right) + \sum_{K=1}^{k}\left\{\left[\sum_{j=1}^{V}\log\left(\Gamma(n_K^{(j)} + \delta)\right)\right] - \log\left(\Gamma(n_K^{(\cdot)} + V\delta\right\}\right.$$

*topicmodels* has the ability to construct a corpus, transform the corpus into a document-term matrix and select the vocabulary from the matrix. The document-term matrix is the input data for topic models. The rows in the matrix correspond to the documents and the columns to the terms. The number of rows is equal to the size of the corpus and the number of columns to the size of the vocabulary. The data pre-processing step involves selecting a suitable vocabulary, which corresponds to the columns of the document-term matrix. Vocabulary can be selected from the matrix by different criteria such as only selecting terms that occur in a minimum number of documents (Griffiths & Steyvers, 2004) or the terms with the highest tf-idf scores (Blei & Lafferty, 2009). See Appendix F for the application of *topicmodels* in RStudio code.

Model Selection

Exploratory analysis of the data will be conducted using LDA to determine the best fitting number of topics in the model. This means that candidate LDA models ranging from one to ten topics will be fit to the data. Model fit indices will be used to select the LDA model with the number of topics that is the best fit from the set of candidate models.

Model Fit. Model fit indices developed from the need to address model fit to the data. There needed to be an external method of identifying the optimal number of topics. The model fit indices assume identifying the best number of topics will lead to selecting the best fitting LDA model. There are two common indices for comparing model fit in LDA models: likelihood-based statistics and information criterion (IC) and similarity measures (Mardones-Segovia et al., 2021). IC indices have been used to analyze best topic structures in a variety of fields (Mardones-Segovia et al., 2021; Schroder et al., 2017; Wang et al., 2016) as well as likelihood-base statistics and similarity measures (Anderson et al., 2020; Mardones-Segovia et al., 2021; Roque et al., 2019). Common fit indices used for LDA model selection are DIC, AIC, BIC, SABIC, average cosine similarity$(\overline{CS})$, and the Jensen-Shannon divergence (JSD) (Mardones-Segovia et al., 2021).

There are popular IC indices for comparing model fit such as deviance information criterion (DIC), Akaike information criterion (AIC), sample-size adjusted BIC (SABIC) and Bayesian information criterion (BIC) values (Silva et al., 2017; Sclove, 1987; Spiegelhalter et al., 2002). These indices also fall in the category of likelihood-based statistics. These IC are composed of two parts: a goodness-of-fit term and a penalty term to denote model complexity (Mardones-Segovia et al., 2021). For IC indices, the topic model with the lowest of IC values is

usually selected as the optimal model. However, the lowest DIC, AIC, SABIC, BIC values do not always guarantee that this is the number of topics that will be selected.

The Deviance Information Criterion (DIC) estimates the maximum likelihood estimation for the deviance "using the average deviance of all parameter estimates" (Callaghan, 2012; Spiegelhalter et al., 2002). Deviance measures uncertainty from the fitted model and suggests the best model will have the lowest uncertainty and therefore, lowest DIC (Callaghan, 2012). DIC values are usually used to evaluate the models' properties and to infer parameter estimates when using a Bayesian algorithm (Silva et al., 2017; Xiong, Choi, Kim, Kwak & Cohen, 2019). This makes DIC the most commonly used model fit index for LDA with smaller DIC values indicating the better fitting model (Xiong et al., 2019). A concern with DIC values is that DIC tends to overfit the data (Callaghan, 2012). Thus, in this study, we use additional fit indices along with the DIC.

AIC estimates the information lost between the true model and candidate models (Clarke, 2007). It is impossible to identify the true model in reality so AIC helps detect the model that has the smallest amount of information loss between the candidate(s) and true model (Vrieze, 2012). Burnham and Anderson (2002) define AIC as:

$$AIC = D' + 2p$$

where *D',* the goodness-of-fit term is the posterior mean of deviance of the LDA model and *2p*, the penalty term, denotes twice the number of parameters estimated in the model. A shortcoming of the AIC index is it does not take into account sample size and performs worse as sample size increases (Burnham & Anderson, 2002).

BIC is an index that does that sample size into account and as a result, tends to perform better as the sample size increases (Cohen & Cho, 2017). BIC imposes a different and stronger

penalty than AIC for the number of parameters in an LDA model and penalizes for over-parameterization. BIC does this by taking into account the natural log of the number of documents in the data multiplied by the number of parameters, $p$ (Mardones-Segovia et al., 2021; Sen & Bradshaw, 2017; Vrieze, 2012). BIC (Cohen & Cho, 2017) is defined as:

$$BIC = D' + p(log(D))$$

SABIC (Sclove, 1987) was developed to account for the shortcomings of both AIC and BIC by reducing the penalty of sample size and adjusting for model complexity. Kim (2014) found that SABIC could be used for small sample sizes and Yang (2016) found that SABIC could be used for large numbers of parameters. SABIC (Sclove, 1987) defines SABIC as:

$$SABIC = D' + p\big((D + 2)/24\big)$$

DIC is a Bayesian IC index that has a penalty term, $p_D$, that penalizes for two terms: the true estimate and the estimated residual information. The true estimate is defined as the posterior mean of the corpus and the estimated residual is defined as the deviance of the posterior (Cohen & Cho, 2017; Mardones-Segovia et al., 2021; Spiegelhalter et al., 2002). Spiegelhalter et al., (2002) define DIC as:

$$DIC = D' + p_D$$

There are also fit indices that are based on similarity measures which evaluate how similar the topics are to one another (Cao et al., 2009). These indices were developed from the need to identify the best topic models structures. The most ideal topic model structure has high within-topic similarity and low between-topics similarity (Cao et al., 2009, Mardones-Segovia et al., 2021). Cao et al., (2009) developed the cosine similarity (CS) measure to identify the distance between topics which is measured as average cosign distance between each pair of words in the candidate models. A CS value close to zero indicates the words between two topics

are not similar while a CS value close to one indicates the words between two topics are similar. Cao et al., (2009) defines CS in the equation below. The numerator denotes the dot product between the log of the word probability distribution for topic $\phi_i$and topic $\phi_j$, and the denominator denotes the dot product of the standard deviation of each topic (Cao et al., 2009, Mardones-Segovia et al., 2021).

$$\overline{CS} = \frac{\sum_{i=0}^{K}\sum_{j=i+1}^{K} CS(\phi_i, \phi_j)}{K \times (K-1)/2}$$

Cao et al., (2009) found that averaging the CS of each candidate model can assess which model produces the most distinct topics. Large CS value, those closer to one, indicates there are similar words between topics while low CS values, those closer to zero, indicate there is not similarity between the words among topics. The lowest average CS value would indicate the best topic model to select in this circumstance.

Deveaud et al. (2014) developed an index that identifies the best candidate model by analyzing the means of the probability of words in each topic. The metric proposed by Deveaud et al., (2014) uses Jensen-Shannon divergence (JSD) to estimate the optimal number of topics. Like the CS index, JSD looks for similarity between pairs of words within topics. JSD is represented in the equation below (Deveaud et al., 2014):

$$JSD = \frac{1}{2}\sum_{v \in V_k \cap V_{k\prime}} p(\phi_{k,v}) \times \log\left(\frac{p(\phi_{k,v})}{p(\phi_{k',v})}\right) + \frac{1}{2}\sum_{v \in V_k \cap V_{k\prime}} p(\phi_{k',v}) \times \log\left(\frac{p(\phi_{k',v})}{p(\phi_{k,v})}\right),$$

where $V_k = \underset{v}{\text{argmax}}[n]\phi_{k,v}$ is the top $n$ word probabilities for topic $\phi_k$. A low JSD value, values closer to zero, indicate similar words in a topic, while a high JSD value, indicates dissimilar words in a topic. Deveaud et al. (2014) expanded upon this metric to include the calculation of semantic similarity between the topics. He does this by summing the JSD between all possible pairs of topics and derives the best model by using the following equation:

$$\widehat{K} = \underset{K}{\operatorname{argmax}} \frac{1}{K(K-1)} \sum_{T_K} JSD$$

where that $T_K$ is the set of candidate topic models. The topic model with the most distinguishable terms and coherent words within topics is the best candidate model. This model would still follow the index from the earlier model with a low JSD indicating similarity and a high JSD value indicating dissimilarity. The most optimal model would have a high JSD value.

Fit indices can provide useful quantitative guidelines to select the best fitting model. However, the selection of the best fitting model should take into account model interpretability as well. Topic interpretability is a concern for model selection (Quinn et al., 2010). In other words, a model that may appear to be the best fit statistically but may not make sense upon interpretation is less likely to be helpful. An exploratory approach that takes into account model fit statistics and human interpretability can be used to determine the best fitting LDA model. The literature suggests using a combination of fit statistics and human judgment to determine the best fitting model (Grimmer & Stewart, 2013; Quinn et al., 2010). Grimmer and Stewart (2013) emphasize the importance of recognizing the results from automated algorithms will always be context specific and thus, will always require human validation of the measures produced by machine learning techniques and making sure the measures correspond with the concept claimed in unsupervised classification. A method of accomplishing validation could be demonstrating that the classification from machines replicates human hand coding (Grimmer and Stewart, 2013).

Akaike Information Criterion (AIC; Akaike, 1974) is another commonly used model fit index for topic model selection (Callaghan, 2012). AIC uses the “asymptotic result of the average log-likelihood” (Callaghan, 2012) to estimate the divergence of the model’s true distribution of data. The smaller the AIC value, the smaller the divergence distance estimated is.

A smaller AIC value is preferred in model selection. Bayesian Information Criterion (BIC; Schwartz, 1978) is also a popularly used model fit index. BIC estimates the approximation of the deviance between the fitted model and the null model (Callaghan, 2012; Raftery, 1995) with a smaller BIC value indicating a better fitting model. A concern with AIC and BIC fit indices is their inability to calculate values if the number of parameters exceeds the number of observations. This is a potential concern as LDA uses a hierarchical model that may result in more parameters than number of observations in a dataset (Callaghan, 2012).

AIC and BIC fit indices appear to be better indicators for simpler models while the DIC fit index appears to be more useful for more complex models (Callaghan, 2012). This would indicate that larger datasets with more documents, longer texts, and more unique words would benefit from using DIC over AIC and BIC. AIC and BIC have been shown to overestimate the number of topics as the sample size increased (Yamanishi, Wu, Sugawara & Okada, 2019).

Cosine similarity (CS) is another LDA model fit index, that examines the distance between topics (Cao et al., 2009). A CS value close to 0 indicates that words between two topics are dissimilar while a CS value closer to 1 indicates that the words between two topics are similar. Thus, a CS value closer to 0 would indicate the topics in the model are dissimilar. CS can then be used to compare candidate models and ideally be used to select the model with the closest value to 0. The sum of JSD is a model fit index that is similar to CS that evaluates the similarity of pairs of words within topics (Deveaud et al., 2014; Mardones-Segovia et al., 2021). Unlike the CS fit index, a value closer to 1 indicates dissimilar topics while a value closer to 0 indicates similar topics. The JSD fit index can be used identify the model with the most distinguishable terms between topics and the most coherent words within topics (Mardones-Segovia et al., 2021).

Mardones-Segovia et al. (2021) identified JSD and CS as the most accurate model fit indices in evaluating the optimal topic model. For the study, CS, JSD, semantic coherence, exclusivity, AIC, BIC, SABIC and DIC were used to evaluate the best topic model fit, with a focus on CS and JSD due to the literature. The other indices, semantic coherence, exclusivity, AIC, BIC, SABIC and DIC were used to provide some additional information and context. The best fitting model would then be selected based on the indices.

While index fit metrics are important in determining the best topic model fit, interpretability is also equally important. Interpretability can be assessed through reading essays that use the highest proportion of each topic. Reading the essays can help modelers classify topics and see which topic structures best classify the data. Due to the large sample size of 5,833 and the literature (Mardones-Segovia et al., 2022), the alpha and beta values were set to 1.

The next step following model selection would be topic structure evaluation. The algorithm was set to extract the 30 words with the highest probabilities for each topic. Once the topics were extracted and through that, the 30 highest probability words for that topic, the topics were examined for potential DSM-5 indicators of negative sentiment and/or risky behavior. The top 3-5 essays that used the highest proportion of each topic were analyzed to get a better sense of the sentiment and theme of each essay. The essays' topic proportions were then used to classify and label the different topics for the dataset.

Topic proportions usage was then used to analyze different responses to the same prompt. One essay will use a proportion of each topic and the total of the proportions will always equal 100%. For example, if the topic structure for the study was determined to be a 4-topic model, a student could use 40% of topic A, 25% of topic B 15% of topic C and 20% of topic D. From this example, it can be seen that student A used the highest proportion of topic A. Topic proportion

usage can be aggregated across cohorts to draw conclusions about similarities and differences between the cohorts. Topics could then be classified, or label based on the top 30 words used and the sentiment of the essays that used the highest proportion of that topic.

As described earlier in the literature review, the sentiment association lexicon was created by assigning scores or weights to words and sentences. For this study, this is the first type of sentiment analysis implemented. Weights were assigned to the predetermined phrases and words from the DSM-5 and KSADS. All the words from the sentiment lexicon that are used in each essay will be counted and totaled for each essay. If a student uses two words from the sentiment association lexicon, the score for that student's essay would be two. If the student uses two words, twice each, their score would be four. The top 5 essays of each topic will be scored using this logic to get a better sense of the sentiment of the essay. The assumption behind this scoring is that essays that have a higher weight will display more potential indicators for negative sentiment, risky behavior or other negative psychological disorders. Essays with a 0 score will have used no words that are taken as signs of potential negative sentiment indicators. Appendix D contains the table of words and phrases that were selected from the DSM-5 and KSADS as potential indicators.

Previous research (noted above) has suggested that negative sentiment and indicators of risky behaviors were likely to be present in only a relatively small number of essays. Furthermore, this use of constructed response answers had not been studied for possible indicators of expressions of this kind of negative sentiment. Thus, it would be important to determine whether or not the LDA would be able to detect these negative kinds of sentiments in what was expected to be a small number of tests.

Next, sentiment analysis and two human raters were used as validation measures for the study.

A second form of sentiment analysis was implemented to provide another lens of looking at the topic model structures. Sentiment analysis was conducted using the *sentimentr* package in RStudio. The *sentimentr* package is designed to quickly calculate text polarity sentiment in the English language at the sentence level and aggregated by entire essay responses for this study. This package utilizes an augmented dictionary lookup to identify a sentiment value from a sentiment dictionary. The sentiment dictionary is composed of four sentiment lexicons from Syuzhet, Bing, NRC and Afinn (Jockers, 2017). A score of -1 indicates negative sentiment and a score of 1 indicates a positive sentiment. The sentiment score can then be compared to the topic model proportion usage to get a better understanding of different sentiments in the essays.

A Pearson correlation was then calculated to identify potential relationships between the topic proportion usage and the sentiment analysis score.

Additionally, two trained essay scorers individually read through the top 10 essays as a validation measure. The essay scorers were both college graduates with bachelor's degrees in social sciences. The training consisted of the following: the scorers were asked to read 10 essays total that already were scored as positive, negative, or mixed by the primary investigator and familiarize themselves with the expectations. They were both then asked to try scoring 5 essays on their own to ensure their scoring matched that of the primary investigator. Once both scorers were able to correctly score all the test essays, they were then asked to score the essays for the study. These trained essay readers were asked to score the essays as negative, mixed, or positive sentiment.

Research question

The goals of this study were to develop and study the use of LDA for detecting the different latent thematic structures that occur in a corpus of answers to CR questions for the purpose of detecting at-risk students in the general student population. More specifically, in this study, topic modeling was used to answer the following research question:

Can topic models detect risky behavior-related or disorder-related negative sentiment in students' written answers to CR questions?

The study will use latent Dirichlet allocation (LDA) topic modeling to analyze students' essays for signs of internalizing disorders such as but not limited to depression, anxiety and negative sentiment.

The corpus used for analysis was comprised of students' answers to the constructed response (CR) items on an assessment of writing developed by the Georgia Center for Assessment, an independent research and development center housed within the University of Georgia. The Center constructed formative assessments aligned with State standards in four subject areas (English language arts - ELA, mathematics, social studies, and science) as well as with the State's end-of-grade standardized tests.

For purposes of this study, ELA tests were analyzed. ELA prompts provide students the most flexibility in their responses. ELA rubrics do not have the same scoring structure as math, science and social studies. Some ELA items are scored dichotomously and others are scored polytomously. Prompts in math and science require students to answer within the context to get points and are scored on a binary scale, right or wrong.

There are three different types of ELA prompts: narrative, argumentative and informational (sometimes referred to as explanatory). The structure of tests depends on the grade levels. Generally, the assessments in this program are comprised of at least one reading passage,

at least 3 selected-response items and at least two CR items (a short-answer item and an extended-response item). The CR extended-response question will be analyzed for purposes of this study.

Research has found that students as young as 9 and 10 can sufficiently express negative sentiment (Fivush et al., 2007; Polce-Lynch et al., 1998). However, students write more as they progress through school and topic modeling requires large amounts of text to produce the best models. As such, the 9th grade essays were selected for analysis. Students in these grades are able to sufficiently express negative sentiment and the prompts for these grades will produce more text than those of lower grades.

Data

The following data were used for this study:

Students' written answers to the CR questions on a standardized formative test in English and Language Arts (ELS) for the 9th grade argumentative prompt in the 2018-2019 school year. The dataset contained responses from 6,038 students. These students are from seventeen school districts and thirty-one schools across the state of Georgia.

There are three types of essay prompts in the ELA tests; narrative, descriptive and argumentative. The rationale for using the argumentative essays is due to the higher potential of negative emotion being elicited with this type of prompt. Essay prompts that lean towards more emotional topics (e.g., identity theft, social media anxiety loss), for example, may prompt students to write more about their own feelings (Jeanne, Gross, Persons & Hahn, 1998) than happier prompts or historical prompts (e.g., the history of baseball or adopting a puppy). Students may also display signs of anxiety, depression and other negative sentiment differently to particular types of prompts depending on their age.

The prompt for the 9th grade argumentative essay was about the impact of social media on peoples' lives. Social media is an important part of most, if not all, teenagers' lives (PrakashYadav & Rai, 2017) which contributed to the suitability of this essay type and prompt for the study. Students are more likely to make write more about a topic, in terms of essay length, and include more personal information when the topic is something that relatable to them (Kennedy, 1994). Since social media has polarizing effects on children (PrakashYadav & Rai, 2017), it could be possible for this topic prompt to elicit negative sentiment and indicators of risky behavior.

The LDA algorithm was used to analyze each dataset. The best fitting topic structure was estimated for the dataset. The pre- and post-process statistics as well as fit statistics for the dataset are listed in Appendix A. The pre- and post-process statistics depict the number of total words, unique words, and the average response length of essays before and after data cleaning.

# CHAPTER 4

## RESULTS

The fit indices indicated models with 2 or 3 topics would best fit the data. CS and JSD values of 3 and 2 were calculated, respectively. AIC, BIC, SABIC and semantic coherence also indicated that a 2-topic model structure would best fit the data.

Two-topic and three-topic models were then compared for interpretability and how well the models classified the data. Results showed that the top 10 essays from each of the three topics in the 3-topic model displayed the similar sentiment. Topic 1 reflected mixed sentiment. Topic 2 reflected predominantly negative sentiment. Topic 3 reflected predominantly positive sentiment.

In the 2-topic model, essays in both topics displayed mixed sentiment towards social media. There was not a clear classification of either topic. Based on the interpretability assessment, the 3-topic model was selected as the best fitting model.

LDA was used to analyze the 2018-2019 9th grade English Language Arts (ELA) argumentative essay dataset. The prompt, shown in Appendix C, is about social networking through social media for teens. The prompt lists 18 statistics, both negative and positive, to provide some context for the students. The prompt includes an additional passage about how identity theft can occur through social media. Students were required to read the introductory text about what social media is and then asked to “write an argumentative essay stating whether or not online social networking has improved the lives of teenagers.”

The top 30 words for each of the 3-topic model that was the best fit to the data are listed

in the table below. The preprocessing statistics, model fit indices values and stop words are given in Appendix A.

Table 1. 9th Grade Argumentative 3 Topic Model with Proportions

| **9th Grade Argumentative 3 Topic Model (2018-2019)** | | | | | |
|---|---|---|---|---|---|
| Topic 1 | | Topic 2 | | Topic 3 | |
| you | 0.070 | teens | 0.026 | media | 0.029 |
| they | 0.041 | network | 0.025 | identity | 0.027 |
| people | 0.023 | media | 0.023 | information | 0.026 |
| media | 0.021 | they | 0.017 | teens | 0.023 |
| get | 0.017 | friend | 0.016 | their | 0.020 |
| not | 0.014 | people | 0.016 | network | 0.018 |
| thing | 0.014 | teenager | 0.016 | theft | 0.017 |
| there | 0.014 | help | 0.014 | not | 0.016 |
| because | 0.012 | more | 0.013 | teenager | 0.016 |
| but | 0.012 | other | 0.013 | passage | 0.014 |
| what | 0.011 | world | 0.012 | they | 0.014 |
| all | 0.010 | use | 0.012 | this | 0.013 |
| know | 0.010 | improve | 0.011 | use | 0.011 |
| just | 0.009 | their | 0.011 | face | 0.009 |
| like | 0.009 | passage | 0.010 | more | 0.009 |
| their | 0.009 | new | 0.010 | online | 0.009 |
| teenager | 0.009 | this | 0.010 | personal | 0.008 |
| out | 0.009 | online | 0.010 | bully | 0.008 |
| some | 0.008 | many | 0.009 | steal | 0.008 |
| think | 0.008 | way | 0.009 | negative | 0.008 |
| make | 0.008 | also | 0.009 | many | 0.007 |
| use | 0.008 | site | 0.009 | state | 0.007 |
| phone | 0.008 | information | 0.008 | about | 0.007 |
| about | 0.008 | about | 0.007 | one | 0.007 |
| don’t | 0.007 | student | 0.007 | people | 0.007 |
| say | 0.007 | relationship | 0.007 | report | 0.006 |
| friend | 0.007 | these | 0.007 | site | 0.006 |
| when | 0.007 | state | 0.007 | time | 0.006 |
| good | 0.007 | connect | 0.007 | victim | 0.005 |
| improve | 0.007 | around | 0.007 | activity | 0.005 |

The words highlighted in yellow are the words that had been identified through the co-occurrence of terms in the DSM-5 and KSADS as potential indicators of negative sentiment or risky behavior. From looking at only the words in the three topic structures, it appears that topic

3 seems to represent a more negative sentiment regarding social media. Topics 2 appears to represent a more positive sentiment towards social media, with words like “friend”, “help”, “connect” and “improve. Topic 1 also has several positive words like “friend”, “good” and “improve.” It is unclear how to classify the sentiment of Topic 1 by looking at only the 30 most used words. Topic 1 could represent a more mixed or neutral sentiment but more context is required to make topic classifications. To get a deeper understanding into the different topics, the 3 essays that used the highest proportion of each topic were read.

The 3 essays that used the highest proportion of each topic are given in the table in Appendix B. Looking at the essays that used the highest proportion of topic 3, they all appear to have the negative sentiment that social media has had a negative impact on teens. The three essays that used the highest proportion of topic 3 all stated that social media has been harmful for teens. The 25 essays that used the highest proportion of topic 3 all conveyed negative sentiment towards social media. The essays that used the highest proportion of Topic 2 appear to display more of a positive sentiment towards social media. The top 25 essays that used the highest proportion of topic 2 all convey a strong positive sentiment towards social media. The essays that used the highest proportion of topic 1 have mixed results. Of the top 25 essays that used the highest proportion of topic 1, 2 essays had mixed sentiment, 7 had positive sentiment and 16 had negative sentiment.

Research Question: Can topic models detect risky behavior-related negative sentiment in students’ written answers to CR questions?

Results suggest that topic modeling using LDA appears to be able to detect negative sentiment through students’ written answers to CR items. This is reflected in the capability of the topic structure to include key words that can help classify topics as reflecting negative or positive

sentiment. The additional reading of essays is required, however, to help more accurately classify topics, as the words reflected in the topic structure may not represent the entire sentiment.

It is not clear, however, whether these are indicators of disorders. For example, the essay below uses the highest proportion of the top 30 words in topic 3 (classified as the negative sentiment topic), included the words “bullying”, “harm”, “neglect” and “theft”. These are words that could be flagged for potential indicators of social anxiety disorder or disruptive and impulse control disorder based on the DSM-5 and KSADS criteria. The essay this student wrote, which is displayed below, obviously expresses negative sentiment towards social media. The essay does not appear to indicate signs of disorders and is just addressing the prompt. The words that are highlighted in red below are words that exist in the sentiment association lexicon that was created for this pilot study (see Appendix D). These are words that are identified as potential indicators for risky behavior, negative sentiment or disorders. In the essay below, there are 13 instances of these words or phrases occurring. So, this essay would be given a score of 13.

While the student used words that occurred in sentiment lexicon and their essay displayed negative sentiment, it would not make sense to flag this student for negative sentiment indicative of risky behavior. This student was using appropriate language from the materials given to address the prompt and does not display any signs that would worry a human reader.

| Essay with the Highest Proportion of Topic 3 (Negative Usage) |
| --- |
| *Social media has massively impacted teenager's lives. Has this improved it, though? Social media has, in fact, not improved teen's lives, if anything, it has made them worse. From less face-to-face activity, cyber bullying, and even identity theft, social media has very much negatively effected teenager's lives. These are just a few of the surplus reasons why the media is not an improvement. The first reason why social media is not an improvement for teenager's everyday live's is that they do not experience much social interaction. According to Passage A, \34% agree that less face-to-face interaction occurs...\".This shows that, with the increased use of social media, real interaction has massively declined. Because of this, many teens struggle with their communication skills. Social media has had a direct impact on face-to-face activity within teens which is, indeed, a negative one. The next reason is that many teens are bullied through a social media platform. It states in Passage A that, \"49% report* |

| |
|---|
| *being a victim of cyber bullying\". This shows that almost half of social media users are bullied. Now, not only has social media impacted teenager's social health, but also effects the mental health of teenage users dealing with cyber bullying. All-in-all, social media has done more harm than good if such a high percentage struggle with being bullied. The final reason to why social media has had a negative impact is that teen's are at a high risk of identity theft. According to Passage B, \"...children under the age of 18 were twice as often victims of identity theft as their parents. Teens are often less cautious about protecting their personal information. Teens often share personal information via sites such as Facebook, Twitter, and Instagram\". This shows that teenager's neglect keeping their information hidden from certain people, which can often lead to identity theft. Because of social media, they are at a much higher risk for this. In conclusion, social media has not improved teenager's lives. It may allow connections and friendships to be made, along with a slight increase in academic performance, but compared to all the negative effects, it cannot be counted as an improvement. Social media has come with lessened social activity within teens, cyber bullying, and an increase in identity thefts. All these negative impacts seem to out weigh the few positive ones. Overall, social media has not improved teenager's lives."* |

Figure 1. Essay with the Highest Proportion of Topic 3 (Negative Usage). The topic proportion of this essay was broken as 98% topic 3 (negative), 0.5% topic 2 (positive) and 1.5% topic 1 (mixed).

Other essays were analyzed for the total number of times that they used words from the sentiment lexicon. It was interesting to see that the essay below used the largest proportion of topic 1. This essay, however, uses stronger, more potent, negative language than the essay above, that uses the largest proportion of topic 3. (See Appendix E for a table comparing the different negative words used in the topic 1 essay versus the topic 3 essay). In terms of sentiment analysis score, this student used indicator words 28 times. This student used over twice as many indicator words than the student from the prior essay and more indicator words than the 5 essays that use the highest proportion of topic 3. Also, due to the nature of the strong potent negative language (highlighted in red below) and for using words such as “bullying”, “kill”, “stab”, and “hang”, words that occur in the DSM-5/KSADS criteria, this student’s essay was flagged for displaying negative sentiment that might be indicative of risky behavior.

| Essay with the Highest Proportion of Topic 1 (Mixed Usage) |
|---|
| *I say social media doesn't make anybodies life better cause if you look at it now you will see that people are saying bad things about each other in there inbox or on Facebook or posting bad videos of people on YouTube. there are even people leaving bad comments saying bad things about people. which really stirs up a lot of drama between people even best friends.* |

*yeah a lot of people will say there making friends and yeah it can give you some valuable information but not everybody is telling the truth about social media. then there cyber bullying its mean and cruel to talk about somebody online i mean it wouldn't make it better if they said it to there face but at least you would have the guts to do it and not say it behind some code name on Facebook. If you could see somebodies inbox you would see some cruel and nasty things in there. they stir up a lot of drama and they start fight. they can even break up friendships. yeah it can sometimes help make friends but you need to watch out cause some people want get close to you just to try and ruin your life.also people like to post ugly pictures of other people or they like to comment on your picture saying you ugly. there is also people on Instagram just like the people on Facebook and they are either better or worse. YouTube there are a lot of haters or people that make bad videos about people. I mean pranks are funny but there are cruel to the person that is getting pranked it might be funny to the other person, but the person that got pranked could be hurt. some people write bad things about other peoples channel but you just got to ignore them people and not let them get in your head and keep making videos. and a lot of people will be your friends just to get some dirt on you or to stab you in the back. they might comment some bad things about you like you suck or delete your YouTube channel nobody watches you.All those things i said before lead up to this cyber bullying. It can get so bad that people kill themselves every year do to some kind of bullying. like it is out outrageous if somebody bullies you to the extent of you killing yourself. like you know its bad if they hang or stab maybe even shoot themselves. like i said in those other paragraphs they will treat you like dirt post mean pictures online about you post mean videos and if you let them do it you* aren't standing up to yourself. and if you let them know they get to you then they wont stop and if they don't stop then delete your account.

Figure 2. Essay with the Highest Proportion of Topic 1 (Mixed Usage). The topic proportion of this essay was broken as 99.4% topic 1 (mixed), 0.3% topic 2 (positive) and 0.3% topic 3 (negative).

It is interesting that an essay conveying strong potent negative sentiment did not use the highest proportion of topic 3. A potential reason for that could be essays that take a negative stance that directly address the prompt might use the highest proportion of topic 3 while essays that contain general negative sentiment that might not as directly address the prompt may use a higher proportion of topic 1. The 30 most used words in topic 3 include words such as “network”, “identity”, and “information” address the prompt. The top 30 words for topic 1 contain words that are neutral and do not directly address the prompt. There is also the potential influence of the potency of the words used in the essays. Using words such as “stab”, “hang” and “kill” have a higher potency meaning than the words, “negatively effected” , “victim” and “harm”.

Semantic analysis

Additional sentiment analysis was then conducted using the *sentimentr* package. Results indicate there were non-zero correlations between two of the topics and the sentiment analysis score. The two topics that displayed a correlation with the sentiment score were the two topics that demonstrated a strong predominant sentiment. The 3 topics in the structure are classified as topic 1: mixed sentiment, topic 2: positive sentiment and topic 3: negative sentiment. The negative sentiment analysis scores are moderately correlated with the topic 3: negative sentiment proportions (-0.48, $p<.05$). The positive sentiment analysis scores are strongly correlated with the topic 2: positive sentiment proportions (0.64, $p<.05$). The mixed sentiment analysis scores were moderately correlated with the topic 1: mixed sentiment proportions (0.23, $p<.05$).

In addition, the essays that had the strongest negative sentiment analysis scores were read and examined. Below are the essays with the strongest negative sentiment. The essays are shorter than the top 5 essays that use the highest proportions of topics 1, 2, or 3. It is interesting to note that these essays use more potent negative language than the essays that use the highest proportion of topic 3. For an example, it is useful to look at essay #2, below. The student is using strong, explicit, negative language and not addressing the prompt at all.

Table 2. Essays with the Strongest Negative Sentiment Analysis Score

| | Essay | Sentiment Score | Topic 1 Usage | Topic 2 Usage | Topic 3 Usage | Sentiment Lexicon Score |
|---|---|---|---|---|---|---|
| 1 | social networking activities have not improved teens life because they are just sitting in their room or they have there face in there phone and they are just texting and they have no face to face convo so therefore they will most likely become antisocial because if it makes them mad and they don't do anything about it it all | -0.7605 | 0.8553 | 0.1316 | 0.0132 | 10 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | bottles up and they for they will take it out on thire family or they will take it out on their friends and that would start a fight they could stat to pick on someone and really upset someone and cause them to harm or worst to them self and harassment they could bother someone all day or sexaully harass them. | | | | | |
| 2 | I'mjust so vary sick and tired of this school they all need to go and to get a life of they own bc all this is just bull shit this is my mother fucking argument with all this shit im sooooooooooooooooooooooooooo fucking done with all this bull shit here at villa rica high school fuck this school | -0.69933 | 0.6875 | 0.2500 | 0.0625 | 2 |
| 3 | the internet has its negative effects and it positive effects because the teenagers can look questions up on the internet for school and things like that will help them on school work and then again it has negative effects because when some bad students get on the internet bullying can happen or they can scam peoples bank accounts or even hack into other peoples profiles on any social media app and then some people make videos that gets them into alot of trouble by the police and even can go to jail for a long time and bullying that happens on the internet is called cyber bullying and sometimes people commit suicide because of other people bullying them | -0.66919 | 0.6923 | 0.2821 | 0.0256 | 7 |
| 4 | i think that networking activities have inproved teenagers lives because most of the time they are inside with they're technology insted of being outside getting in trouble and getting accocieated with illegal stuff, but in the oher hand networking activities have ruined a | -0.66863 | 0.6875 | 0.2500 | 0.0625 | 4 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | lot of lifes too by stealing they're identity. | | | | | |
| 5 | i think social network is not really safe for teenagers because most people get bully,get there identity steal alot of people has fake to be someone on social media ,alot of people get hurt. | -0.63387 | 0.8148 | 0.0741 | 0.1111 | 4 |
| 6 | No because things that teenagers are doing are getting them kidnapped, abused and it causes suicide. | -0.60634 | 0.6667 | 0.2500 | 0.0833 | 3 |

The student who wrote essay 2 in the table above would not have triggered the flags from the sentiment lexicon score or from using the largest proportion of Topic 1 or Topic 3. However, when the sentiment score was added to the analysis, the essay was flagged due to having the second highest negative sentiment analysis score.

It is interesting to note that essays that have the strongest negative sentiment scores also have the largest proportion of use of topic 1 and not topic 3. It appears that topic 3 may detect essays that contain negative sentiment towards the social media prompt while topic 1 may be able to detect essays that use strong potent negative sentiment or essays that do not address the social media prompt. The essays in Table 3 follow the pattern of using stronger language or use explicit language and ignore the prompt completely.

The prompt requires students to take a stance on social media and it should be expected that some students will take a negative stance. However, there is a difference between writing “social media is bad because it has a negative impact on my social health” and “social media is bad because it makes people want to stab, shoot and kill themselves”. “Social media is bad because it has a negative impact on my social health” addresses the prompt while taking a negative stance. “Social media is bad because it makes people want to stab, shoot and kill themselves” still addresses the prompt but uses stronger negative language, like “stab”, “shoot”

or “kill”. The second statement uses words that depict sentiment of higher activity and potency than the first statement.

Human raters

The results from the individual raters matched the topic modeling and sentiment analysis scores 100% of the time for the 2 topics that could be clearly classified as positive and negative. There was a 60% match rate for rater one for the first topic of mixed sentiment and a 50% match for rater two.

The major takeaways from the study results are:

(1) LDA topic modeling does appear to be capable of detecting some negative sentiment in students’ responses to a constructed response item. There was a specific individual topic, topic 3, that was predominantly associated with essays that strictly display negative sentiment towards the essay prompt.

(2) In terms of identifying indicators of risky behavior or disorders, topic modeling was able to identify indicators of risky behavior or disorders as well as essays that should be flagged based on the indicators. In Figure 1, the student used words that could be flagged as potential indicators of major depression or DMDD (based on the DSM-5 and KSADS criteria). This student had a sentiment lexicon score of 13 which was higher than the scores of all the essays in Table 2. (It is important to note the length of the essay response played a factor in the sentiment lexicon score count for the response in Table 2 in this instance.) This student would not be flagged for risky behavior.

(3) There are a couple of essays from the study that might be flagged for indicators of potential risky behavior. It is evident, however, that such essays occur with relatively low frequency. This result is consistent with findings from Ophir et al. (2020). A couple of essays

may not be extremely problematic on their own, especially given the context of the prompt, but flagging these essays for potential indicators of future risky behavior, disorder or negative sentiment can provide one more safeguard for students.

(4) The results from the study indicate topic modeling has the potential and capability to differentiate between the type and potency of negative sentiment. Topic modeling appears to be able to detect essays that weak-potent negative sentiment directed towards social media and displays this in the form of Topic 3 in the model. Essays that displayed strong potent and active negative sentiment used the largest proportion of Topic 1. Essays that used the highest proportion of Topic 3 did not use the strong potent negative sentiment that is seen in Topic 1.

(5) The study also highlights the importance of human raters in validating and reading essays. The topics appear to be able to differentiate between negative language specifically addressing the social media prompt and negative language that does not address the prompt or negative language that is more explicit. Human rater confirmation is needed to confirm which essays were appropriately flagged for concerning indicators and which essays simply address the prompt.

(6) When used together, LDA and sentiment analysis can provide more information than when used as two separate analyses. If only the topic structures were looked at, it would be harder to classify the topics that show mixed sentiment. If only sentiment analysis was examined, the results may not be meaningful since polarity is heavily skewed if there is only one sentence or a few words. The two analyses together create an additional way to view the essays and identify which essays should be flagged.

# CHAPTER 5

# DISCUSSION

Overall, these results indicate that a combination of topic modeling, sentiment analysis and individual rater score can identify positive and negative sentiment from individual essays. These results also show that the topic structure may be capable of differentiating between the subject of negative sentiment as well as negative sentiment potency.

The results indicated that it is possible to detect negative sentiment using LDA. This can be seen in the essays that used the highest proportion of the topics that are classified as having a high preponderance of negative sentiment. Essays that displayed potential indicators of risky behavior or disorders did not occur with great frequency, but two instances were detected. These findings align with the literature (Ophir et al., 2020) and confirm that it is very uncommon but possible to detect these potential indicators. Detecting indicators of social disorders in is rare and there was not an expectation to identify many essays for risky behavior. The identification of two instances that required flagging are considered a success in this study as this is students that can be potentially helped.

Topic 1, the “mixed” sentiment topic, poses interesting results. This topic had essays with mixed sentiment could arguably be said to have been correctly classified. There were essays that did not take a definitive stance on the social media argument and that reflected in how the sentiment analysis and the human raters scored the essay. The sentiments for some of the essays were mixed/neutral. It is interesting that essays that used the highest proportion of this topic contained both strongly negative sentiment and strongly positive sentiment. This is also the topic

that contained the most flagged essays that this study was hoping to identify. These results prompt questions about why the mixed topic, Topic 1, is able to pick up on the most essays that have indicators of social disorders, but Topic 3, the negative sentiment topic, is not able to do this?

It could be argued that Topic 1 was detecting essays with strong potency and ignoring the polarity of the essays. The essays that used the highest proportion Topic 1 used words that took more active stances than essays that used words from the prompt material which were more passive. For example, the prompt located in Appendix C talks about the negative impact of social media using the phrases “victim of cyberbullying” or “negatively affect grades”. These are low potency and more passive words that show up in the most used words in Topic 3. These phrases also appear in essays that use the highest proportion of Topic 3. The literature has shown that students are taught to draw on prompt material to build their argument and topic models are able to detect when students use this kind of writing strategy (Duong et al., 2019). The prompt never uses strong, potent, and active words like “hang”, “kill”, or “stab”. The topic model could also be separating the negative sentiment by two layers: (1) potency and activity of negative language and (2) whether the negative sentiment in the essay draws upon the prompt material and addresses the prompt.

Additionally, the results prompt questions about adding more flagging words to the sentiment lexicon. Adding curse words or explicit language might be helpful in picking up essays that depict indicators of social disorders. Curse words are not words that are commonly seen in essays and students using those words may be doing it for emphasis, jokingly or in a negative context.

When used together, topic modeling and sentiment analysis have the capability of casting a wider net to detect essays with social disorder indicators. Adding extra layers and flags to model to detect negative sentiment indicative risky behavior can provide more support to students. Topic modeling interpretation can also strongly benefit with the addition of a sentiment analysis score to increase interpretability and topic labeling.

Sentiment analysis and human rater results further confirmed that LDA was able to successfully separate predominately negative and positive essays into different topics. The results support the literature in the requirement to have human raters as the last layer of validation. Grimmer and Stewart (2013) emphasize the importance of recognizing the results from automated algorithms will always be context specific and will always require human validation of the measures produced by machine learning techniques and making sure the measures correspond with the concept claimed in unsupervised classification.

Grimmer and Stewart (2013) argue that successful validation of a model would be demonstrating that the classification from machines replicates human hand coding and that is what the human raters were able to do in the study for two out of the three topic models. For the Topic 1, where there was not a complete agreement, the sentiment of the essays could be classified as mixed which further supports the need for human raters. The essays that used the highest proportion of Topic 1, the mixed sentiment topic, had mixed sentiments. It makes sense that the human raters would not completely agree on the sentiment classification as the topic modeling algorithm could not either. Writing contains nuances and can be interpreted in different lights depending on the audience (Hunter & Docherty, 2009).

Fortgang et al. (2021) found that feelings of social isolations can be predictive of suicidal thoughts and there has been an increase in feelings of social isolation since the onset of COVID-

19. Research has shown there is a need for more outreach and treatment to address this (Fortgang et al., 2021). Using topic modeling as a potential screening tool can be a method of outreach in schools.

While the results from the study did not identify any essays that specifically displayed signs of suicidal thinking, this was expected as texts that display this kind of thought are uncommon (Fortgang et al., 2021; Ophir et al., 2020). However, the results from the study show that it is possible able to identify and flag essays with strong negative sentiment using topic modeling and sentiment analysis.

The results of the study show that topic modeling can detect different themes of negative sentiment in students' writing and has the potential to be used as a screening tool with the right structures around it. In the context of this study, the structures that helped topic modeling provide better results were sentiment analysis and human raters.

Future research. Future research on using topic modeling as a screening tool would need to focus on threshold recommendations for sentiment analysis and topic modeling proportion usage in students' essays. Specifically, how much negative sentiment would trigger a flag that would require human raters to read the essay to confirm the topic modeling results?

It might also be useful to focus future topic modeling research on essays that require students to take a stance on something that is significant to them. An example of an essay prompt could be the increasing "eco-anxiety" about climate change and its effect on mental health since that has been a growing concern with Gen Z (Reyes et al., 2021; Ross & Rouse, 2022). This could be the type of essay prompt that elicits strong negative or positive sentiment about their future and is something that could affect them directly.

Future Research

Finally, future research on how indicators of specific mental illness and disorders display themselves in text could make topic model interpretation more meaningful and robust. There is currently not specific research on the presentation of mental disorders through text and prior research shows that different mental disorders present different indicators. The lack of research on this topic was the reason for the creation of the negative sentiment lexicon. Therefore, further work is needed to explore the best way to detect negative sentiment, risky behavior and mental illness and disorder indicators in text analysis.

# REFERENCES

Abels, P., & Verduyn, C. (2009). Computerized CBT for Adolescent Depression ("Stressbusters") and its Initial Evaluation Through an Extended Case Series. *Behavioral and Cognitive Psychotherapy, 37.*

Adedoyin, O. B., & Soykan, E. (2020). Covid-19 pandemic and online learning: the challenges and opportunities. *Interactive Learning Environment.* https://doi.org/10.1080/10494820.2020.1813180.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, (6), 716-723.

Allison, V.L., Nativio, D.G., & Mitchell, A.M. (2013). Identifying Symptoms of Depression and Anxiety in Students in the School Setting. *The Journal of School Nursing, 30(*3), 165-172. https://doi.org/10.1177%2F1059840513500076

American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Anderson, D., Rowley, B., Stegenga, S., Irvin, P. S., & Rosenberg, J. M. (2020). Evaluating content-related validity evidence using a text-based machine learning procedure. *Educational Measurement: Issues and Practice, 39,* 53–64.

Bae, Y., & Lee, H. (2012). Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitters. *Journal of the American Society for Information Science and Technology, 63*(12).

Beck, A.T. (1967). *Depression: Clinical, Experimental and Theoretical Aspects.* Harper & Row.

Blei, D.M. (2012). Probabilities topic models. *Communications of the ACM, 55*(4), 77-84.

Bonifacci, P., Candria, L., & Contento, S. (2008). Reading and writing: what is the relationship with anxiety and depression. *Reading and Writing, 21*. https://doi.org/10.1007/s11145-007-9078-6.

Boyd-Graber, J., Mimno, D., & Newman, D. (2014). Care and feeding of topic models: Problems, diagnostics, and improvements. In E. M. Airoldi, D. Blei, E. A. Erosheva, & S. E. Fienberg (Eds.) *Handbook of mixed membership models and their applications* (pp. 225-254). Boca Raton, FL: CRC Press.

Bougie, J. R. G., Pieters, R., & Zeelenberg, M. (2003). Angry customers don't come back, they get back: The experience and behavioral implications of anger and dissatisfaction in services. Open access publications from tilburg university, Tilburg University

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer.

Callaghan, L. (2012). *An evaluation of latent Dirichlet allocation in the context of plant-pollinator networks*. [Unpublished master's thesis]. University of Guelph.

Cao, J., Xia, T., Li, J., Zhang, Y., & Tang, S. (2009). A density-based method for adaptive LDA model selection. *Neurocomputing, 72*, 1775–1781.

Cao, W., Fang, Z., Hou, G., Han, M., Xu, X., Dong, J., & Zheng, J. (2020). The psychological impact of the COVID-19 epidemic on college students in China. *Psychiatry Research, 287*.

Cardozo-Gaibisso, L., Kim, S., Buxton, C.A., & Cohen, A.S. (2019). Thinking beyond the score: Multidimensional analysis of student performance to inform the next generation of science assessments. *Journal of Research in Science Teaching, 57*, 856-878.

Chaplot, D., Rhim, E., & Kim, J. (June 2015). Predicting Student Attrition in MOOCs using Sentiment Analysis and Neural Networks. *Presented at the 17th International Conference on Artificial Intelligence in Education.*

Chen, Y., Yu, B., Zhang, X., & Yu, Y. (2016, April). Topic modeling for evaluating students' reflective writing: A case study of pre-service teachers' journals. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge,* ACM, Edinburgh, UK, 1-5.

Cherry, C., Mohammad, S.M., & De Brujin, B. (2012). Binary classifiers and latent sequence models for emotion detection in suicide notes. *Biomedical Informatics Insights, 5*(Suppl. 1), 147.

Cohen, A. S., & Cho, S.-J. (2017). Information criteria. In W. Vander Linden (Ed.), *Handbook of item response theory, volume 2: Statistical tools* (pp. 363–378). Monterey, California: CRC Press.

Deveaud, R., SanJuan, E., & Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document Numérique, 17*, 61–84.

Das, S.R., Kim, S., & Kothari, B. (2017). Zero-Revelation RegTech: Detecting Risk through Linguistic Analysis of Corporate Emails and News. *SSRN Electronic Journal*. https://ssrn.com/abstract=2960350.

DeSocio, J. & Hootman, J. (2004). Children's mental health and school success. *Journal of School Nursing, 20,* 189-196.

Dooey, P. (2008). Language testing and technology: problems of transition to a new era. *The Journal of EUROCALL, 20*(1), 21-34.

Drain, T. S., Grier, L.E., & Sun, W. (2012). Is the Growing Use of Electronic Devices Beneficial to Academic Performance? Results from Archival Data and a Survey. *Issues in Information Systems, 13*(1), 225-231.

Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology, 38*(1), 188-230.

Duong, E., Cohen, A., Mellom, P., Hixon, R., & Weber, J. (2019, April). *Examining the Impact of the Instructional Conversation on Students' Writing through Topic Modeling Analysis.* Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada.

Ellis, W.E., Dumas, T.M., & Forbes, L.M. (2020). Physically isolated but socially connected: Psychological adjustment and stress among adolescents during the initial COVID-19 crisis. *Canadian Journal of Behavioural Science/Revue Canadienne des sciences du comportement, 52*(3), 177-187. https://psycnet.apa.org/doi/10.1037/cbs0000215.

Elmer, T., Mepham, K., & Stadtfeld, C. (2020). Students under lockdown: Comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland. *PLoS ONE, 15*(7),

Fivush, R., Marin, K., Crawford, M., Reynolds, M., & Brewin, C.R. (2007). Children's narratives and well-being. *Cognition and Emotion, 21*(7), 1414-1434.

Foltz, Peter W, Walter Kintsch, and Thomas K Landauer. 1998. "The Measurement of Textual Coherence with Latent Semantic Analysis." *Discourse Processes* 25 (2-3): 285–307.

Fortgang, R.G., Wang, S.B., Millner, A.J., Reid-Russell, A., Beukenhorst, A.L., Kleiman, E.M., Bentley, K.H., Zuromski, K.L., Al-Suwaidi, M., Bird, S.A., Buonopane, R., DeMarco, D., Haim, A., Joyce V.W., Kastman, E.K., Kilbury, E., Lee, H-I. S., Patrick Mair1, Nash,

C.C., Onnela, J-P., Smoller, J.W., & Nock, M.K. (2021). *Increase in Suicidal Thinking During COVID-19*. Clinical Psychological Science, 9*(3),* 482-488.

Frayne, A., & Wade, T.D. (2006). A comparison of written emotional expression and planning with respect to bulimic symptoms and associated psychopathology. *European Eating Disorders Review*, *14*(5), 329-340.

Geiger, D., & Heckerman, D. (1997). A characterization of the Dirichlet distribution through global and local parameter independence. *The Annals of Statistics, 25*, 1344–1369.

Greenberg, L.S., & Watson, J.C. (2006). *Emotion-focused Therapy for Depression.* American Psychological Association.

Grimmer, J., & Stewart, B.M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *18*(1), 1-35.

Grun, B., & Hornik, K. (2011). topicmodels: an R Package for Fitting Topic Models. *Journal of statistical software*, *40*(13).

Hara, N., & Kling, R. (2003). Student distress in a web-based distance education course: an ethnographic study of participants' experience. *Online Journal of Distance Education, 4*(2).

Hassan, A., Abu-Jbara, A., & Radev, D. (2012). Extracting signed social networks from text. In Workshop Proceedings of TextGraphs-7 on Graph-based Methods for Natural Language Processing, pp. 6–14.

Havigerova, J.M., Haviger, J., Kucera, D., & Hoffmannova, P. (2019). Text-Based Detection of the Risk of Depression. *Front Psychol*, *10(*513). https://dx.doi.org/10.3389%2Ffpsyg.2019.00513.

Hendron, M. & Kearney, C.A. (2016). School Climate and Student Absenteeism and Internalizing and Externalizing Behavioral Problems. *Children and Schools, 38*(2), 109-116.

Hornik, K., & Grün, B. (2011). topicmodels: An R package for fitting topic models. *Journal of statistical software*, *40*(13), 1-30.

Hunter, K & Docherty, P. (2009). Reducing variation in the assessment of student writing. *Assessment & Evaluation in Higher Education,* 36*(1),* 109-124.

Jeanne, M., Gross, J.J., Persons, J.B., & Hahn, J. (1998). Mood Matters: Negative Mood Induction Activates Dysfunctional Attitudes in Women Vulnerable to Depression. *Cognitive Therapy and Research, 22,* 363-376.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools Applications,* 78, 15169–15211.

Jockers, M. (2017). Package 'syuzhet'. https://github.com/mjockers/syuzhet.

Kaufman, J., & Schweder, A. E. (2004). The Schedule for Affective Disorders and Schizophrenia for School-Age Children: Present and Lifetime version (K-SADS-PL). In M. J. Hilsenroth & D. L. Segal (Eds.), Comprehensive handbook of psychological assessment, Vol. 2. Personality assessment (pp. 247–255). John Wiley & Sons, Inc..

Kennedy, B.L. (1994). The Role of Topic and the Reading/Writing Connection. *TESL-EJ,* 1*(1).*

Kim, S., Kwak, M., Cardozo-Gaibisso, L.A., Buxton, C.A., & Cohen, A.S. (2017). Statistical and qualitative analyses of students' answers to a constructed response test of science inquiry knowledge. *Journal of Writing Analytics, 1,* 82-102.

Kim, S.-Y. (2014). Determining the number of latent classes in single-and multiphase growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal, 21*, 263–279.

Kirelli, Y. & Arslankaya, S. (2020). Sentiment Analysis of Shared Tweets on Global Warming on Twitter with Data Mining Methods: A Case Study on Turkish Language. https://pubmed.ncbi.nlm.nih.gov/32963511.

McLoone, J., Hudson, J.L., & Rapee, R.M. (2006). Treating Anxiety Disorders in a School Setting. *Education and Treatment of Children, 29*(2), 219-242.

Landauer, Thomas K, Peter W Foltz, and Darrell Laham. 1998. "An Introduction to Latent Semantic Analysis." *Discourse Processes* 25 (2-3): 259–84.

Li, X. G., Li, S. M., Jiang, L. R., & Zhang, S. B. (2014). Study of english pronunciation quality evaluation system with tone and emotion analysis capabilities. Applied Mechanics and Materials, 475, 318–323.

Liu, B. (2010). Sentiment Analysis: A Multi-Faceted Problem. *IEEE Intelligent Systems, 25*.

Loades, M.E., Chatburn, E., Higson-Sweeny, N., Reynolds, S., Shafran, R., Brigden, A., Linney, C., McManus, M.N., Borwick, C., & Crawley, E. (2012). Rapid Systemic Review: The Impact of Social Isolation and Loneliness on the Mental Health of Children and Adolescents in the Context of COVID-19. *Journal of the American Academy of Child & Adolescent Psychiatry, 60(*1), 6-7.

Lynch, S.M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists.* New York: Springer.

Mardones-Segovia, C., Kim, H-K., Hong, M., Wheeler, J., & Cohen, A. (2022). *Comparison of Estimation Algorithms for Latent Dirichlet Allocation*. Quantitative Psychology (pp. 27-37). https://doi.org/10.1007/978-3-031-04572-1_3.

Mardones, C., Wheeler, J.M., Cohen, A.S., & Choi, H.-J. (2020, in preparation). *Model selectin for latent Dirichlet allocation with short answers or small numbers of topics.*

Mohammad, S. (2012). Portable features for classifying emotional text. In Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies (pp. 587–591), Montréal, Canada.

Mohammad, S. (2015). Challenges in Sentiment Analysis. *A Practical Guide to Sentiment Analysis*. http://www.saifmohammad.com/WebDocs/sentiment-challenges.pdf.

Mohammad, S. (2016). 9 – Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text. *Emotion Measurement.* 201-237. https://doi.org/10.1016/B978-0-08-100508-8.00009-6.

Mohammad, S. & Yang, T. (2011). Tracking Sentiment in Mail: How Genders Differ on Emotional Axes. *In Proceedings of the ACL 2011 Workshop on ACL 2011 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), June 2011, Portland, OR*. https://arxiv.org/abs/1309.6347.

Mouthami, K., Devi, K.N., & Bhaskaran, V.M. (2013). *Sentiment analysis and classification based on textual reviews. Proceedings of the 2013 International Conference on Information Communication and Embedded Systems (ICICES)*, Chennai, India.

Muris, P., Meesters, C., van Melick, M., & Zwambag, L. (2001). Self-reported attachment style, attachment quality, and symptoms of anxiety and depression in young adolescents. *Personality and Individual Differences, 30*(5), 809-818.

Nahar, V., Unankard, S., Li, X., & Pang, C. (2012). Sentiment analysis for effective detection of cyber bullying. In Web Technologies and Applications, pp. 767–774. Springer.

Neviarouskaya, A., Prendinger, J., & Ishizuka, M. (2009). Compositionality principle in recognition of fine-grained emotions from text. In Proceedings of the proceedings of the third international conference on weblogs and social media (ICWSM-09) (pp. 278–281), San Jose, CA.

Noble, K., Hurley, P., & Macklin, S. (2020). COVID-19, employment stress and student vulnerability in Australia. Mitchell Institute for Education and Health Policy, Victoria University.

Nobles, A.L., Glen, J.J., Kowsari, K., Teachman, B.A., & Barnes, L.E. (April 2018). Identification of Imminent Suicide Risk Among Young Adults using Text Messages. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, April 2018, Chicago, IL.*

Nock, M.K. (2008). Actions speak louder than words: An elaborated theoretical model of the social functions of self-injury and other harmful behaviors. *Applied and Preventative Psychology, 12*(4), 159-168.

Nunez, G.H. (2020). Using Text Mining to Gauge Student Sentiments about Hands-on Activities in a Protected Agriculture Course. *HortTechnology, 30*(1), 129-135.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

Ophir, Y., Rikochinski, R., Asterhan, C.S.C., Sisso, I., & Reichart, R. (2020). Deep neural network detect suicide risk from textual facebook posts. *Scientific Reports, 10*.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. (1957). The measurement of meaning. University of Illinois Press.

Paul, MJ., & Dredzem M. (2011). You are what you Tweet: Analyzing Twitter for public health. *ICWSM, 20,* 265-272.

Pennebaker, J.W., Mehl, M.R., & Niederhoffer, K.G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54(1),* 547-577.

Polce-Lynch, M., Myers, B.J., Kilmartin, C.T., Forssmann-Falck, R., & Kliewer, W. (1998). Gender and Age Patterns in Emotional Expression, Body Image, and Self-Esteem: A Qualitative Analysis. *Sex Roles, 38*.

Ponweiser, M. (2012). Latent Dirichlet allocation in r (Doctoral dissertation). Retrieved from https://epub.wu.ac.at/3558/1/main.pdf

Quinn, K.M., Monroe, B.L., Colaresi, M., Crespon, M.H., & Radev, D.R. (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, *54*(1), 209-228.

Raftery, A. (1995). Bayesian model selection in model research. *Sociological Methodology, 25,* 111-163.

Ramesh, A., Goldwasser, D., Huang, B., Daume III, H., & Getoor, L. (2014). Understanding MOOC discussion forums using seeded LDA. In *Proceedings of 9th Workshop on Innovative Use of NLP for Building Educational Applications*. Baltimore, MD: 28-33.

Reich, J., Stewart, B., Mavon, K., & Tingley, D. (2016). *The civic mission of MOOCA: Measuring engagement across political differences in forums*. Paper presented at the Third ACM Conference on Learning, Edinburgh, UK. Retrieved from https://doi.org/10.1145/2876034.2876045.

Reyes, M.E.S., Carmen, B.P.B., Luminarias, M.E.P. (2021). An investigation into the relationship between climate change anxiety and mental health among Gen Z Filipinos. *Current Psychology*. https://doi.org/10.1007/s12144-021-02099-3

Ross, A.D., Rouse, S.M. (2022). (Young) Generations as Social Identities: The Role of Latino*Millennial/Generation Z in Shaping Attitudes About Climate Change. *Political Behavior*, 44, 1105–1124. https://doi.org/10.1007/s11109-020-09649-8

Roque, C., Cardoso, J. L., Connell, T., Schermers, G., & Weber, R. (2019). Topic analysis of road safety inspections using latent Dirichlet allocation: A case study of roadside safety in Irish main roads. *Accident Analysis & Prevention, 131*, 336–349.

Rude, S., Gortner, E. -M., & Pennebaker, J. (2004). Language use of depressed and depression vulnerable college students. *Cognition & Emotion*, 18(8), 1121–1133.

Rude, S., Gortner, E-M., & Pennebaker, J. (2010). Language use of depressed and depression-vulnerable college students. *Cognition and Emotion, 8*¸ 1121-1133. https://www.tandfonline.com/doi/abs/10.1080/02699930441000030?casa_token=PqMUrqyUfv4AAAAA:SxM9sj07JYV6zHcdA-ki5mVMbBxu49zi3RTzNh-7S1T7LcUmfagIkgh1EvH5tQ7cUawLaKNtdAPp.

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39(*6), 1161.

Runchin, V., Sukhodolsky, D., Vermeiren, R., Koposov, R., & Schwab-Stone, M. (2006). Depressive symptoms and associated psychopathology in urban adolescents: A cross-cultural study of three countries. *The Journal of nervous and Mental Disease, 194*, 106-113.

Schroder, N., Falke, A., Hruschka, H., & Reutterer, T. (2017). Analyzing browsing and purchasing across multiple websites based on latent Dirichlet allocation. *ALLDATA* 2017, 40, 44.

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6, 461-464.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*, 333–343.

Sen, S., & Bradshaw, L. (2017). Comparison of relative fit indices for diagnostic model selection. *Applied Psychological Measurement, 41*, 422–438.

Sentiment Analysis. (2020). In *Oxford Online Dictionary.* Retrieved January 16, 2022, from https://en.oxforddictionaries.com/definition/sentiment_analysis.

Shermis, M. D., & Burstein, J. (Eds.). (2013). *Handbook of automated essay evaluation: Current applications and new directions*. Routledge.

Shorter, E. (2013). The History of DSM. In: Paris J., Phillips J. (eds) Making the DSM-5. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-6504-1_1.

Silva, F.F., Jerez, E.A.Z., de Resende, M.D.V., Viana, J.M.S., Azevedo, C.F., & Lopes, P.S., Nascimento, M., de Lima, R.O., & Guimaraes, S.E.F. (2017). Bayesian model combining linkage and linkage disequilibrium analysis for low density-based genomic selection in animal breeding. *Journal of Applied Animal Research,46*(1), 873-878.

Sit, J.W.H., Chung, J.W.Y., Chow, M.C.M., & Wong, T.K.S. (2005). Experiences of online learning: students' perspective. *Nurse Education Today, 25*(2), 140-147.

Son, C., Hegde, S., Smith, A., Wang, X., & Sasangohar, F. (2020). Effects of COVID-19 on College Students' Mental Health in the United States: Interview Survey Study. *Journal of Medical Internet Research, 22*(9).

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & Ven Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4). https://doi.org/10.1111/1467-9868.00353.

Suero Montero, C., & Suhonen, J. (2014). Emotion analysis meets learning analytics: Online learner profiling beyond numerical data. In Proceedings of the 14th koli calling international conference on computing education research (pp. 165–169). ACM.

Syed, S., & Spruit, M. (2018). Selecting priors for latent Dirichlet allocation. In *2018 IEEE 12th international conference on semantic computing* (pp. 194–202).

Thelwall, M. (2017). Gender bias in sentiment analysis. *Emerald Insight*. https://www.emerald.com/insight/content/doi/10.1108/OIR-05-2017-0139/full/pdf?casa_token=_9xnbiMSK5cAAAAA:F5w7iZ4FxQjnxJ-pn94dlqBwj_ff0cwgcLQGIo7zyIeqBnsMG5OfqYSs5p8W-wrvSkUdh0IXsWoqGB3MQkH_n2UNMuFZI6-lUoxyXSpdTwRMIvNXWs4

Tucker, S.Y. (2012). Promoting Socialization in Distance Education. *Turkish Online Journal of Distance Education, 13*(1), 174-182.

Valenti, A.P., Chita-Tegmark, M., Tickle-Degnen, L., Bock, A.W., & Scheutz, M.J. (2019). Using topic modeling to infer the emotional state of people living with Parkinson's disease. *Assistive Technology,* DOI: 10.1080/10400435.2019.1623342.

Vayansky, I. & Kumar, S.A.P. (2020). A review of topic modeling methods. *Information Systems, 94*. https://doi.org/10.1016/j.is.2020.101582.

Ventirozos, F.K., Varlamis, I., & Tsataronis, G. (April 2017). Detecting Aggressive Behavior in Discussion Threads Using Text Mining. *Presented at the 18th International Conference on Computational Linguistics and Intelligent Text Processing*.

https://www.researchgate.net/publication/315743770_Detecting_Aggressive_Behavior_in_Discussion_Threads_Using_Text_Mining.

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods, 17*, 228–243.

Walker, S.P., Chang, S.M., Vera-Hernandez, M., Grantham-McGregor, S. (2011). Early Childhood Stimulation Benefits Adult Competence and Reduces Violent Behavior. *Pediatrics, 127*(5), 849-857.

Walsh, M. (2004). Markov Chain Monte Carlo and Gibbs Sampling. *Lecture Notes for EEB 581*, University of Arizona.

Wamboldt. (n.d.). Describing relationship patterns in DSM-5: A preliminary proposal. In *Contemporary directions in psychopathology: scientific foundations of the DSM-5 and ICD-11 /* (pp. 565–576). Guilford Press.

Wang, Y., Luo, J., Niemi, R., Li, Y., & Hu, T. (2016). Catching fire via" likes": Inferring topic preferences of Trump followers on twitter. *Proceedings of the International AAAI Conference on Web and Social Media, 10*(1), 719-722. https://doi.org/10.1609/icwsm.v10i1.14778.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 english lemmas. Behavior Research Methods, 45(*4*), 1191–1207.

Weist, M., Rubin, M., Moore, E., Adelsheim, S., & Wrobel, G. (2007). Mental health screening in schools. *Journal of School Health, 77,* 53-58.

Wheeler, J.M., Lee, J., Choi, H.-J., Xiong, J., & Cohen, A.S. (2020). A semi-confirmatory latent Dirichlet allocation topic model. Proposal submitted to the 2021 annual meeting of the National Council on Measurement in Education.

Widiger, T.A., & Clark, L.A. (2000). Toward DSM—V and the classification of psychopathology. *Psychological Bulletin, 126*(6), 946–963. https://doi.org/10.1037/0033-2909.126.6.946.

Xiang, R., Long, Y., Wan, M., Gu, J., Lu, Q., & Huang, C-R. (2020). *Affection Driven Neural Networks for Sentiment Analysis.* Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), 112–119.

Xiong, J. Choi, H-J., Kim, S., Kwak, M., & Cohen, A.S. (2020). *Topic Modeling of Constructed Response Answers on Social Study Assessments*. The Annual Meeting of the Psychometric Society.

Yamanishi, K., Wu, T., Sugawara, S., & Okada, M. (2019). The decomposed normalized maximum likelihood code-length criterion for selecting hierarchical latent variable models. *Data Minng and Knowledge Discovery, 33,* 1017-1058.

Yang, C.-C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics & Data Analysis, 50*, 1090–1104.

Yildirim, Ilker. (2012). Bayesian Inference: Gibbs Sampling. University of Rochester. http://www.mit.edu/~ilkery/papers/GibbsSampling.pdf.

# APPENDIX

## APPENDIX A: TABLE OF 9TH GRADE ARGUMENTATIVE STATISTICS FOR 2018-2019 SCHOOL YEAR

| **9th Grade Argumentative Statistics (2018-2019)** | | | | |
|---|---|---|---|---|
| | # of Documents (D) | # of terms in vocab (W) | Total # of tokens in the data (N) | N/D |
| Pre-processing | 6038 | 5487 | 1965852 | 325.58 |
| Post-processing | 5832 | 4348 | 1195150 | 204.93 |
| Stopwords: "and" "have" "social" "live" | | | | |
| **Priors:** Iterations: 10,000 \| burn-in: 5,000 \| alpha: 1 \| beta: 1 \|<br>**Model Fit Statistics**: CS: 3\| JS: 2\| Semantic Coherence: 2\| Exclusivity: 5\| AIC: 2\| BIC: 2\| AICC: 10\| SABIC: 2\| DIC: 10 | | | | |

# APPENDIX

## APPENDIX B: EXAMPLE ESSAYS THAT USED THE HIGHEST PROPORTION OF EACH TOPIC

| Topic | Essay |
|---|---|
| Topic 1: Mixed | I say social media doesn't make anybodies life better cause if you look at it now you will see that people are saying bad things about each other in there inbox or on Facebook or posting bad videos of people on YouTube. there are even people leaving bad comments saying bad things about people. which really stirs up a lot of drama between people even best friends. yeah a lot of people will say there making friends and yeah it can give you some valuable information but not everybody is telling the truth about social media. then there cyber bullying its mean and cruel to talk about somebody online i mean it wouldn't make it better if they said it to there face but at least you would have the guts to do it and not say it behind some code name on Facebook. If you could see somebodies inbox you would see some cruel and nasty things in there. they stir up a lot of drama and they start fight. they can even break up friendships. yeah it can sometimes help make friends but you need to watch out cause some people want get close to you just to try and ruin your life.also people like to post ugly pictures of other people or they like to comment on your picture saying you ugly. there is also people on Instagramjust like the people on Facebookand they are either better or worse. YouTube there are a lot of haters orpeople that make bad videos about people. I mean pranks are funny but there are cruel to the person that is getting pranked it might be funny to the other person, but the person that got pranked could be hurt. some people write bad things about other peoples channel but you just got to ignore them people and not let them get in your head and keep making videos. and a lot of people will be your friends just to get some dirt on you or to stab you in the back. they might comment some bad things about you like you suck or delete your YouTubechannel nobody watches you. All those things i said before lead up to this cyber bullying. It can get so bad that people kill themselves every year do to some kind of bullying. like it is out outrageous if somebody bullies you to the extent of you killing yourself. like you know its bad if they hang or stab maybe even shoot themselves. like i said in those other paragraphs they will treat you like dirt post mean pictures online about you post mean videos and if you let them do it you aren't standing up to yourself. and if you let them know they get to you then they wont stop and if they don't stop then delete your account. |
| | I think that a lot of kids have lost there lives on there social networks.But you should not have to put some much stuff in when you make your make or snapchator facebook.Becauseit makes you put in all of your life in when you make it why does it make you in some muck stuff.Like they make put in were live you |

| | |
|---|---|
| | name how old are you should not have to do all of that to have facebook or snap chat. That is why a lot of kids today die because you have to put to muck in. Yes, they need to make so changes to they were you do not have to put some muck of your life on the phone. I do not care what you say they now stuff on phone this day in time can hurt you and your family because there are some mean people in the world today that you can not trust.They will hurt you and your family because they do not care about you.They do a lot of stuff to get want they want from you and it is really sad.That people will do the stuff that they do to get what they want from of your family.They will kill you for this stuff that they can go to the store and get they are that low and cold.So when you do make your facebook be safe with it do not let your whole life get off. And do not put were your life on your facebook. Were all the people on there can see were your life.Causes those people kill you of hurt so do not do that.This what I think about your Facebook and the other stuff. So so do not put your whole life on the facebook and stuff if you want your family and friends to see it you go to there house and saw them.Because it is a lot safer doing that were thosebad people cannot see were life ok so be safe with your life do not do stupid stuff with that phone that you keep in your hand all the time. |
| | Yes they have they have made our lives a lot better because we have something to do when we are bored and don't have anything else to do.It is good when you need to contact people when you need something or when you need help because you in danger. Sometimes social media can get you in bad trouble if you aren't doing the right thing on the app.You can go to jail for somethings you do on the apps.Need to just watch what you do because they can see everything that you do on your phone. When you get bored you can get on the apps and play video games when you have nothing else to do. Most on the games on the phone are okay but some are bad and some show stuff that is nasty. I wouldn't let one of your friends get on your phone on less you are watching them way they are on it.Because if they get on something bad you could get in big trouble or looked up something on your phone that they shouldn't look at. When you are on your phone make sure that you are talking nice to people or you could get in trouble or they could end up hurting there self and then you would feel bad or you could get in trouble.It is just always good to be nice to people and watch what you are doing on your phone. |
| Topic 2: Positive | I believe that online social networking activities have improved teenagers' lives. In paragraph two of Passage A, the writer claims, \The Internet and online social networking sites offer teens opportunities to connects with others around the world.\" Social networking has provided ways for the younger generation to interact and participate in being aware about recent world events, such as politics. Also in paragraph two, the last line states, \"Because of their access to instant information via smartphones and computers, today's teens may appear more informed than previous generations.\" This connects with the idea that because of social network sites teens are more engaged in world events and such. Another reason would be that people have a chance to interact with a variety of people. People make new friends. In Passage A, again in paragraph two, it states, \"These \"plugged-in\" teenagers have specific expectations of online social networking sites and of the friends they meet on these sites.\" This supports the idea that teens are able to make new friends and have a diverse group of friends. But, however, |

| | |
|---|---|
| | this could be dangerous and lead to identity theft. My final reason would be that the education system is now dependent on technology. In Passage A, paragraph two, it states, \"These sites are virtual spaces where teens meet, chat, perform research, enjoy music, view or post videos, share and comment on photos, exchange personal stories, and much more.\" Morley focused on the \"perform research\", essays, applications, and even assignments like the one I am currently taking, is done so on a computer or another technological device." |
| | Social networking activities are improving teenagers' lives. Social has benefits like instant communication, make friends all around the world, and to see the world in other peoples eyes. The passage says, \In fact, social media has become the most popular form of instant communication among teens with access to the Internet.\" (Passage A) Teens can instantly communicate with someone if they need help, or need to talk. In the passage it said, \"Teens around the world insist that these online communities promote interactions among family and friends, invite new friendships, build relationships with friends in other countries, and offer support to individuals.\" (Passage A) Social networking has helped teens make friends, talk to family members that are far away, and promote businesses. Passage A quotes, \"These sites arevirtualspaces where teens meet, chat, perform research, enjoy music, view or post videos, share and comment on photos, exchange personal stories, and much more.\" (Passage A) Social media can help teens see how other countries live, look like, and religions. Even though social networking is helpful in many ways people can get their identity taken by using social media." |
| | Yes,I think that online social networking activities have improved teenagers' lives.In paragraph two sentence two it says that,\The Internet and online social networking sites offer teens opportunities to connect with others around the world.\"i belive it does,it also in paragraph two sentences four through seven does this ,\"has become the most popular form of instant communication among teens with access to the Internet. These sites arevirtualspaces where teens meet, chat, perform research, enjoy music, view or post videos, share and comment on photos, exchange personal stories, and much more. These â€œplugged-inâ€• teenagers have specific expectations of online social networking sites and of the friends they meet on these sites. Teens around the world insist that these online communities promote interactions among family and friends, invite new friendships, build relationships with friends in other countries, and offer support to individuals. Because of their access to instant information via smartphones and computers, todayâ€™s teens may appear more informed than previous generations." |
| Topic 3: Negative | A common question asked by many adults, \Has online social networking activities improved teenagers' lives?\". Socials networks have made life easier, but have made online services dangerous. With the rise of identity theft and hacking, virtual services have become a risk. Others may argue that social media has mostly supported the teenage lifestyle, but are ultimately wrong. The amount of risks outlaw the amount of benefits of social media.Certain young adults may find social media, YouTube, or other websites very addicting. An article called \"Information about Social Networking\" found that about \"20% report that social networking activities have negatively affected their grades\". This may be a small percentage, but if people count all the students in the world, this is a huge number of children. |

| | |
|---|---|
| | Technology can often be a distraction for teenagers in their daily and social life. The same passage took a survey and it concluded that about \"34% agree that less face-to-face interaction occurs today\" and \"32% text or chat online during family meals.\" If this were continue into their adult life, it may be difficult for them to make friends and find good jobs. These facts can cause a huge toll on teenagers' adult life.Most teens use dangerous and shady websites unknowing of the dangerous trap they are stepping into. Many and most social media require personal information and \"teens often share personal information via sites such as Facebook, Twitter, and Instagram\" (Identity Theft and Social Media). These social networks often promise a strong security, but hackers often find ways to get through it. It would be much safer if people used fake information of social media. Due to young adults carelessness, about \"28% have had personal information unknowingly shared with companies\" (Information and Social Networking). This personal information can be retrieved by a random employee and used for malicious intent. With the dangers of hacking and id" |
| | Social media has massively impacted teenager's lives. Has this improved it, though? Social media has, in fact, not improved teen's lives, if anything, it has made them worse. From less face-to-face activity, cyber bullying, and even identity theft, social media has very much negatively effected teenager's lives. These are just a few of the surplus reasons why the media is not an improvement. The first reason why social media is not an improvement for teenager's everyday live's is that they do not experience much social interaction.According to Passage A, \34% agree that less face-to-face interaction occurs...\".This shows that, with the increased use of social media, real interaction has massively declined. Because of this, many teens struggle with their communication skills. Social media has had a direct impact on face-to-face activity within teens which is, indeed, a negative one. The next reason is that many teens are bullied through a social media platform. It states in Passage A that, \"49% report being a victim of cyber bullying\". This shows that almost half of social media users are bullied. Now, not only has social media impacted teenager's social health, but also effects the mental health of teenage users dealing with cyber bullying. All-in-all, social media has done more harm than good if such a high percentage struggle with being bullied. The final reason to why social media has had a negative impact is that teen's are at a high risk of identity theft. According to Passage B, \"...children under the age of 18 were twice as often victims of identity theft as their parents. Teens are often less cautious about protecting their personal information. Teens often share personal information via sites such as Facebook, Twitter, and Instagram\". This shows that teenager's neglect keeping their information hidden from certain people, which can often lead to identity theft. Because of social media, they are at a much higher risk for this. In conclusion, social media has not improved teenager's lives. It may allow connections and friendships to be made, along with a slight increase in academic performance, but compared to all the negative effects, it cannot be counted as an improvement. Social media has come with lessened social activity within teens, cyber bullying, and an increase in identity thefts. All these negative impacts seem to out weigh the few positive ones. Overall, social media has not improved teenager's lives." |

| | |
|---|---|
| | Online social networking activities have not improved teenagers' lives. Online social networking influences many negative aspects on our lives. Online social networking activities can lead to identity theft, it affects academic performance, and it could lead to cyberbullying. Teenagers are very vulnerable to identity theft. In Passage B the writer states, \One of the ways identity thieves gain information is via social media sites. With only limited information in the profile (name, place of birth, and year), a thief has ample information to steal a person's identity.\" In this way teens are easily targeted through social media. One of the statistics from Passage A says that, \" 68% remain vulnerable to hacking, viruses, and other malware.\" which shows us that there are still a large amount of teenagers that can become involved in identity theft. Social media affects teenagers' academic performance as well. 20% of teenagers report that social networking activities have negatively affected their grades. Also while doing school work 35% admit to plagiarizing online sources. This could lead to academic dishonesty and could have you removed from an AP or Honors class." |

# APPENDIX

## APPENDIX C: 9TH GRADE ELA ARGUMENTATIVE PROMPT

Passage A: Information about Social Networking

Social networking refers to the use of a website to connect people with similar interests in an online group format. Social media is so popular that teens connect daily—or many times daily—using social networking sites like Facebook, Instagram, and Twitter.

In recent years, social media has, by and large, replaced traditional communication tools such as home telephones, hand- written notes, and emails. The Internet and online social networking sites offer teens opportunities to connect with others around the world. In fact, social media has become the most popular form of instant communication among teens with access to the Internet. These sites are virtual spaces where teens meet, chat, perform research, enjoy music, view or post videos, share and comment on photos, exchange personal stories, and much more. These "plugged-in" teenagers have specific expectations of online social networking sites and of the friends they meet on these sites. Teens around the world insist that these online communities promote interactions among family and friends, invite new friendships, build relationships with friends in other countries, and offer support to individuals. Because of their access to instant information via smartphones and computers, today's teens may appear more informed than previous generations.

According to the 2014 Forrester's North American Consumer Youth Survey, the following statistics indicate that teens (ages 12-17) spend at least some time each day using social media. Consider the following statistics:

- 40% spend an average of one hour per day on social media.
- 67% regularly use Facebook.
- 34% agree that less face-to-face interaction occurs today.
- 98% have viewed or posted YouTube videos.
- 52% use Instagram daily.
- 46% agree sites encourage social awareness and community involvement.
- 22% visit Twitter weekly.
- 49% learn false information online.
- 28% have had personal information unknowingly shared with companies.
- 57% get to know fellow students at school better.
- 52% state increases in friends/relationships.
- 20% report that social networking activities have negatively affected their grades; 59% note an increase in academic performance.
- 32% text or chat online during family meals.
- 49% report being a victim of cyberbullying.
- 64% listen to new music on YouTube exclusively.
- 35% admit to plagiarizing online sources.
- 85% post comments on others' pictures/comments.
- 68% remain vulnerable to hacking, viruses, and other malware.

Page 2 of 5 Research indicates that social networking has had both positive and negative impacts on our lives. Using these resources responsibly needs to be a part of what we do every day.

Source: Forrester's North American Consumer Youth Survey

Use the following passage for question #1, #2, #3, #4, #5.

Passage B: Identity Theft and Social Media

Growing up, most of us were likely taught not to speak to strangers, not to give out our names and addresses to people we didn't know, or in general, to keep personal information to ourselves. But today, people have profiles on social media sites, smart phones capture our daily movements using GPS signals, and information about almost anyone can be found by just Googling a name on the internet. Is it no wonder that identity theft has become a major concern of people all over the world today and a growing area of crime?

In 2012, the Federal Trade Commission reported that more than 360,000 complaints of identity theft were registered. In 2014, Georgia ranked second per capita in reporting identity theft complaints. These complaints included everything from being asked to visit a scam website in an email to being the target of identity theft. Identity thieves are continuously finding new avenues to steal personal information.

One of the ways identity thieves gain information is via social media sites. With only limited information in the profile (name, place of birth, and year), a thief has ample information to steal a person's identity. Doing a status update, however innocent this may seem, can leave you open to a variety of crimes. Making the statement that you are going away for the week on vacation alerts a criminal that your home will be vacant. It can also make an identity thief aware that your bank account may be accessed from remote locations via your phone or a public computer.

While people are becoming savvier about protecting themselves, many people are not aware that some sites ask for sensitive information that could be used in inappropriate ways to either steal your identity or to use your identity to commit crimes in malicious ways. The

following information is often included in your profile, or is disclosed in conversations on social media sites:

- Your full name, including your middle name
- Your date of birth, which is often required
- Your home town
- Your relationship status
- The schools you attended and/or graduated from
- Your pet's name
- Your religious affiliations, hobbies, clubs, etc.

Who are those most susceptible to identity theft? Statistics from the Federal Trade Commission show that the group with the highest percentage of identity theft victims are aged 20-29. Some of the reasons for this seem to be that this group of people is less likely to keep a regular check on their bank accounts and credit card activity. The records indicate that they are also less careful when sharing information on social media.

In 2011, a Carnegie Mellon study of more than 40,000 children reported that children under the age of 18 were twice as often victims of identity theft as their parents. Teens are often less cautious about protecting their personal information. Teens often share personal information via sites such as Facebook, Twitter, and Instagram. The information shared often includes hints to things that a person might use to answer security questions for secure sites. Teens often have unused social security numbers. Thieves might sell the social security number to illegal immigrants so that they can obtain a fake ID for employment. Teens are also less cautious about sharing a credit card number with a friend to buy tickets for a concert or when they go out to eat. If the friend writes down the number and is not careful about destroying it, the number can be

compromised. When teens use Wi-Fi, it is easy for a thief to obtain personal information. Smart phones store and transmit personal data. If the phone is lost, the information can be retrieved and used for malicious purposes. Importantly, teens often don't understand the ramifications of identity theft and often consider themselves 'immune' to being a target.

The impact identity theft has on things like applying for college scholarships, buying your first car, getting your first job, or your first place to live could be an obstacle that could take years to overcome. An example of a teen who was caught up in identity theft turmoil. I found myself sitting in an attorney's office with my parents, waiting to meet with someone who I hoped could help me. Someone has stolen my identity and I am trying to get into college. When my college loan application was rejected, I found out information that led me to the conclusion that someone was using my personal information.  I need help!

Apparently, someone had stolen my social security number and has been using it for a while. I lost my wallet a couple years ago, but I never reported my social security card as stolen. I also filled out all the information in the Profile on Facebook and used my phone to make purchases all the time. I had no idea when or how the person had connected my information, but this unknown thief has created a lot of problems for me. Once I started investigating, with my parents help, we found that the person had applied for credit cards in my name, bought a car, and some furniture. I had contacted the credit card companies, but I have to go through some type of legal process to prove that I am really me and that I am not the one who ran up all those bills. Once we talked to the attorney, we realized that it was going to take some time and cost some money to try to clear my name and set things straight. He worked really hard to help me talk to the college loan offices so that I could apply for loans to go to school. It took an extra year before the matter had been resolved to the point that I could get the loan, and I was finally able to

start school. The attorney is still working to clear up the credit card debt. What a way to start my life after graduation!

How do we prevent becoming victims of identity theft? Just a couple of easy steps can help when using social media sites. These will not protect against all identity theft, but will be a good place for everyone to start to be proactive and responsible.

• Use different passwords for every site you use.

• Don't share passwords with anyone.

• Use a password to access your phone.

• Keep track of credit card and bank statements regularly.

• Use credit protection services if you can to monitor your accounts.

• Shred all mail that has any identifying information on it instead of discarding it in trash

• Be responsible and protect your identity. There should be only one you!

#5 Constructed Response

Have online social networking activities improved teenagers' lives? Write an argumentative essay stating whether or not online social networking has improved the lives of teenagers. Be sure to:

• Introduce your claim

• Acknowledge and counter alternate or opposing claims

• Develop your claim and counterclaims fairly to point out the strengths and limitations of both using evidence from BOTH passages

• Use words, phrases, and clauses to clarify relationships between claim and reasons, between reasons and evidence, and between claim and counterclaims

• Provide a concluding statement or section that follows from and supports the argument

# APPENDIX

## APPENDIX D: SENTIMENT LEXICON

Table of Potential Indicator Words and Phrases for Risky Behavior, Negative Sentiment and Disorders from the DSM-5 and KSADS

| Disorder | Characteristics of Disorder | Key words/potential indicators<br>Also known as **Sentiment Lexicon** |
|---|---|---|
| Generalized Anxiety Disorder | • Excessive anxiety and worry (apprehensive expectation), occurring more days than not for at least 6 months about a number of events or activities (such as work or school performance)<br>• Individual finds it difficult to control the worry<br>• The anxiety and worry are associated with three (or more) of the following 6 symptoms: Restlessness or feeling on edge, easily fatigued, difficulty concentrating or mind going blank, irritability, muscle tension, sleep disturbance<br>• Anxiety, worry or physical symptoms cause significant distress or impairment in social, occupational or other important areas of functioning<br>• 12-month prevalence of GAD among adolescents is 0.9% and 2.9% for adults in the general US community<br>• Lifetime morbid risk is 9% and females are twice as likely as males to experience GAD<br>• In children and adolescents with GAD, anxieties and worries often concern quality of their performance or competence in | • Worry, worried, sick, feel well, headaches, stomach aches, aches, school days, stupid, up-tight, tense, can't relax, on edge, jumpy, anxious,<br>• Can't stop thinking about the past, need to be good at everything, hate school, anxiety about leaving home |

| | | |
|---|---|---|
| | school or sporting events, even when their performance is not being evaluated by others.<br>• GAD may be over diagnosed in children | |
| Social Anxiety Disorder | • Marked fear or anxiety about one or more social situations in which the individual is exposed to possible scrutiny by others<br>• The individual fears that he or she will act in a way or show anxiety symptoms that will be negatively evaluated<br>• The social situation almost always provokes fear or anxiety<br>• Social situations are avoided or endured with intense fear or anxiety<br>• The fear or anxiety is out of proportion to the actual threat posed by the social situation and to the sociocultural context<br>• The anxiety is persistent for 6 months or more and causes clinically significant distress or impairment in social, occupational or other important areas of functioning<br>• Social communication skills developed appropriately but are not utilized because of anxiety, fear or distress about social interactions<br>• Median onset of SAD is 13 years old and 75% have had an age onset between 8 and 15 years old | • Shy, can't say anything, hard to talk, hard to talk in class, fear of talking, fear, embarrassed, don't like being around people, nervous, self-conscious, uncomfortable, speaking in front of others, in front of others, performance situations, social events, can't talk, hate school, pretend to be sick, sick, anxiety about leaving home, avoid people, avoid meeting people, hate speaking in public |
| Panic Attack | • An abrupt surge of intense fear or intense discomfort that reaches a peak <u>within minutes</u> and during which time four or more of the following symptoms occur: palpitations (pounding heart or increase heart rates), sweating, trembling or shaking, sensations of shortness of breath or smothering, feelings of choking, chest pain or discomfort, nausea or abdominal | • Same as panic disorder |

| | | |
|---|---|---|
| | distress, feeling dizzy, unsteady, or faint, chills or heat sensations, paresthesia (numbness or tingling sensations), derealization (feelings of unreality) or depersonalization, fear of losing control, fear of dying,<br>• At least of the attacks has been followed by 1 month or more of one or both of the following: persistent concern or worry about additional panic attacks, a significant maladaptive chance in behavior related to the attacks<br>• The disturbance is not due to drugs or other medical conditions and cannot be better explained by another mental disorder<br>• Two types of panic attacks: expected and unexpected<br>• Expected attacks have an obvious cue<br>• More common in females than males but both present with the same symptoms<br>• Panic attacks are uncommon,<br>and unexpected panic attacks are rare, in preadolescent children. Adolescents might be less willing than adults to openly discuss panic attacks, even though they present with episodes of intense fear or discomfort. | |
| Panic Disorder | • Recurrent unexpected panic attacks<br>• Median age at onset for panic disorder in the US is 20-24 years old<br>• A small number begin in childhood<br>• Panic disorder should not be diagnosed if full-symptom (unexpected) panic attacks have never been experienced. | • Heart racing, out of the blue, randomly felt, anxious, nervous, frightened, chest pain, sweating, trembling, shaking, can't breathe, dizziness, fear of dying, losing control, fear, numbing, throwing up, nausea, frozen, freeze, flashbacks |

| | | |
|---|---|---|
| | • Panic attacks that occur as a symptom of other anxiety disorders are expected. However, if the individual experiences unexpected panic attacks as well and shows persistent concern and worry or behavioral change because of the attacks, then an additional diagnosis of panic disorder should be considered | |
| Disruptive Mood Dysregulation Disorder | • Severe recurrent temper outbursts manifested verbally that are grossly out of proportion in intensity or duration to the situation or provocation.<br>• temper outbursts are inconsistent with developmental level.<br>• The temper outbursts occur, on average, three or more times per week.<br>• The mood between temper outbursts is persistently irritable or angry most of the day, nearly every day, and is observable by others (e.g., parents, teachers, peers).<br>• Prior Criteria have been present for 12 or more months. Throughout that time, the individual has not had a period lasting 3 or more consecutive months without all of the symptoms in Criteria A-D.<br>• Criteria A and D are present in at least two of three settings (i.e., at home, at school, with peers) and are severe in at least one of these.<br>• The diagnosis should not be made for the first time before age 6 years or after age 18 years.<br>• Criteria 1-5 need to be made before age 10<br>• There has never been a distinct period lasting more than 1 day during which the full symptom criteria, except duration, for a | • Cranky, irritable, angry, easy to get angry, punch, kick, beat, hurt, set off, arguments, fight, |

| | | |
|---|---|---|
| | manic or hypomanic episode have been met. | |
| Major Depression (p160) | • Five or more of the following symptoms have been present during the same 2-week period and present a chance from previous functioning: at least one of the symptoms is either depressed mood or loss of interest or pleasure<br>• 1. Depressed moods of the day, nearly every day as indicated in subjective report or reported by others<br>• Markedly diminished interest or pleasure in all or almost all activities, most of the day, nearly every day<br>• Significant weight loss when not dieting or decreased/increased appetite every day<br>• Insomnia or hypersomnia nearly every day<br>• Psychomotor agitation or retardation nearly every day<br>• Fatigue or loss of energy nearly every day<br>• Feelings of worthlessness or excessive or inappropriate guilt nearly every day<br>• Diminished ability to think or concentrate or indecisiveness nearly every day<br>• Recurrent thoughts of death, recurrent suicidal ideation without a specific plan or a suicide attempt or a specific plan for committing suicide<br>• Symptoms cause clinically significant distress or impairment in social, occupational or other important areas of functioning | • Depressive: sad, blue, down, empty, depressed, gloomy, unhappy, feels like crying, miserable, agitated<br>• Irritability/anger: annoyed, irritated, cranky, mad, angry, temper, lost your temper, hurting myself, killing, torturing<br>• Not interested, bored, concentrate, enjoy less,<br>• Suicidal Ideation: dying, killing, plan (in relation to killing), actually tried to kill, tell in advance, want to die, hurt yourself, hurt, hurt others, hopeless<br>• Bad feelings, tearful, come and go, loneliness, unloved, alone, friendless, no one cares, ugly, hate myself, unfair, bad luck, punishment, guilty, helpless, eating all the time, don't feel like eating |
| Trauma- and Stressor-Related (PTSD) (p265) | • Must be older than 6 years old<br>• Persistent and exaggerated negative beliefs or expectations | • I have always had bad judgment<br>• People in authority can't be trusted<br>• I can't trust anyone ever again |

| | | |
|---|---|---|
| | about oneself, others, or the world (e.g., “I am bad,” “No one can be trusted,” ‘The world is completely dangerous,” “My whole nervous system is permanently ruined”).<br>• Another form is persistent and exaggerated negative expectations regarding important aspects of life applied to oneself, others, or the future (e.g., “I have always had bad judgment"; "People in authority can't be trusted") that may manifest as a negative change in perceived identity since the trauma (e.g., "I can't trust anyone ever again"; Criterion D2). Individuals with PTSD may have persistent erroneous cognitions about the causes of the traumatic event that lead them to blame themselves or others (e.g., "It's all my fault that my uncle abused me") | • It's all my fault that my uncle abused me<br>• See “again and again”, think again and again, hear again and again<br>• Stop bad thoughts, get rid of bad thoughts<br>• Nightmares, can’t sleep<br>• Reminded, avoided, hard to trust, threaten, hurt, hard to trust, alone<br>• Concentrating, hard to concentrate, anxious, |
| Disruptive Mood Dysregulation Disorder (DMDD) | DMDD is closely related in longitudinal studies to depression. Common for kids to express irritability for depression instead of sadness or loss of pleasure/interest | • annoyed, angry, lie, liar, break promises, con, fight, hit someone, hit, kill, bully, threaten, call names, make fun, push, kick, stolen, |

# APPENDIX

## APPENDIX E: COMPARISON OF SENTIMENT LEXICON USAGE

Table Comparing Negative Indicator Sentiment Lexicon Between Essays that Used the Highest Proportion of Topic 1 to Topic 3

| Negative Sentiment Lexicon in Each Essay | |
|---|---|
| Topic 3 Essay | Topic 1 Essay |
| Bullying | Bad |
| Theft | Cyber bullying |
| Negatively effected | Mean |
| Victim | Fight |
| Harm | Ruin your life |
| Theft | Ugly |
| Neglect | Hurt |
| | Stab you in the back |
| | You suck |
| | Nobody watches you* |
| | Killing |
| | Hang |
| | Stab |
| | Shoot |
| | Treat you like dirt |

*In the context of the essay, this statement refers to nobody wanting to give another person attention and they're neglected

# APPENDIX

## APPENDIX F: R PACKAGE *topicmodels* APPLICATION

Sample Code Application

```
library(topicmodels) #load library
data = read.csv(file='', header=T, sep=",", fill=T) #read in data

#clean data using desired methods
#Transform data into the format required
get.terms <- function(x) {index = match(x, vocab)
  index <- index[!is.na(index)]
  rbind(as.integer(index - 1), as.integer(rep(1, length(index))))}
documents <- lapply(new.doc.list, get.terms)
del.doc = which(sapply(documents,length)<10)
dat1 =data[-del.doc,]

#load stemming list
#remove stop words based on TFIDF to create final dataset from model training and testing
pstP_dtm0_total<-DocumentTermMatrix(dat1, control=list(stemming=F, minWordLength =
1))
two0_total<-findFreqTerms(pstP_dtm0_total,5)
pstP_two.total<-pstP_dtm0_total
term_tfidf <-tapply(pstP_two.total$v/row_sums(pstP_two.total)[pstP_two.total$i],
pstP_two.total$j, mean)*log2(nDocs(pstP_two.total)/col_sums(pstP_two.total > 0))
cv<-as.numeric(term_tfidf[order(term_tfidf)[28]])
cv<-quantile(term_tfidf,0.0020)
pstP_two2.total <- pstP_two.total[,term_tfidf >= cv]
pstP_two3.total <- pstP_two2.total[row_sums(pstP_two2.total) > 1,]
training_dataset <- pstP_two3.total[col_sums(pstP_two3.total) > 1,]

pstP_two.test<-DocumentTermMatrix(dat1)
pstP_two.test2 <- pstP_two.test[,term_tfidf >=cv]
pstP_two.test3 <- pstP_two.test2[row_sums(pstP_two.test2) > 1,]
test_dataset <- pstP_two.test3[col_sums(pstP_two.test3) > 1,]

#set parameters for generating topic models with different numbers of topics
sequ = seq(2, 7, 1) # in this case a sequence of numbers from 2 to 10, by ones.
alpha = 1
delta = 1
SEED = 3423
iter = 10000
```

```
keep = 1
burnin = 5000
# use lapply function in topicmodels to generate topic models using given parameters
train <- lapply(sequ, function(k) LDA(training_dataset, k = k, method = "Gibbs",
control=list(alpha=alpha, delta=delta, seed=SEED, burnin=burnin, iter = iter, keep = keep)))
test<- lapply(train, function(l) LDA(test_dataset, model=l, k = sequ, method = "Gibbs",
control=list(alpha=alpha, delta=delta, seed=SEED, burnin=burnin, iter = iter, keep = keep)))

#select number of topics and specify
k = 3 #number of topics

#use different seed and LDA function in topicmodels to generate topics structures for specific
k
SEED = 7845652
m <-LDA(pstP_two3.total, k=k, method = "Gibbs", control=list(alpha=alpha, delta=delta,
seed=SEED, burnin=burnin, iter=20000))

#Post processing to extract topic structures
n1<-table(m@z)[1]
n2<-table(m@z)[2]
n3<-table(m@z)[3]
Topic_k<-topics(m,3)
Terms_k<-terms(m,30)
term.prob<-posterior(m, pstP_two3.total)$terms
list.m.t1<-order(term.prob[1,], decreasing=TRUE)[1:30]
list.m.t2<-order(term.prob[2,], decreasing=TRUE)[1:30]
list.m.t3<-order(term.prob[3,], decreasing=TRUE)[1:30]
m.t1<-term.prob[1,list.m.t1]
m.t2<-term.prob[2,list.m.t2]
m.t3<-term.prob[3,list.m.t3]

#data1, data2, and data3 are the topic structures
data1<-round(as.data.frame(m.t1),3)
data2<-round(as.data.frame(m.t2),3)
data3<-round(as.data.frame(m.t3),3)
```