12-2023

Did That Help? Harnessing Natural Language Processing to Uncover Characteristics of Peer Feedback and Their Impacts on Teamwork-Skills Learning.

Annamaria V. Wolf
Clemson University, avw@clemson.edu

DID THAT HELP? HARNESSING NATURAL LANGUAGE PROCESSING TO UNCOVER CHARACTERISTICS OF PEER FEEDBACK AND THEIR IMPACTS ON TEAMWORK-SKILLS LEARNING.

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Industrial-Organizaational Psychology

---

by
Annamaria Wolf
December 2023

Accepted by:
Dr. Marissa Shuffler, Committee Chair
Dr. Jannifer Ogle
Dr. Allison Trayler
Dr. Patrick Rosopa

ABSTRACT

Peer Evaluation Systems (PESs) allow members of student teams to provide one another with computer-mediated feedback in the form of qualitative, open-ended comments. The current research leverages unsupervised Natural Language Processing (NLP), namely Biterm Topic Modeling (BTM) and sentiment analysis, to uncover latent topics and degree of positivity and negativity expressed in peer feedback, respectively. BTM results revealed a 6-topic model that was reliably replicated over 10 Gibbs initializations 80% of the time. Topics were labeled *Timely Communication, Idea Generation, Coordination & Adaptation, Work Quality, Team Support & Focusing,* and *Work Accountability*. Qualitative comparison suggests that these topics demonstrate significant overlap with concepts detailed within existing teamwork and feedback frameworks. Sentiment analysis indicated that peer feedback had a predominantly neutral to positive valence orientation, but that the analysis had limited accuracy. Significance testing evaluating the impact of the topic of feedback, feedback valence, and feedback length on outcome measures of students' learning of teamwork skills were entirely non-significant. These results are discussed, along with discussions of NLP's potential to expand existing theories and frameworks with data-driven techniques, and to provide educators with rapid, high-level insight from PESs that support student learning outcomes.

Keywords: feedback, peer evaluation system, teams, engineering, topic modeling, sentiment analysis

DEDICATION

I dedicate this dissertation to the incredible friends I have made during graduate school. The genuine connections and shared experiences we have had truly been the best part of this academic adventure. In moments of difficulty or disheartenment, you were a trusted shoulder to lean on, supporting me with the practical challenges of life or intellectual quandaries, and giving me confidence and courage to press on. Your advice, encouragement, and shared laughter have made my academic journey more than a scholarly pursuit; it has been a rich era of personal growth in many facets. To each friend who has played a part in this chapter of my life, I express my deepest gratitude.

AKNOWLEDGMENTS

I extend my deepest gratitude to my supervisor, Dr. Marissa Shuffler, who has provided unwavering empowerment of my ideas and encouragement throughout my dissertation research. I am eternally thankful for the countless opportunities to excel she has offered me during my graduate school career. I owe the development of my competencies as an Industrial-Organizational psychologist and teams' researcher, in large part, to her mentorship and dedication to my success.

To each of my committee members, it has been my pleasure to collaborate with you on this dissertation research. Thank you for lending your specialized expertise and insights, particularly within the fields of team science, learning and development, civil engineering pedagogy, and statistics and machine learning. This academic work has been elevated by your mentorship, creative insights, and constructive feedback.

I extend my heartfelt appreciation to my peers who, on several occasions, openly shared their wisdom and opinions on technical matters related to my research and often engaged in stimulating, thought provoking discussions. I am thankful to have learned from these individuals, and to have shared in the spirit of mutually supportive professional collaboration.

Finally, I would like to thank Dr. André Bittermann and his collaborators for their provision of open-source R code which has been invaluable to this research and responding to my email inquiries on this subject. This experience underscores the profound impact that open-source science has had on making research more accessible, collaborative, and ultimately, enhancing the depth and breadth of my work.

TABLE OF CONTENTS

Table of Contents (Continued) Page

Table of Contents (Continued).. Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

The ability to work effectively in a team environment is necessary across a range of fields and professional roles, including within the field of engineering (O'Neill & Salas, 2018). As such, it is important that university curriculums strive to prepare students for the demands of their professional careers by equipping them with teamwork skills and competencies (Donia et al., 2018; Petkova et al., 2021). Peer feedback has an important role to play in the development of these skills in higher education (Ohland et al., 2012); feedback from ones team members is shown to support learning of teamwork skills, and ability to perform effectively in teams (Brutus et al., 2013; Brutus & Donia, 2010). Feedback is also important for team functioning as a whole, known to be a crucial factor in driving effective team processes and outcomes (DeShon et al., 2004; Gabelica et al., 2012; London & Sessa, 2006). More generally, the ability to give and receive feedback is a central feature of many common workplace practices, such as performance appraisals, 360-feedback programs, and team debriefs exercises (Bracken et al., 2001; Murphy & Cleveland, 1991; Tannenbaum & Cerasoli, 2013).

In response to the growing demand for teamwork competencies in the workforce, University programs are implementing structured systems to support teamwork skills learning and development (Baker, 2008; Hughes & Jones, 2011; Ohland et al., 2012). An important component of many of these systems is the Peer Evaluation System (PES), which support teamwork learning by providing structured mechanisms through which team members can exchange formative feedback regarding one another's teamworking skills, behaviors, and abilities (e.g., O'Neill et al., 2019). The majority of research of PES' impact on learning and development focus on the quantitative components of the systems (e.g., Brutus et al., 2013; Donia et al., 2018; O'Neill et al., 2019).

Qualitative feedback in the form of written comments serve an important purpose for student learning, by for instance contextualizing and providing further insight into quantitative scores and assessments, and direction for improvement (Hattie & Timperley, 2007). Yet, as compared to the quantitative components of PES, the impact of qualitative feedback appears to be largely under-researched (see Brutus et al., 2013 for an exception). The information-rich, complex nature of student's qualitative data makes it notoriously more difficult to analyze (Hodges & Stanton, 2007). Traditional qualitative analysis, in general, invovles laborious and time-consuming human coding to draw insights about themes, patterns, or categories represented within the data (Creswell, 2008).

Natural Language Processing (NLP) techniques hold potential to expediate and automate the insights that can be drawn from qualitative data within PES. This is especially true for unsupervised NLP, wherein discovering patterns, themes, and insights within texts are possible without applying predefined, human-generated labels or categories (Vajjala et al., 2020). *Topic Modeling* (TM) and *Sentiment Analysis* (SA) are two NLP techniques that can be applied in an unsupervised manner. TM can identify hidden, 'latent' topics, within a collection of text documents (Vayansky & Kumar, 2020), and in its simplest form, SA evaluates the valence (e.g., degree of positive, negative, or neutral tone) expressed in text (Mite-Baidal et al., 2018). These techniques are already showing potential as a method to rapidly draw insights about student feedback, for the purposes of support educational programs. For instance, they can be used to discover what students are primarily discussing and whether commentary is generally positive or negative, when leaving feedback and evaluations of teachers, courses, or peers (Gottipati et al., 2017; Nanda et al., 2021, 2022; Sun & Yan, 2023).

As compared to traditional methods of qualitative coding, NLP methods can draw insights automatically and, at times, rapidly. This capability could be harnessed by educators wishing to gain high-level insight text data of feedback exchanged between peers/team-mates, to support, understand, and positively impact student learning of teamwork skills and course design. However, further research is needed to implement these NLP techniques for this purpose, and identify and extend the extent of its capabilities and limitations. For instance, research is needed to determine whether NLP can be used in combination with statistical methods to draw conclusions about the impact of qualitative feedback characteristics on student learning outcomes. This constitutes the overarching aim of the current research.

The subsequent sections provide background information to illustrate the importance of peer/team-member feedback on individual and team development, the impact of PESs to support students' teamwork skills development, NLP techniques and uses in pedagogical research, the proposed aims and associated research questions and methodologies.

Background

Peer Feedback and Teamwork Skills Development

Teams are identifiable when at least two or more individuals which are operating in a similar organizational context, have at least one shared objective or goal which requires their interdependent collaboration to achieve (Kozlowski & Bell, 2003; Salas et al., 1992). Team-based work is becoming the foundation of many industry processes (Mathieu et al., 2019; Rapp et al., 2021; Salas et al., 2015). No wonder, that reports suggest that 'teamwork' is one of the top desirable skills rated by employers (Kommission, 2011). Educational authorities within certain fields, such as the Accreditation Board for Engineering and Technology, has named teamwork, and the ability to work effectively in teams, as a key desirable outcome for any accredited engineering program (ABET, 2021). Graduates within engineering disciplines also rated teamwork as one of the most critical ABET-identified competencies to their success in professional settings (Passow, 2012).

In response to the rising industry demands for graduates with professional teamwork skills, university course curriculums within engineering and technology are increasing team-based, hands-on, applied learning. Such experiential learning opportunities are necessary, since interpersonal and teamwork skills are not effectively transferred through lectures or textbook reading (Veine et al., 2020). Yet, providing the opportunity to work in teams is not, in and of itself, sufficient for the development of teamwork skills (Oakley et al., 2004), but also relies on high-quality, formative feedback to supplement these experiences (Veine et al., 2020).

Feedback can originate from internal source, such as self-evaluations, as well as external sources such teachers, managers, and peers (Murphy & Cleveland, 1991). Feedback can be directed to individuals or teams (London & Sessa, 2006), and be used for different purposes like evaluating performance or amount of knowledge held (i.e., summative feedback) or providing direction and guidance for learning (i.e., formative feedback) (Scriven, 1967).

According to Feedback Intervention Theory, external feedback may yield benefits because gaining information about how others see you reveals one's blind-spots and, potentially, provides insight regarding how to adapt actions and behaviors to enhance performance (Kluger & DeNisi, 1996; Pienaar & Nel, 2017). Formative peer feedback has been cited as particularly beneficial in evaluating and developing interpersonal skills and teamwork skills, since peers, colleagues, or team members are likely to have more opportunities to observe and interact with the target than teachers, coaches, or managers (Falchikov & Goldfinch, 2000; London & Smither, 1995; Millis & Cottell, 1998). This is supported by research, such as Petkova et al.'s (2021), a mixed methods study that found a strong relationship between evaluations received by students from their teammates mid-semester and increased ratings of teamwork effectiveness later in the semester. Interviews with students suggested that this was because peer feedback unveiled what behavior changes were needed for them to work better in the team.

Team-Member Feedback and Team Development

The current research focuses on feedback directed towards the individual team members by another team member, however, feedback can be directed towards the team-level (for a review see Gabelica et al., 2012). This section briefly describes the differences between individual-level and team-level feedback and showcases the importance feedback can play for the performance of not only individuals, but the overall team. Although this will not be the focus of the current study, it illustrates why providing effective feedback, and working effectively in a team may be considered complimentary skills. This section does not provide an exhaustive understanding of relative impacts of individual-level, team-level, and combined feedback on team and individual performance outcomes. For a deeper understanding on this, the reader is directed towards the following literature: DeShon et al., 2004; Gabelica et al., 2012; Handke et al., 2022; London & Sessa, 2006).

Feedback has played an important role in so-called, Team Development Interventions (TDIs), which aim to support effective team performance and reduce breakdowns in teams, which are dreadfully common and can have disastrous consequences (Bell & Kozlowski, 2011). TDIs include myriad activities and programs (Shuffler et al., 2018), including, e.g., teamwork-skills-oriented trainings (Salas & Cannon-Bowers, 2001), team building to strengthen interpersonal bonds and synchrony (Dyer et al., 2007). Team-member feedback is inherent to many such TDIs but are especially central to one TDI known as team debriefs (Ellis & Davidi, 2005; Shuffler et al., 2018), which invovles teams jointly reflecting on recent activities or performance episodes, to understand what lead to positive or negative outcomes (Tannenbaum & Cerasoli, 2013).

Team debriefs are unique from individually focused feedback in that debriefs are at the team level with open discussion between all members (Ellis & Davidi, 2005; Tannenbaum & Cerasoli, 2013). However, individually targeted feedback may occur within debriefs, or in addition to debriefs. For instance, team members may provide one another with feedback related to how their specific actions or behaviors contributed to team outcomes (London & Sessa, 2006). This can play a critical role in various aspects of improving and maintaining team effectiveness (Dominick et al., 1997; Gabelica et al., 2012). Individually targeted feedback can make members aware of what the teams' standards are, and how their behaviors are aligned with these standards and expectations (Kapp, 2009). It can provide the opportunity for members to hold one another accountable to the team, and reduce free-riding behavior (Anson & Goodman, 2014; Bacon et al., 1999; Croy & Eva, 2018; Millis & Cottell, 1998).

Clearly, feedback has a multifaceted impact in individual and team development; not only can individual-level teamwork skills benefit from peer feedback (Brutus & Donia, 2010; Petkova et al., 2021), peer feedback reinforces individuals' performance as teammates and can also support the performance of the team as a whole (Gabelica et al., 2012; Kapp, 2009). Thus, skills in effective feedback provision and teamworking are complimentary. Also, feedback is a central part of working professional life; feedback comes into play during performance appraisals (Murphy & Cleveland, 1991), which increasingly includes peer feedback such as through 360-degree feedback programs

(Bracken et al., 2001), or activities such as team debriefs that specifically target team performance and necessitate team-member feedback (Tannenbaum & Cerasoli, 2013).

The university setting is an excellent time to begin equipping individuals with interpersonal skills needed for providing feedback and working in teams. However, it should not be assumed that merely providing the opportunity for students to work in groups and teams, and asking them to exchange feedback, will automatically result in learning and growths in these areas (Petkova et al., 2021). Similarly, the opportunity to practice giving and receiving feedback will not automatically benefit student development without the provision of guidance and direction that supports high-quality feedback exchanges (e.g. Chen et al., 2004). The next section will explore what effective, high-quality feedback may entail.

Feedback Effectiveness and Quality

The provision of feedback is not always guaranteed to be beneficial. In Kluger and DeNisi's (1996) seminal meta-analysis of feedback and performance research, one third of the studies they reviewed found a negative relationship between feedback and performance, which is reiterated in later research (Krenn et al., 2013). A large body of research has been dedicated to understanding what makes written feedback helpful or unhelpful for learning and development, and why. The answer to this question is complex, and it is not within the score of the current research to consider the full extent of this issues. However, the following section will provide some overviews of the key topics that have been considered as relevant to consider for feedback effectiveness and quality.

Feedback Focus

Kluger and DeNisi's (1996) meta-analysis found that feedback may have different focuses, which may impact the extent to which it benefits the recipient's performance. Broadly, feedback may focus on learning of tasks, motivation to complete tasks, or personal attributes or characteristics of the learner. The authors conclude that feedback that is focused on the former of the three categories, task completion, are most likely to result in improved performance, while feedback targeting the self would garner the least positive effect on performance. In addition to focusing on tasks or processes, rather than the self and individual characteristics, high-quality feedback generally includes an element of future-orientation that includes actionable, specific directions for ways to improve (Fong et al., 2019; Hattie et al., 2021; Hattie & Timperley, 2007; Mandouit & Hattie, 2023).

Feedback Positivity Vs Negativity

One area that has been heavily researched, but is rife with mixed results, concerns the impact of feedback valance, or the extent that feedback is positive, negative, or neutral. One meta-analysis compared studies with clashing theories regarding the impact of feedback valence on learning motivation (Fong et al., 2019). Herein, some studies suggested that negative, or critical feedback, may boost motivation to learn by empowering individuals with information needed to enhance competencies and achieve goals, while other studies assumed that negative feedback would damage learner's self-perception and sense of competence, thus reducing their intrinsic desire to improve. Overall, the meta-analysis found that negative feedback was less motivating than positive feedback, but that these effects were buffered when feedback is focused on the task rather than the person and when feedback gives directions for how to improve.

Other studies paint a similarly mixed and context-dependent picture when it comes to the impact of feedback valence. A controlled experimental study found that feedback valence had a strong significant effect on reactions to feedback and intention to improve, such that positive feedback resulted in positive reactions while negative feedback resulted in neutral reactions, and negative feedback resulted in stronger intentions to improve (Bloom & Hautaluoma, 1987). Research of the impact of teacher's feedback on student demonstrates that a balance of positive *and* negative feedback, encouragement *and* constructive criticism, is important for learning (e.g., Anson & Anson, 2017; Ferris, 2014). This trend is echoed in the context of peer-feedback by one study where over 10,000 students rated the usefulness and quality of 100,000 written peer-evaluations. Here, it was found that although negative feedback was generally rated more harshly than positive feedback, feedback that was too positive was also deemed as less useful (Wind & Jensen, 2017).

Feedback Length

Some research has studied the impact of feedback length on perceptions of feedback quality. Wind and Jensen's (2017) research found that longer feedback was perceived more positively by recipients, which aligns with previous research that found shorter feedback was perceived as being of lower-quality as compared to average (Wessa & De Rycker, 2010) and longer feedback (Zong et al., 2021). Another study that compared positive and lengthy, positive and concise, negative and lengthy, negative and concise feedback found that the first category of positive and lengthy was best preferred by student raters, who indicated that feedback of this nature was most helpful, most detailed, and most motivating (Häkkinen & Ramadan, 2023). Interestingly, there is evidence that lengths of peer-to-peer feedback provided had a positive association with learning and performance transfer of the feedback giver more than the receiver (Yu & Schunn, 2023). From this section, we can summarize that longer feedback is generally considered to be of higher quality, perhaps especially when it is positive, and perhaps because it is more likely to contain some of the other critical aspects of high-quality feedback such as having specific and actionable insights and guidance, with a focus on tasks and processes rather than individual characteristics.

Peer Evaluation Systems for Teamwork Skills Learning

Knowing that there may be a complex interplay of factors that impact feedback quality, and that provision of feedback itself is not guaranteed to result in positive learning outcomes (e.g., Kluger & DeNisi, 1996) makes guidance ever-more important in educational settings aiming to support the development of teamwork competencies by utilizing peer (or team-member) feedback. Recognizing that educators require support in developing students' teamwork competencies and enabling structured peer-feedback to support individual and team development, team science scholars have created standardized peer evaluation systems towards this end (Planas-Lladó et al., 2021). These digital platforms, which have gone through rigorous validation processes (e.g., Loughry et al., 2007; Ohland et al., 2012; O'Neill et al., 2019), allow instructors and students to garner insights into the performance of their team and individual members, usually by having members complete standardized measures intended to capture team-work related processes and performance (e.g., Ohland et al., 2012). These evaluation systems have been used in employment settings as well as within higher education. For instance, the Comprehensive Assessment of Team Member Effectiveness (CATME) is a rating system that allows team members to evaluate one another's teamwork behaviors and contributions, and can be used as a tool for feedback or/ performance evaluation (Loughry et al. 2007).

These evaluation systems have been shown to benefit student teams that use them with regards to their teamwork skills development, and their abilities to provide quality feedback to teammates. For instance, research suggests that students who were repeatedly exposed to standardized peer evaluation systems over the course of several semesters experienced incremental increases in performance within teams, along with boosts in confidence when evaluating, observing, and providing feedback (Brutus & Donia, 2010; Donia et al., 2018). There is evidence that more confident student raters also provided more future-oriented, specific teamwork skills feedback (Brutus et al., 2013), which aligns with characteristics of effective and helpful feedback as previously discussed (e.g., Fong et al., 2019; Hattie & Timperley, 2007; Kluger & DeNisi, 1996). In addition to supporting individual outcomes, these feedback systems also appear to support the team's overall performance and sense of well-being (ONeill et al., 2020). The benefits observed in higher education settings, even show evidence of effectively transferring to the workplace in the form of better organizational citizenship behaviors (Donia et al., 2018).

Quantitative Feedback

Feedback between teammates within peer evaluation systems are typically provided through one of two methods (or both): the first is through evaluations of members' teamwork skills using a structured rating scale, and the second is through open-ended comments. The latter is of primary interest to the current research, but the former is far more prevalent and well-researched, including the previously mentioned CATME system (Loughry et al. 2007), and numerous other standardized rating scales and rubrics designed to assess teamwork skills (see e.g., Hughes & Jones, 2011 for information on VALUE and AACU rubrics). The quantitative scales within these PES are well grounded in theory, and have been evaluated on their reliability and validity (e.g., Baker, 2008). The benefits of their use, some of which have been previously discussed, include the ability for students to provide structured evaluations of one another's teamwork competencies in a standardized manner. These rating scales can be implemented for a variety of purposes; for instance, they may demonstrate to students the standards they are expected to meet within the team (Hughes & Jones, 2011; Kapp, 2009), and, of course, act as a feed-back mechanisms to inform students where they could improve their teamwork skills during group projects. Receiving a rating on pre-specified teamwork competencies from several sources (i.e., their team members' ratings, and even self-ratings), allows students to cross-compare feedback, and track score changes over time (Baker, 2008). They can also provide instructors with a quick way to identify students or team members that may be struggling within the team (O'Neill et al., 2019).

Qualitative Feedback

Many PESs include the option to provide qualitative feedback to student teammates. Research suggests that such open-ended, written comments serve an important purpose for learning and development. In professional settings where performance appraisals are used, narrative, open-ended, qualitative feedback delivered by peers or supervisors is considered more closely by the recipient than quantitative feedback, and might compensate for the decontextualized, non-directive nature of numerical feedback (David, 2013; Smither & Walker, 2004). A meta-analysis of feedback in academic settings, indicates that receiving written, qualitative feedback better predicts performance and motivation as compared to merely receiving quantitative feedback (e.g., a grade) (Koenka et al., 2021).This is echoed by well-known theories of feedback, which espouse the power of qualitative feedback to provide students with richer, deeper, more specific insights regarding their behavior and behavior-change needs (Hattie & Timperley, 2007).

Harnessing Qualitative Feedback Within Peer Evaluation Systems

Generally, most of the research on PES has focused on the quantitative components of these systems in relation to individual or team outcomes. Although some studies mention that their research participants indeed made use of the qualitative components of the platform (e.g., Brutus & Donia, 2010; Donia et al., 2022). There is rarely discussion regarding the characteristics or impact of the qualitative feedback provided between teammates. Some notable exception includes Brutus et al., (2013), who analyzed qualitative features of the written feedback and found that more experienced, senior students provided feedback that had a higher valance (i.e., more positive), was more reinforcing of positive behaviors, and more specific and future oriented. Another study comes from Sridharan and Boud (2019), who analyzed how features of the qualitative comments (i.e. degree of praise vs criticism, and future-orientation vs past-orientation) mediated relationships between quantitative peer ratings and summative assessments at the end of the course. Their key findings included that qualitative comments were primarily positive and past-oriented, focused on individual characteristics and processes, and seemed to nullify the advantageous impacts of quantitative rating on performance outcomes.

The norm with these studies is that they mention providing students with the option of exchanging qualitative feedback through PES, and may even include guidance on best-practices when delivering open-ended feedback (e.g., advising that the tone of comments is constructive and respectful, the content is specific, behavior-focused, and future-oriented) (e.g. O'Neill et al., 2019), yet do not make analyses of the characteristics or impact of this feedback. Thus, it appears that qualitative feedback is an understudied and underutilized component of many PESs. One likely reason for this is that labor and time needed to draw meaningful insights from qualitative data far exceeds that required for quantitative data. In this regard, Natural Language Processing (NLP) holds potential as a tool to facilitate this process.

CHAPTER TWO

NATURAL LANGUAGE PROCESSING

NLP is a branch of the Artificial Intelligence field that enables computers to understand, analyze, and even generate human language. The capabilities of NLP are vast and range from very basic tasks like summarizing word counts, analyzing the types and frequencies of words, to advanced techniques like language translations and text generation (such as ChatGPT), NLP covers a wide spectrum of abilities. Some NLP techniques are more time-consuming than others. For instance, when NLP techniques are *supervised*, they demand more involvement and decision making from human researchers, usually in the form of labeling data to train algorithms for specific tasks. On the other hand, *unsupervised* NLP explores patterns and structures in unlabeled data without the need for human influence. The current research focuses on unsupervised NLP techniques, which may be ideal to support the generation of rapid, low-effort insights into student feedback in PES. Before describing research studies that have used NLP technologies to do just that, let us establish an overview of specific NLP techniques that will be closely considered in the current proposal.

Topic Modeling

When TM is applied to text, it can result in identification if hidden patterns, themes, or constructs known as “topics”, within the text. There are several types of TM approaches, among the most popular is Latent Dirichlet Allocation (LDA) which was first developed by Blei et al., (2003). LDA, and most TMs, are statistical models that views documents as consisting of random combinations of hidden topics that are probabilistically distributed across all the words in the document. More in-depth descriptions of LDA are covered in several sources (e.g., Vayansky & Kumar, 2020). In simple terms, LDA works by treating text documents as a so-called “bag of words,” which are analyzed to identify patterns of word co-occurrence across the documents, or bags. The entire sample of documents is termed a corpus. TM ultimately uses co-occurrence patterns to map word distributions into the topics. As an example, an effective TM would map words such as “fur”, “tail”, “bark”, “fetch”, “paw” onto the topic “dog”.

When TM is *unsupervised,* topics are discovered within documents inductively without pre-defining labels or categories, which is the case for *supervised* NLP models (e.g., classification models) which are then trained to identify these categories in new, unfamiliar documents (Mohr & Bogdanov, 2013). However, it is up to the researcher to provide qualitative labels to the topics identified by the TM algorithm, and even decide how many topics the algorithm should extract. In our example of a TM that mapped the words “fur”, “tail”, “bark”, “fetch”, “paw” as belonging to the same topic, it would be up to the researcher to assign a meaningful name or label for this topic, in this case “dog”.

<u>Sentiment Analysis</u>

Sentiment analysis (SA) is a natural language processing technique used to automatically identify, extract, and interpret the emotional tone, attitudes, and opinions expressed within text data. One of the fundamental features of SA, invovles classifying words or sentences in text in terms of how their valence, that is, how positive, negative, or neutral they are. SA can also be applied to extract more advanced insights about the text, such as what emotional categories appear to be present (e.g., anger, trust, happiness, etc.). Typically, the unsupervised approaches to SA that generates valence by using sentiment lexicons, which is essentially a list of words associated with pre-defined valence scores. In combination with other linguistic rules, the lexicon-based algorithms detect the valance of never-before seen texts (Mite-Baidal et al., 2018).

It's important to note that sentiment's accuracy heavily depends on the quality of the sentiment lexicon it uses, and that if words within these lexicons are not present in the text being analyzed, sentiment may not be accurately captured. SA may miss contextual ques present within texts (e.g., humor, sarcasm, etc.) which may lead to misunderstanding of valence and less accurate results for complex sentences (Cambria et al., 2017).

Natural Language Processing of Qualitative Student Feedback

Open-ended, qualitative data generated by students are considered information-rich, yet labor intensive to analyze and interoperate as compared to quantitative ratings of short, close-ended questions (Hodges & Stanton, 2007). Such has been the case for research evaluating qualitative feedback within PESs supporting team-skills development (e.g. Brutus et al., 2013; Sridharan & Boud, 2019), or research of qualitative feedback regarding teacher and course evaluations (e.g., Brockx et al., 2012; Yüksel & Başaran, 2020). However, advancements in NLP technology have the potential to expedite this process using quantitative algorithms to extract meaningful themes and patterns from large corpuses of text data. This is demonstrated through several studies. Sun and Yan (2023) used an NLP technique known as topic modeling (TM) to identify common topics or themes in students' feedback during teacher evaluations. These latent topics aligned with past research of students' teacher evaluations. Excitingly, nascent research on peer feedback for teamwork evaluation are also applying TM to qualitative comments within PES to extract meaningful insights. Nanda et al., (2022) preformed TM on peer feedback exchanged before and during the COVID-19 pandemic and found that latent topics had better overlap with quantitative dimensions used to evaluate team-members teamworking abilities (i.e., using the CATME rating scale) during pre-pandemic time-periods, but that additional latent topics emerged after the massive shift to virtual teaching.

Another NLP techniques known as Sentiment Analysis (SA) is increasingly being used automatically detect the valence of student feedback (for review see Pinargote-Ortega et al., 2023). SA has been applied to ascertain the positivity to negativity of comments within teacher evaluation systems left by students (Balahadia et al., 2016), and of peer evaluations in student teams using PES (Alsharif et al., 2022). In one study, SA was used to detect linguistic changes in student's peer feedback after the implementation of a feedback intervention designed to guide and improve group reflection (Leshed et al., 2007). SA has also been combined with TM approaches to identify the major themes within students' course-evaluation commentary (i.e., topic), and the valance of comments inherent to these themes (i.e., sentiment) (Gottipati et al., 2017, 2018; Nanda et al., 2021).

Topic Modeling with Short Texts

Reports suggest that Topic Modeling may be less effective in extracting meaningful topics when applied to documents with short texts, as compared to when documents are longer. This is especially the case for LDA, because LDA depends on word co-occurrence within a document to identify patterns and themes. Brevity of words the LDA algorithm with less to work with, resulting in data sparsity in short texts that is less likely to render meaningful topics (Albalawi et al., 2020). Hence, using LDA may be a limitation seeing as previous reports indicate students spent around 5-10 minutes providing feedback, including qualitative commentaries (O'Neill et al., 2019). LDA was the TM of choice in several of the studies which analyzed qualitative student feedback, specifically, Nanda et al., (2021, 2022) and Sun & Yan (2023). Out of these three studies, only Nanda and colleagues (2021) report that average length of written responses for the course-evaluation questions posed in the study: the average length of students' feedback was between 17 and 16 words, and included over 130,000 responses for each of the three questions posed. Thus, while this research had large corpuses (the total number of responses per question), the average document sizes (the individual responses) were small.

Progress is being made to identify TM techniques that are more effective for short texts, such as within the realms of social media posts (Meddeb & Romdhane, 2022), and customer reviews (e.g., Pietsch & Lessmann, 2018). One unsupervised TM approach that shows promise for short texts is Biterm Topic Modeling (BTM). BTM, developed by (Yan et al., 2013), differs from LDA in the sense that it focuses on pairs of words within documents, disregarding their order, and models the co-occurrence of word pairs. Research suggests that this helps in contexts where information is limited, like short texts; by focusing on word pairs, BTM can capture nuances that might be missed by LDA when analyzing very short or noisy text data. Thus, BTM could provide promising avenues to enhance topic modeling for the purposes of understanding what student communicate about via their prevision of qualitative feedback.

CHAPTER THREE

THE CURRENT RESEARCH

As previously discussed, both TM and SA have been successfully applied to understand student feedback in educational settings, specifically students' courses and teaching evaluations (e.g. (Balahadia et al., 2016; Gottipati et al., 2018) and, more relevant for the current research, evaluation of team members in PESs (e.g., (Alsharif et al., 2022; Nanda et al., 2022). The current research focuses on qualitative feedback data from teams of civil engineering students who are enrolled in a university curriculum that places strong emphasis on team-based learning and teamwork skills development, which is facilitated through the use of a PES designed to team learning. The current research aims to replicate the success of these previous studies by harnessing NLP techniques to investigate feedback exchanged within this context. The following sections outline the overarching research aims proposed in the current research, along with related research questions.

Decerning Latent Topics in Peer Feedback using Topic Modeling

The first aim of the current research is to use unsupervised TM to identify latent topics dormant within qualitative peer/team member evaluations. Although TM has successfully been used in similar study contexts (e.g., Nanda et al., 2022), this aim is approach cautiously because there is no guarantee that TM will result in discernable topics. While unsupervised TM has the potential to draw speedy insights into qualitative text, there may be trade-offs in time and effort if latent topics are not apparently meaningful. The effectiveness of TM depends on various factors, namely the quantity and quality of the text, the tuning of hyperparameters and text-preprocessing, and the choice of TM approach to name a few. Hence, the first research question in this study is posed simply as:

Research Question 1 – *To what extent can Topic Modeling be used to extract meaningful topics from open-ended qualitative feedback provided between team members in student teams.*

<u>Mapping Latent Topics on to Concepts Within Existing Research</u>

The second aim entails exploring whether latent topics within team-member feedback map on to existing constructs known through research of feedback quality and impact, or research of team science. This aim is pursued due to its relevance for previous research, like that of Sun and Yan (2023) who found that TM outputs of student feedback on teacher evaluations aligned with outcomes found in qualitative studies of similar contexts. Also Nanda et al., (2022), who found that topics mapped on to constructs measured thought the quantitative components of the PES, specifically within CATME tool. (see also Debortoli et al., 2016 where TM were also mapped onto constructs in existing theory). Although the aim is not to make a-priori predictions of what topics may emerge, there are some possible constructs that could feasibly emerge based on our knowledge of the research context. The following paragraphs briefly touch on some of these constructs.

Given that the qualitative feedback is generated in a classroom context based on team-based learning and focused on the development of teamwork skills, we may see constructs related to teamwork competencies, team roles, or team processes emerge as topics. In the current context, the PES utilizes *ITPMetrics* for peer feedback evaluations, which measures constructs/competencies related to *commitment*, *capabilities*, *communication, focus,* and *standards,* which will be further described in the methods. Perhaps, latent topics in qualitative feedback will reflect these topics from quantitative feedback within the PES, similar to what Nanda et al (2022) found in their research.

The team context also means that students will be interdependent towards achieving common goals, and will likely have unique roles to play in team outcomes (Kozlowski & Bell, 2003). Perhaps, emergent topics will allude to the common roles students played in their teams, which in turn may indicate their contributions to team processes. There are numerous taxonomies of team roles and associated behavioral markers (Driskell et al., 2017; Mathieu et al., 2015), as well as team processes (Crawford & LePine, 2013; Mathieu et al., 2014) which these topics could overlap with. As an example, perhaps TM reveals when peers are demonstrating behaviors associated with *action, transition,* and *interpersonal* behaviors as outlined by Marks et al.'s, 2001 taxonomy of action-transition process phases of teamwork. For instance, peers may demonstrate contributions such as goal specification (an action process), monitoring and backing up team members (a transition process), or affect management (an interpersonal process).

Given that the purpose of the PES is to facilitate peer-feedback, emergent topics may overlap with research that has de-constructed feedback characteristics, which perhaps also relate to feedback quality. Recall that research of feedback suggest that feedback may be oriented more or less towards the learner individual characteristics and behaviors vs the outcomes of a task (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Similarly, the feedback may differ in its specificity, advice and guidance, or future-orientation (Fong et al., 2019; Hattie et al., 2021). It is possible that latent topics also differ on some of these characteristics, such as being more person or task focused, the extent to which specific and actionable strategies for improvement are provided, etc.

Given that the exploratory nature of this study is not conducive to specifying a-priory hypothesis regarding which topics will emerge, yet there remains a possibility that topics will replicate previous models and theories of feedback and learning (e.g., Hattie & Timperley, 2007) and/or teamwork (e.g. Marks et al., 2001), the second research question is posed:

Research Question 2 – *Do latent topics reflect existing models, theories, or constructs within learning & feedback and/or teamwork research?*

Latent Topics and Teamwork Skills Learning

The current research aims to make a unique contribution to the literature by investigating whether receiving feedback on latent topics appear to impact students' individual learning outcomes when it comes to teamwork skills development. To this author's knowledge, research has yet to combined unsupervised TM and SA techniques with more predictive statistical analyses to model the relationship between qualitative peer/team member feedback characteristics and student learning outcomes. Thus, in addition to existing research which measures the impact of quantitative feedback provided through PES on individual learning outcomes (Donia et al., 2018), this research aims to measure the impact of receiving qualitative feedback on different topics. Such an analysis is possible because, in the research context under investigation, the PES is utilized several times during the course a semester at which points both qualitative and quantitative peer feedback is solicited (e.g., Brutus et al., 2013). This provides an opportunity to statistically analyze the relationship between the topics of qualitative peer feedback provided at time 1, and changes in composite scores quantitative teamwork competencies between time 1 and time 2.

The characteristics of latent topics are unknown until RQ1 and RQ2 are answered, making it currently impossible to hypothesize a plausible relationship between feedback topic and learning outcomes. If significant relationships between qualitative peer feedback and student learning are found, this information would be valuable for educators to know what kind of feedback to encourage within students. Hypothetically speaking, if a latent topic demarcating specific, future-focused feedback is significantly related to student learning while a latent topic demarcating generic, past-oriented comments is not, then this would provide a strong signal that educators should encourage and guide students to provide feedback that likens the former topic. In another scenario, if emergent latent topics demarcate feedback oriented towards different team roles or processes, and receiving more feedback regarding some roles and processes is more beneficial to learning outcomes, this may indicate that peer feedback is more useful in supporting individual development when it is focused on specific behaviors, roles, or processes but not others. To conclude, the third research question is posed as follows:

Research Question 3 – *Do qualitative feedback topics received by students have a significant impact on learning of teamwork skills?*

Feedback Valence and Length, and Teamwork Skills Learning

The fourth aim entails incorporate additional NLP tools such as sentiment analysis and word count techniques to assess whether there is a relationship between valence and length of qualitative peer feedback with students' learning outcomes. This aim is interesting given that feedback valence and feedback lengths have been heavily researched in relation to feedback quality and impact indicators. Yet, research has yet to investigate whether the valence and length of feedback received from team members through PES has an impact on teamwork skills learning. Should significant relationships be found, this may provide educators with further guidance on how rubrics and requirements students might want to follow when providing their peers with feedback, to maximize its effectiveness.

When it comes to whether positive or negative valence is impactful for peer learning, the research discussed in this proposal was mixed (i.e., see page 9 within this document), and largely determined by subjective ratings of usefulness. Based on previous research, it appears both positive and negative feedback may have benefits on motivation and learning (Anson & Anson, 2017; Wind & Jensen, 2017). Taken together, a curvilinear relationship between the valence of feedback and learning outcomes is hypothesized. Specifically, overly positive and overly negative feedback is predicted to have a detrimental impact on learning outcomes, and learning outcomes will improve as feedback valence approaches a more neutral point on the spectrum. In addition, the current research hypothesizes that feedback length will have a significant, positive main effect on increased learning, as has been demonstrated consistently in past studies (e.g.

Häkkinen & Ramadan, 2023; Wessa & De Rycker, 2010; Zong et al., 2021). There might be, in addition, an interaction between the feedback length and feedback valence, such that feedback that is longer and positive or longer and negative is more effective than feedback that is shorter and positive or shorter and negative. In sum, the following research question and hypothesis is posed:

Hypothesis 1 – *Feedback valence will demonstrate a curvilinear relationship with students' learning of teamwork skills such that learning increases as feedback approaches neutral valences.*

Hypothesis 2 – *Students who receive feedback with a greater word count will demonstrate increased learning on teamwork skills than students who receive shorter feedback.*

CHAPTER FOUR

RESEARCH CONTEXT

Sample & Data Collection

This research was conducted at a large academic university in the southeastern United States. After excluding individuals who provided verbatim feedback to all members of their team, indicative of putting little thought or effort in providing unique feedback, the sample included 516 civil engineering undergraduate students and 135 teams with the team size being 3-4 students on average. These students partook in a curriculum spanning the sophomore, junior, and senior years of civil engineering course program. Some of the students within this sample were apart of several teams in different classes throughout the years.

This innovative curriculum model was supported by an NSF grant entitled Revolutionizing Engineering Departments, where the primary objective is to provide students with a comprehensive understanding of the broader aspects of civil engineering while fostering the development of essential skills including professionalism, communication, and teamwork. Courses are heavily focused on semester long hands-on, real-world projects undertaken in teams. The average team size consisted of 4 members.

Twice during a semester, students provided feedback about the performance of their fellow team members using the Individual and Team Performance (ITP) Metrics platform (itpmetrics.com). The platform permits feedback to be given on quantitative scales related to team-member performance, and qualitative, open-ended feedback. When it comes to providing qualitative feedback, some minor instructions on how to provide feedback in a courteous, specific, and respectful manner was displayed to students in the system (Appendix A). Feedback was reviewed by instructors to ensure that no obscene, overtly offensive, or derogatory comments have been left by students. Feedback is anonymized and distributed to feedback recipients. This occurs during the semester mid-term and final, typically separated by around 6 weeks.

Quantitative and Qualitative Data

ITP Metrics Peer Feedback

ITP Metrics has various capabilities which include the ability to measure one's conflict management style, reflect on various aspects of one's team's performance. The current research utilizes ITP Metrics Peer Feedback survey, permit students to give and receive feedback to their teammates. This survey was adapted from, and validate against, the Ohland et al.,'s (2012) BARS (behaviorally anchored rating scale) version of the Comprehensive Assessment for Team Member Effectiveness (CATME) (O'Neill et al., 2019), which was originally developed by Loughry et al., 2007. The original CATME was developed through extensive literature searches to identify the most relevant clusters of contributions team-member make, and the survey was validated through exploratory and confirmatory factor analysis in two large samples of students. Ohland et al., (2012) provided additional research to confirm the validity and reliability of this 5-dimension CATME survey. The survey dimensions include *commitment* to the team's work; *capabilities* such as having relevant knowledge, skills and abilities relevant to the team's goals; effective *communication* within the team; *focus* such as keeping the team on track towards their goals; *standards* as in emphasizing and expecting high quality within the team (Appendix B) (O'Neill et al., 2019). Within ITP Metrics, students rate one another on a scale of 1 (anchored as 'to no extent') to 5 (anchored as 'to a considerable extent') on five separate teamwork competencies.

ITP Metrics automatically generates average scores across peer-ratings for each of the 5 domains, as well as a grand composite score that aggregated ratings across all dimensions. The extent of students' learning and development regarding their teamwork skills, as rated by their peers, will be operationalized the composite score at the finals while controlling for the composite score at the mid-terms.

Text Data

As previously mentioned, students received qualitative feedback from each of their peers through the IPT Metrics platform (Appendix A). Since our participant sample consists of 516 civil engineering students belonging to teams averaging 3 – 4 members, the number of there were a total of 1385 text entries or feedback entries from participants. Hence, the size of our corpus for NLP consisted of 1385 documents, after removing participants who provided the same feedback for all of their team members. Taking the average of document lengths within the corpus of 1385 documents revealed that the average length of peer feedback was 27 words.

CHAPTER FIVE

PROCEDURE AND ANALYSIS

This research relies on qualitative data for NLP, and quantitative data for statistical analysis. Both NLP and statistical analyses were facilitated using the programming language and software environment, "R.". Table 1 provides an overview of the data and analysis being used to address each research question and hypothesis.

Table 1

Overview of Data and Analytical Approach Used for Each Proposed Research Question and Hypothesis

	Data	**Analytical Approach**
Research Question 1	Qualitative Feedback Data	Biterm Topic Modeling, using semantic coherence and exclusivity trade-off scores for model selection and reliability measures for model validation.
Research Question 2	Qualitative Feedback Data	Literature review to cross-reference latent topics with existing research.
Research Question 3	Latent TM Topics ITP Metrics Peer Feedback Survey	Using mixed-effects multiple linear regression, where the independent variables are latent topics gathered at time 1, and the dependent variable is scores in peer-rated teamwork skills at time 2 controlled for at time 1.
Research Question 4 & Hypothesis 1	Qualitative Feedback Data ITP Metrics Peer Feedback Survey	Using mixed-effects polynomial regression, where the independent variables are word count of qualitative feedback and quadratic valence scores of qualitative feedback at time 1, and the dependent variable is scores in peer-rated teamwork skills at time 2 controlled for at time 1.

Biterm Topic Modeling

The current research utilized BTM via the R package “btm” (Wijffels & Yan, 2023) due to the superior ability of this modeling approach when dealing with short text documents.

Text Pre-Processing

The text data underwent pre-processing practices commonly used to reduce variability and dimensionality in NLP (Maier et al., 2018). Text cleaning and pre-processing entailed removing stop-words entailed tokenization, which involves splitting the text into individual words or tokens. Tokens are converted to lowercase. English stop-words, (e.g., "and", "the", "in"), were removed, along with symbols, punctuation, and white spaces. Additionally, removal of custom stop-word was conducted, which mainly included names of participants mentioned in the text data along with words the words “work”, “job”, “good”, and “great” which were highly frequent within the corpus and posed challenges to topic distinction and interpretation.

Pre-processed data was converted into a Document-Feature Matrix (DFM), which represents documents as rows and tokens as columns, showcasing their frequency in each document. This DFM then underwent lemmatization. Lemmatization transforms words into "lemmas", which is their base or dictionary forms. For example, the word "running" is lemmatized to "run." Lemmatization may enhance text analysis by providing precise and interpretable word forms (Denny & Spirling, 2018).

Biterm Specifications

As previously mentioned, a biterm is a pair of words co-occurring in a short context window within a document. BTM permits us to specify which words are of most interest to consider in biterms, and which size to set the context window (Yan et al., 2013). For this research, "nouns", "adjectives" and "verbs" were specified as relevant biterms. BTM also permits adjustment of the word co-occurrence window, thus specifying how far apart biterm pairs are permitted to be. Window size may have different implications for BTM modeling, since a smaller window would capture more immediate word relationships, whereas a larger window captures a broader context but might include less directly related word pairs.

Several BTM models were tested, each with varying word co-occurrence windows specified. Specifically, a window of 5, 10, 15, and 26 was tested. The value of '26' was specifically tested since the median length of a student's feedback was 26 words. Thus, a word co-occurrence window considers the entire length of the average comment left by students for their peers. Furthermore, previous research suggests that this is around the number of words (i.e., between 20 - 29 words) that it takes to express singular thoughts in oral or written form (Einhorn, 1978).

Model Search, Evaluation, & Validation

An important part of topic modeling is how many topics (i.e., *K*) should be included in the model. Often, researchers must test and compare several models under different conditions and in an iterative process before arriving at an ideal Topic Model (Zhao et al., 2015). The process can be aided by running additional, goodness-of-fit, analyses, such as topic coherence which provides scores indicating how semantically meaningful and coherent the words within a topic are, and thus how well they form a meaningful theme (e.g. Stevens et al., 2012). Unfortunately, the BTM package does not currently include built-in functions for computing and comparing such goodness-of-fit values across models within varying topic. This is unlike other topic modeling R packages, such as STM which has the built-in ability to identify optimal topics based on these calculations (Roberts et al., 2019). Fortunately, (Bittermann and colleges (2021) have developed R code by for their research on twitter mining for psychological concepts, that permits evaluation of models via exclusivity and semantic coherence trade-off scores. They follow best practices outlined by Maier et al. (2018) for semantic coherence and exclusivity evaluations, and includes an additional evaluation measure of topic reliability, which Bitterman et al. (2021) adapted from research by (Niekler, 2016) Niekler (2016). Topic reliability entails generating several models (in this case 10) with the same number of topics but with different Gibbs initializations (see Griffiths & Steyvers, 2004), and calculating cosine similarities between the top words in each of the topics within the original topic models, and each of the models generated from the random seeds. Specifying a threshold of 0.8 would identify which of the topics replicable 80% of the time over different initializations.

R code and BTM selection approach was adapted for the current research. BTMs with a *K* equal to between 5-11 topics were considered during the search for an ideal model. This range was chosen based on preliminary analyses, which suggested that a model within this range would provide the most parsimony and interpretability. Models exceeding 11 topics were considered too difficult to comprehend in an applied setting. The model with optimal number of *K* was underwent further validation using Bitterman et al.'s (2021) approach model reliability evaluation . Only the topics that demonstrated 80% replicability of the 5 top words within topics across 10 initializations were included in the final model.

An additional metric used to guide the selection of the final model was the distribution of topic representation across the corpus, which is indicated by topic *theta*. More specifically, *theta* values represent the distribution of topics for a single document, a subset of documents, or all of the documents within a corpus. The latter indicates the relative distribution of topics in over the whole corpus. Models with a more even distribution of topics were preferred, as opposed to models where some topics were much more highly represented within the text than others. It was felt that a model with a more even topic distribution would aid in the reliability of future regression analyses theta scores.

Model Comparison and Semantic Interpretation

A final model was selected based on the scores of model quality and topic validation, along with qualitative analysis of the topic's semantic interpretability. A handful of candidate models with similar quality scores were evaluated and, where appropriate, compared. This process entailed inspecting the 15 top words representative of the topics, as well as the documents most highly representative of the topic. Topic representation or proportion within a document is indicated by high theta values.

By analyzing the most frequent words in a topic, along with the documents highly associated with a topic, it is possible to see whether there are coherent themes across documents with similarly high theta scores for a particular topic. Model interpretability was considered poor if topics appeared to have high overlap, marked by repetition of top words and similar themes represented top documents. Models were also considered poor if there were too few topics, or if it was difficult to interpret the embedded themes and ascribe a succinct label to the topic.

Description & Labeling of Final BTM

Labels were ascribed to each of the topics in the final topic model by close evaluation of highly representative documents. Themes that persisted across most documents were considered the primary descriptions of the topic, while themes that ran across some of the documents were considered secondary topic descriptions. The effort was made to consider a mix of highly representative documents in terms of both positive feedback and constructive criticism, as well as past-oriented and future-oriented feedback.

Comparison of Final BTM with Existing Frameworks

Once topics were labeled and described, the model was compared with existing frameworks and models of teamwork and feedback. Specifically, the constructs measured in ITP Metrics quantitative peer evaluation scales (O'Neill et al., 2019), and the constructs represented in Marks et al.'s (2001) team framework of transition and action processes were mapped on to topics where fitting. These frameworks are respectively depicted in Appendix B and Appendix C. Topics were also interpreted in light of Hattie and Timperley's (2007) model of effective feedback, depicted in Appendix D.

Valance Analysis & Word Count

Valence analysis was conducted using the *sentimentr* (Rinker, 2021) in R, which has reportedly outperforms other packages when calculating valence at the sentence and document level (Misuraca et al., 2020; Naldi, 2019). Students' feedback was first separated by sentence in order to obtain, so-called, sentiment score calculations for each sentence in the text data. Sentiment scores which fall into three classes: positive (> 0), neutral (= 0), and negative (< 0). Sentiment scores were then averaged across sentences belonging to single documents, hence single sentiment scores were created for each of the 1385 text entries or feedback entries included in the data. *Sentimentr* simultaneously generated word count for the each of the 1385 text documents.

Regression Analysis

Relationship of Feedback Topic to Student Learning Outcomes

To investigate the relationship between feedback topics and subsequent student learning outcomes, a mixed- effects multiple regression analysis was employed. The independent variables were averaged theta scores for each identified topic across all documents (i.e., feedback) associated with a single participant. Hence, average theta scores for each of the six topics in the BTM were produced for each of the 516 participants in the sample. These averaged theta scores represented the prevalence of each topic in the student's feedback, as provided by their team-mates/peers. The dependent variable was the students' scores on ITP Metrics peer-feedback of teamwork competencies average across all team members, taken at the course finals. The rating provided at the mid-term was included as a control variable. Due to the hierarchical nature of the data where individuals belong to teams, a mixed-effects model was chosen to account for within-group (fixed effects) and between-group (random effects) variations.

Relationship of Feedback Valence and Length to Student Learning Outcomes

To tests relationships between the valence of feedback received by students, the length of feedback received, and subsequent learning outcomes, a mixed- effects polynomial multiple regression analysis was employed. Here, independent variables included the average of feedback valence scores and word-count for each of the 516 unique participants. To test hypothesis 1, the regression included a quadratic term of valence score. The dependent variable was the students' scores on ITP Metrics peer-feedback of teamwork competencies average across all team members, taken at the course finals.

Table 2

Overview of Quantitative and Qualitative Comparison and Evaluation of Top Biterm Topic Model

Word Cooccurrence Window	Best Performing Between *K* 5-11	Semantic Coherence Score	Exclusivity Score	Semantic Coherence Exclusivity Trade-off	Number of Topics Reliably Replicated	Topic Theta Scores	Interpretability
5 words	10	-17.224	27.004	0.454	5	T1: 0.322, T2: 0.109, T3: 0.053, T4: 0.157, T5: 0.091	Fair. Some topic overlap
10 words	10	-18.280	28.269	0.597	6	T1: 0.335, T2: 0.063, T3: 0.093, T4: 0.061, T5: 0.106, T6: 0.122	Good
15 words	10	-17.110	27.634	0.513	4	T1: 0.100, T2: 0.101, T3: 0.259, T4: 0.061	Too few topics
26 words	9	-17.741	27.149	0.655	6	T1: 0.131, T2: 0.100, T3: 0.138, T4: 0.100, T5: 0.092, T6: 0.289	Good

CHAPTER SIX

RESULTS

Research Question 1

To address the first research question, *to what extent can Topic Modeling be used to extract meaningful topics from open-ended qualitative feedback provided between team members in student teams,* several BTM models with different word cooccurrence windows (i.e., 5, 10, 15, 26) were created and compared. Table 2 reports the evaluation metrics for the best performing models for each of the word cooccurrence model iterations, the number of topics that were reliably replicated for these models, and the interpretability of these topics. Each model had decent interoperability, but model with the word cooccurrence window of 5 had topic overlap, while 15 had too few reliable topics. Models with word cooccurrence windows of 10 and 26 had the highest quality scores, and both produced 6 reliable topics. When compared, both models had interpretable topics that were, in fact, very similar across models. Ergo, roughly the same topics were replicated in word cooccurrence windows of 10 and 26. The model with the word cooccurrence of 26 was retained as the final model due to the slightly elevated semantic coherence and exclusivity trade-off score (i.e., 0.655) and the slightly more even distribution of topic theta (see Table 2).

The 6 topics that emerged within our final BTM in turn include *Timely Communication, Idea Generation, Coordination & Adaptation, Work Quality, Team Support & Focusing,* and *Work Accountability.* Each of the topics demonstrated similar representation within the text data, aside from Work Accountability which represented to a slightly higher extent (see Table 2). Table 3 demonstrated each of these topics in term of the top 10 words and their proportion within the topics (i.e., beta values), several examples of feedback representative of the topic as indicated through high theta values, and descriptions of the topics' primary and secondary foci. Both a primary and secondary description of the topics were included to account for both the central unifying theme across examples, and secondary themes mentioned. In conclusion, it appears that topic modeling, to a sizable extent, can be used to extract meaningful topics from open-ended qualitative peer feedback between student team members.

Research Question 2

The second research question asked whether *latent topics reflect existing models, theories, or constructs within learning & feedback and/or teamwork research?* Table 4 demonstrates which aspects of the (O'Neill et al., 2019) and Marks et al. (2001) teamwork frameworks, as well as the Hattie and Timperley (2007) framework of feedback appear to overlap with topics in the BTM. The following section describe the results of the BTM topics overlap with, first, the teamwork frameworks and, second, the feedback framework.

Topic Overlap with Teamwork Concepts

Cross-comparison of the BTM topic description and feedback examples with constructs in the teamwork frameworks suggest that each of the topics has demonstrable overlap with elements in both frameworks. These findings were validated through conversations with SMEs in teams research. Furthermore, a single topic from the BTM may demonstrate elements of several constructs within a singular framework. For instance, the Idea Generation topic has aspects of “transition”, “action”, and “interpersonal” processes as described by Marks et al. (2001).

The same is seen for O’Neill et al.’s peer-feedback assessment framework; the Coordination & Adaptation topic overlaps with “focus” , “communicating with team members”, “capabilities”, and “commitment to team’s work”. It is interesting to note that each of our topics did not generally overlap with every sub-element of the concepts in O’Neill et al.’s. For example, while the Timely Communication topic overlapped with two elements of the “communicating with team members” construct: “exchanges information with teammates in a timely manner” and “effectively communicates and openly shares information”, the Work Accountability topic overlaps solely with effectively communicates and openly shares information.

It appears that, when it comes existing teamwork theories and constructs, the topics found by our BTM have substantial overlap. However, singular BTM topics do not map exclusively or uniformly on to topics in either team framework. Instead, the topics appear to represent a mixture of concepts represented in these frameworks. This is perhaps unsurprising given that qualitative feedback is known to be nuanced. It appears that peers consider both “action”, “transition”, and “interpersonal” aspects of their team member’s work and contributions, when providing feedback.

It is interesting to not that some elements of both teamwork frameworks under consideration were not clearly addressed by the topics. For Marks et al., “mission analysis”, described as the “interpretation and evaluation of the team's mission,

Table 3

Label, Top Terms and Betas, Feedback Examples and Thetas, and Primary and Secondary Description of the Six BTM Topics

Topic label	Top Lemmatized Terms & Beta Values	Feedback Examples with Topic Theta Value	Topic Description
Timely Communication	need (0.04) timely (0.03) manner (0.03) information (0.03) question (0.03) ask (0.02) help (0.02) complete (0.02) member (0.02) time (0.02)	"timely response to messages and requests for data. reliable" (0.59) "At times did not respond to questions via email or text in a timely manner. Could be clearer in conveying information. Did an [sic] good job conveying information in the report and presentations." (0.52) “You did a great job at helping me understand specifics, and exchanging information to ensure the drawing specifications match architectural specifications." (0.45) "Could use work in developing more knowledge regarding the subject, be more communicative" (0.39)	**Primary description:** timeliness of team members' information communication and the ability to convey information effectively. **Secondary description:** ability to communicate about subject matter expertise and complete work in a timely manner.
Idea Generation	idea (0.04) project (0.03) design (0.03) need (0.02) help (0.02) think (0.02) lot (0.02) communicate (0.01) keep (0.01) get (0.01)	"Dresses very spiffy, communicates well, strong speaker, provides great feedback on parking lot designs" (0.67) "The glue that keeps the group together, brings great ideas for the group parking lot design" (0.49) "[Name] is good at problem-solving and often provides great suggestions for ways to improve our methodology." (0.48) "Enthusiastic about projects, and provides new ideas during group discussion and project planning." (0.47)	**Primary description:** the generation of ideas and solutions, to solve problems and find solutions for the team project. **Secondary description:** Topic also involves bringing positivity and professionalis m to conversations and providing feedback.

Table 3 (continued)

Coordination & Adaptation	design (0.04) site (0.03) need (0.03) make (0.03) able (0.02) project (0.02) change (0.02) work (0.02) help (0.01) get (0.01)	"Able to be flexible with short notice changes." (0.59) "[Name] did a great job with the site layout and communicating changes and updates." (0.47) "[Name] did a great job getting information about the site from [Name] and staff at clemson university. He was able to get occupancy data from fike which has been useful for the design. He has also been working closely with the other members when we needed to change some aspects of the design or coordinate features on the site." (0.40) "Impressed with how quickly you were able to adapt and implement changes in the site design into your own design." (0.40)	**Primary description:** Topic primarily describes the ability to adapt their work in response to needed project changes, and to effectively communicate to the team about changes. **Secondary description:** ability to coordinate with others, both inside and outside the team, to address needed changes.
Work Quality	report (0.04) time (0.04) conceptual (0.03) presentation (0.02) help (0.02) think (0.02) lab (0.02) member (0.01) class (0.01) meeting (0.01)	"-Assists team in identifying problems/solutions; improves quality of our deliverables. -Procrastination negatively impacted conceptual design report" (0.66) "[Name] did a great job preparing the structural and architectural layouts and had very captivating slide transitions/animations. Also, he gave a great presentation and handled very well under the time pressure when turning the report in." (0.42) "He did an excellent job participate in team work and discussion and submit lab reports and all" (0.42) "I think that some of your work thus far has slightly impeded the quality of the work we've been submitting. I think dedicating more effort into quality assurance will go a long way in aiding our group. (especially in the context of lab reports)." (0.40)	**Primary description:** the extent that team members produce high quality work, particularly regarding project deliverables such as reports, presentations, and meeting minutes. **Secondary description:** participation in discussion and contributing to solution seeking and problem resolution.

Table 3 (continued)

Team Support & Focusing	help (0.03) keep (0.03) task (0.03) lab (0.02) project (0.02) make (0.02) take (0.02) sure (0.02) member (0.02) appreciate (0.02)	"Leader. Excellent employee. Keeps us progressing forward." (0.68) "Team player, helped with meeting minutes and other group tasks including creating sheets and organizing docs. " (0.60) "[Name] does a good job assisting on all the labs, especially the autoCAD centered labs where he helps [Name] follow the instructions through AutoCAD" (0.45) "[Name] is also helpful in keeping the group on task during labs and tries to keep progress moving when slow downs occur. [Name] takes initiative in assigning work tasks and completing assignments without being asked/told to. Overall, [Name] is very effective in keeping the group organized and on task." (0.43)	**Primary description:** the extent to which team members supported the team and kept the team on track through activities related to management, organization, and helping behaviors. **Secondary description:** specific activities related to taking meeting minutes, organizing and managing files, taking initiative on tasks, supporting others with tasks, and boosting morale.
Work Accountability	make (0.03) time (0.03) get (0.03) need (0.02) sure (0.02) complete (0.02) help (0.02) keep (0.02) project (0.01) work (0.01)	"Needs to work on managing his time and remembering due dates." (0.50) "Always made sure the team knew your progress and went through your changes with the team. Did a good job staying on track and made sure to complete action items in time" (0.40) "[Name] was great in maintaining his individual deadlines and helping to organize submittal documents in a timely fashion." (0.35) "[Name] completed his individual requirements, but I think the teamwork requireme nts could've been more equally distributed. He stays quiet rather often, but with the level of interdependencies in this project, team and inter-team [comment cuts off]" (0.32)	**Primary description:** the extent to which team members stayed accountable to the team and met expectations by completing individual-level tasks and requirements on time. **Secondary description:** the extent that team members contributed to team-level as well as individual-level group needs.

formulation, and main tasks as well as planning operative environmental conditions and team resources available for mission execution" was not clearly covered. Perhaps, this relates to the timing of the feedback solicitation period, which occurred at the mid-term where many of the team projects would be completed. This would be well past the point of class teams' initial planning phase, where "mission analysis" might be most relevant. Indeed, since "action" processes appeared to be addressed more frequently within the topics, these processes might have been considered as being more important to comment on during peer feedback. Another construct in Marks et al.'s framework that was not addressed was "conflict management", described as "establishing conditions to prevent, control, or guide team conflict before it occurs and working through task and interpersonal disagreements among team members." Although feedback did touch on topics related to finding solutions to problems, no topic directly addressed conflict management. Perhaps, this was not an element that peers felt comfortable or relevant to provide feedback on, or the topic was not prevalent enough to emerge in our model.

Regarding the O'Neill et al. framework, each of the high-order constructs overlapped with the BTM. However, not each of the lower-order elements of each construct was directly addressed within the BTM. Namely, for the "communicating with team members" construct the sub-elements of 'requests feedback regularly and incorporates feedback from team members', and 'seeks appropriate team input before taking action' was not strongly represented . Under the "capabilities" construct, "seeks to gain the knowledge, skills, and abilities needed by the team" and "learns about other teammates' tasks and roles" was not strongly represented. Under the "commitment to the team’s work” the “prepared for team meetings” was not directly represented.” Under “emphasizing high standards” both “shows confidence in the team’s ability to perform” and “believes that the team will achieve high standards” was not directly represented.

Topic Overlap with Feedback Concepts – Future & Past Orientation

(Hattie & Timperley, 2007) feedback model posits that effective feedback involves both “feedback” and “feedforward” which respectively targets past behaviors and their effectiveness in achieving a goal, and beneficial behaviors and changes to make in the future. Cross-comparison of the BTM topic description and feedback examples with the Hattie and Timperley's (2007) framework and discussion with SMEs, (i.e., university course teachers and researchers), suggest that topics in the topic model do not reliably describe whether peer evaluation targeting past behaviors (i.e., “feedback”) or future behaviors (i.e., “feedforward”) based on their descriptions. However, examples of “feedback” comments appear to be much more numerous than “feedforward” comments. For an example, take the feedback highly representative of the Coordination & Adaptation topic: *“[Name] did a great job with the site layout and communicating changes and updates.”* (see Table 3).

Despite having made the effort to find representations of both past-oriented and future-oriented comments, as well as positive feedback and constructive criticism, the examples provided in Table 3 are predominantly of a past-oriented nature. This is especially the case for Idea Generation, Coordination & Adaptation, Team Support & Focusing. However, a higher proportion of future-oriented constructive criticism appears to be present in feedback targeting: Timely Communication (e.g., *“Could use work in*

Table 4

Demonstration of Topic Overlap with Components of the O'Neil et al., Marks et al., and Hattie & Timperley Frameworks

Topic label	Topic Description	O'Neill et al. (2019) Peer Feedback Survey Constructs	Marks et al.'s (2001) Taxonomy of Team Processes	Hattie & Timperley (2007) Model of Effective Feedback
Timely Communication	**Primary description:** timeliness of team members' information communication and the ability to convey information effectively. **Secondary description:** ability to communicate a bout subject matter exper tise and complete work in a timely manner.	**Communicating with team members**: • "Exchanges information with teammates in a timely manner." • "Effectively communicates and openly shares information." **Capabilities:** • "Demonstrates the capabilities needed for the team to perform."	**Action - Coordination:** • "Orchestrating the sequence and timing of interdependent actions."	Some indication of future-oriented, constructive feedback present. **Self-regulation Level** • "Self-monitoring, directing, and regulating of actions." **Process Level** • "The process needed to understand/perform the task(s)."
Idea Generation	**Primary description:** the generation of ideas and solutions, to solve problems and find solutions for the team project. **Secondary description:** Topic also involves bringing positivity and professionalism to conversations and providing feedback.	**Communicating with team members**: • "Effectively communicates and openly shares information." **Emphasizes high standards:** • "Encourages and motivates the team." **Focus:** • "Provides meaningful, growth-oriented, and regular feedback to members."	**Transition - Strategy formulation and planning:** • "Development of alternative courses of action for mission accomplishment." **Action - Team monitoring and backup:** • "Providing a teammate verbal feedback or coaching." **Interpersonal - Motivating and confidence building:** • "Generating and preserving a sense of collective confidence, motivation, and task-based cohesion with regard to mission accomplishment." **Interpersonal - Affect management**: • "Regulating member emotions during mission accomplishment."	Predominantly past-oriented, positive feedback. **Process Level** • "The process needed to understand/perform the task(s)." **Self level** • "Personal Evaluations and affect (usually positive about the learner)."

Table 4 (continued)

Coordination & Adaptation	**Primary description:** Topic primarily describes the ability to adapt their work in response to needed project changes, and to effectively communicate to the team about changes. **Secondary description:** ability to coordinate with others, both inside and outside the team, to address needed changes.	**Focus:** • "Monitors issues that may effect the team and notices problems." • "Helps the team plan and organize work and anticipates issues." **Communicating with team members**: • "Effectively communicates and openly shares information." **Commitment to the team's work:** • "Contributes appropriately to the team's work." • "Demonstrates commitment to the team's work." **Capabilities:** • "Demonstrates capabilities needed for the team to perform." • "Seeks to gain knowledge, skills, and abilities needed by team."	**Transition - Strategy formulation and planning:** • "Development of alternative courses of action for mission accomplishment." **Action - Systems monitoring:** • "Tracking team resources and environmental conditions as they relate to mission accomplishment." **Action - Monitoring progress toward goals:** • "Tracking task and progress toward mission accomplishment, interpreting system information in terms of what needs to be accomplished for goal attainment, and transmitting progress to team members."	Predominantly past-oriented, positive feedback. **Task Level** • "How well the tasks are understood/performed." **Process Level** • "The process needed to understand/perform the task(s)." **Self-regulation Level** • "Self-monitoring, directing, and regulating of actions."

Table 4 (continued)

Work Quality	**Primary description:** the extent that team members produce high quality work, particularly regarding project deliverables such as reports, presentations, and meeting minutes. **Secondary description:** participation in discussion and contributing to solution seeking and problem resolution.	**Commitment to the team's work:** • "Contributes appropriately to team's work." • "Keeps deadlines and delivering complete, accurate work." **Capabilities:** • "Demonstrates capabilities needed for the team to perform." • "Seeks to gain knowledge, skills, and abilities needed by team."	**Transition - Goal specification:** • "Identification and prioritization of goals and subgoals for mission accomplishment."	Some indication of future-oriented, constructive feedback present. **Task Level** • "How well the tasks are understood/performed." **Process Level** • "The process needed to understand/perform the task(s)." **Self-regulation Level** • "Self-monitoring, directing, and regulating of actions."
Team Support & Focusing	**Primary description:** the extent to which team members supported the team and kept the team on track through activities related to management, organization, and helping behaviors. **Secondary description:** specific activities related to t aking meeting minutes, organizing and ma naging files, taking initiative on tasks, supporting others with tasks, and boosting morale.	**Commitment to the team's work:** • "Contributes appropriately to the team's work." • Demonstrates commitment to the team's work." **Emphasizes high standards:** • "Encourages and motivates the team." • "Cares about the quality of team's work."	**Action - Team monitoring and backup:** • "Helping a teammate behaviorally in carrying out actions, or assuming and completing a task for a teammate." **Action - Motivating and confidence building:** • "Generating and preserving a sense of collective confidence, motivation, and task-based cohesion with regard to mission accomplishment." **Interpersonal - Affect management:** • "Involves regulating member emotions during mission accomplishment, including (but not limited to) social cohesion, frustration, and excitement."	Predominantly past-oriented, positive feedback. **Task Level** • "How well the tasks are understood/performed." **Process Level** • "The process needed to understand/perform the task(s)." **Self-regulation Level** • "Self-monitoring, directing, and regulating of actions." **Self level** • "Personal Evaluations and affect (usually positive about the learner)."

Table 4 (continued)

Work Accountability	**Primary description:** the extent to which team members stayed accountable to the team and met expectations by co mpleting individual-level tasks and requirements on time. **Secondary description:** the extent that team members contributed to team-level as well as individual-level group needs.	**Commitment to the team's work:** • "Demonstrates commitment to the team's work". • "Keeps deadlines and delivering complete, accurate work." **Communicating with team members**: • "Effectively communicates and openly shares information."	**Action - Monitoring progress toward goals** • "Tracking task and progress toward mission accomplishment, interpreting system information in terms of what needs to be accomplished for goal attainment, and transmitting progress to team members."	Some indication of future-oriented, constructive feedback present. **Process Level** • "The process needed to understand/perform the task(s)." **Self-regulation Level** • "Self-monitoring, directing, and regulating of actions."

developing more knowledge regarding the subject, be more communicative"; Work Quality (e.g., *"… I think dedicating more effort into quality assurance will go a long way in aiding our group. (especially in the context of lab reports)"*) and Work Accountability (e.g. *"Needs to work on managing his time and remembering due dates."*).

Topic Overlap with Feedback Concepts – Feedback Level

Hattie and Timperley's (2007) framework showcases that feedback (either future or past-oriented) may target the level of the task, the process, a person's self-regulation, or the person themselves. Consensus from SMEs suggests concluded that none of the six topics within the BTM cleanly overlap with any of the four levels of Hattie and Timperley's model. Instead, BTM topics appear to contain a mixture of several of the four levels. Furthermore, several levels may at times be communicated by a single feedback comment. Take for example the final example in for Team Support & Focusing (see Table 3). The first sentence of this feedback, (i.e., *"[Name] is also helpful in keeping the group on task during labs and tries to keep progress moving when slow downs occur"*) appears to target the process through which the team members supports in task completion. The second sentence (i.e., *"[Name] takes initiative in assigning work tasks and completing assignments without being asked/told to*") appears to be commenting on the self-regulation level, since they discuss the team member's ability to effectively self-regulate in terms of completing tasks without outside prompting. The final sentence in this feedback (i.e., *Overall, [Name] is very effective in keeping the group organized and on task"*) provides feedback of the task-level by a simple appraisal of the team member's performance regarding the task of keeping the group organized.

At a high level, topics in the BTM demonstrate some greater overlap with certain of Hattie & Timperley's 4 feedback levels than others. Timely Communication appears to target the self-regulation level in terms of team member's tendency to (not) communicate promptly, and the process level in terms of the manner and (in)effectiveness of communication. The Idea Generation topic overlaps with the process level in the sense that team members put their ideas and solutions forward towards task achievement, and this topic overlaps with the self level due to the element of personal evaluations regarding the team members' positive attitudes. The Coordination & Adaptation topic appears to contain feedback about self-regulation in terms of the team members' ability to adapt their actions in response to changes, and feedback about the process and task level in terms of describing how well team members understood what was needed in order to achieve specific, newly defined tasks. Work Quality demonstrates task and process level feedback in terms of commenting on how and why team members produced high quality (or didn't) in relation to specific tasks, and self-regulation feedback is implicated when comments discuss self-how team members monitored and directed their actions (or didn't) to do so. Team Support & Focusing also demonstrate task, process, and self-regulation level feedback targeting members' the methods and abilities to achieving specific tasks that ultimately helped the process of keeping the team on track, and self level due to personal praise about positive or helpful team member characteristics. Finally, the Work Accountability topic appears to mainly target the self regulation level since it provides feedback on the impact of team members' conduct and behavior assisted in (un)successful staying on target, and the process level when examples were given of how these actions aided (or did not aid) in task achievement.

Research Question 3

The final research question asked whether *qualitative feedback topics received by students had a significant impact on learning of teamwork skills.* The mixed-effects multiple regression suggests that there is no significant relationship between any of the six topics provided by peers as feedback at time 1 and students' teamwork abilities as measured by peers at time 2, when controlling for teamwork abilities as measured by peers at time 1 (see Appendix E). The interclass correlation value of 0.324 suggests that 32.4% of variance in the model is attributable to differences between teams, confirming the appropriateness of using a mixed-effect model where group ID was placed as a random effect. Despite having non-significant effects, the model as a whole accounts for 69.8% of variance seen in the dependent variable (i.e., *Pseudo-*$R^2 = 0.698$), and just the fixed effects, or interdependent and control variables account for 55.3% of variance (i.e., *Pseudo-*$R^2 = 0.553$). Tests of assumptions of normality of residuals suggest that data was normally distributed but contained some outliers (Appendix F) , while assumptions of homoscedasticity of residuals appeared to be violated (Appendix G).

Comparison of Peer Feedback Scores Between Time 1 and Time 2

The results of the third research question suggest no significant relationship between the qualitative feedback received by peers at time 1, and quantitative peer ratings at time 2. The results suggested a significant relationship existed between the control variable (averaged scores of peer feed survey at time 1) and the dependent variable (averaged scores of peer feed survey at time 2) (see Appendix E). To further explore this relationship, a paired samples T-test was conducted to test whether peer feedback survey scores changed between the midterm and finals. The results indicate no statistically significant difference between scores at time one (mean = 4.44) and scores at time 2 (mean = 4.47), i.e., $t(514) = -1.4747$, $p = 0.1409$). The 95% confidence interval for the mean difference ranged from -0.064 to 0.009, and the sample estimate for the mean difference was -0.028.

Sentiment Analysis

When divided into single sentences for sentiment analysis, feedback documents rendered 2,781 distinct sentences. Sentiment analysis on these produced valence scores for each sentence. The minimum (i.e., -0.750), median (i.e., 0.280), and maximum (i.e., 1.075) valence scores indicate that the range is positively skewed in this dataset. This suggests that most sentences were positively skewed. Figure 1 displays the distribution and interquartile range of averaged valence scores for the entire set of 1,385 documents; this suggests that the valence of feedback falls between 0.21 and 0.46, suggesting the valence scores across the documents are skewed in the neutral-to positive directions.

Valence score accuracy was evaluated by reading the sentences with the top 5 most positive and negative, and neutral scores. Table 5 displays these sentences, their valence scores, and a human assessment. It appears that the accuracy of the sentiment scores is limited and at times poor. For instance, the sentence "*solid teammate, no complains*" was given a score of -.38, where this comment imparted positive praise.

Figure 1

Distribution of Average Valence Score per Document Including Interquartile Range and Central Tendency

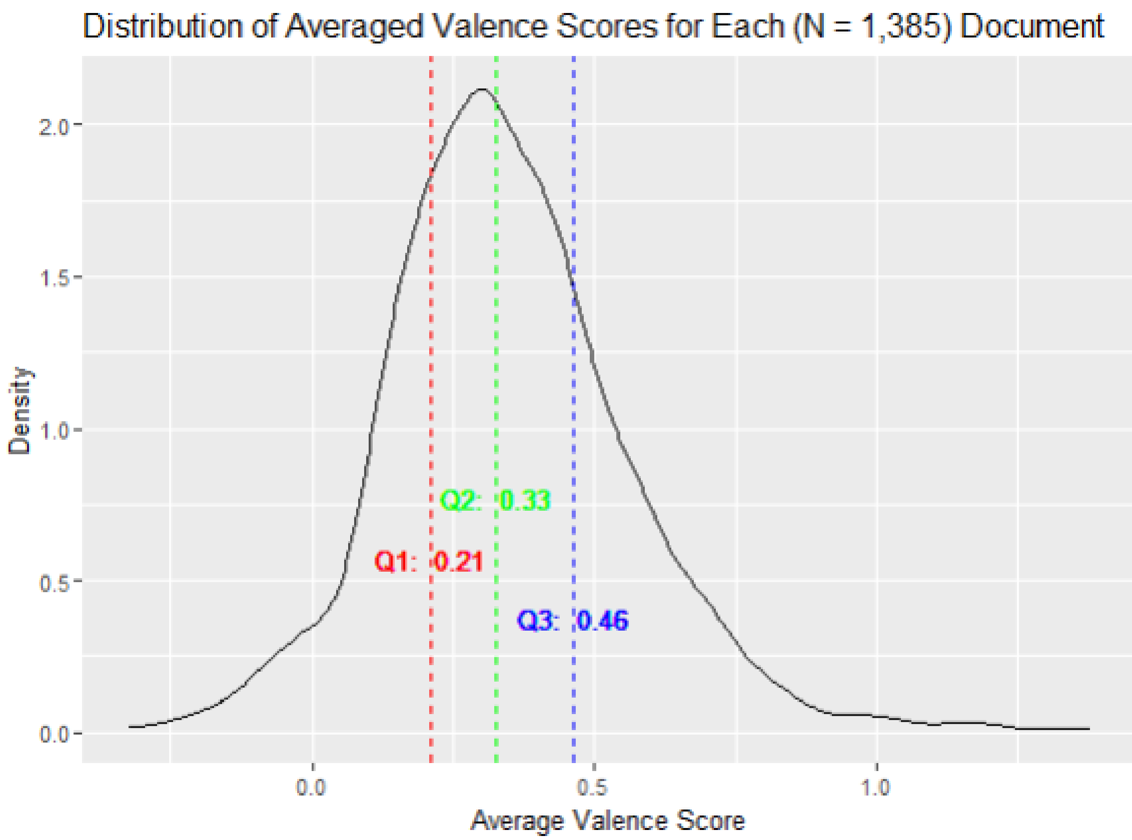


Table 5

Interpretation of Sentences with the Most Negative, Neutral, and Positive Valence Scores

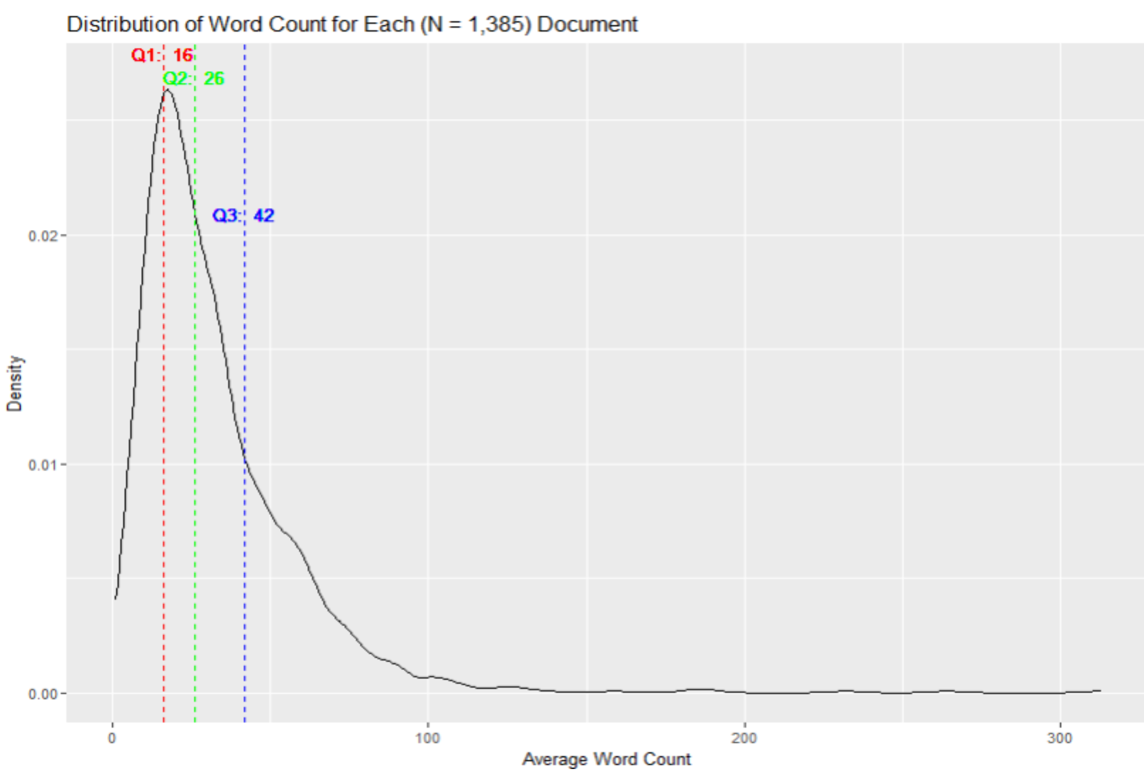
Sentence	Valence Score	Human Assessment of Score
"absentee."	-.75	Constructive criticism.
"My only complaint would be, I wish David would let us know how the rest of the group can help him complete his work more often."	-.70	Constructive criticism. Maybe be scored too negatively relative to other comments
"Bob's computer has been an issue, but even with the times we live in, computer issues are not an excuse."	-.62	Constructive criticism. Maybe be scored accurately relative to other comments
"I'd say that the rest of the group could definitely pick up some slack since you have really been grinding with CAD."	-.41	Positive praise. Criticism is not directed towards member.
"Solid teammate, no complaints"	-.38	Positive praise. Scored inaccurately as constructive criticism
"We are usually on the same page and can fix any issues relatively quickly."	.000	Nuetral praise.
"She still pulled her own weight to help the group out."	.000	Positive Praise.
"The entire team would benefit from putting more effort outside the lab."	.000	Neutral criticism directed towards team.
"Out of all the group members, she is most likely to know when things are due and what else needs to be done before then."	.000	Positive praise. Peprahs scored too negatively.
"If you believe something should be done a different way say something."	.000	Neutral criticism and suggestion for improvement. Score may be accurate.
You excel pages are well organized and easy to follow."	+.93	Positive praise. May be scored too positively
"You are really good at asking about others' progress and assuring everyone about your progress as well."	+.93	Positive praise. May be scored too positively
"Great team asset, super helpful with autocad."	+1.00	Positive praise. May be scored too positively
"You were great at ensuring your work was high quality and working with peers/consultant(s) to improve that quality."	+1.07	Positive praise. May be scored too positively
"I think you are pretty knowledgeable in the material and have helped guide us pretty well."	+1.08	Positive praise. Scored too positively relative to other sentences

Feedback Length

The range of feedback length, determined by number of words per document, ranged from 1 to 313 words, with a median of 26. The range and central distribution of feedback length per document is visualized in Figure 2 demonstrates that the inter-quartile range of word length is between 16-42 words; feedback larger than 100 words in length were infrequent in the dataset.

Figure 2

Distribution of Average Number of Words per Document Including Interquartile Range and Central Tendency



Hypothesis Testing

Hypothesis 1 stated that *feedback valence will demonstrate a curvilinear relationship with students' learning of teamwork skills such that learning increases as feedback approaches neutral valences.* Hypothesis 2 stated that *feedback with a greater word count will have significant, positive, linear relationship with increased learning on teamwork skills.* The mixed-effects polynomial regression showed a non-significant relationship between the quadratic valence scores and the dependent variable, leading to the rejection of hypothesis 1. Furthermore, there was a non-significant relationship between the word count and dependent variable, leading to the rejection of hypothesis 2 (see Appendix H). Scatterplots of the relationships of Peer Feedback Survey scores at time 2 with valence at time 1 (Figure 3) and feedback length at time 1 (Figure 4) indicate neither linear nor curvilinear relationships, rather they suggest feedback scores in the data are generally high, regardless of the feedback length or valence.

The interclass correlation value suggests that 31.5% of variance in the model is attributable to differences between teams, confirming the appropriateness of a mixed-effect model. Just the fixed effects, or interdependent and control variables account for 55.2% of variance (i.e., $Pseudo\text{-}R^2 = 0.552$) most of which is attributable to the control variable. Residuals were generally normally distributed (Appendix I), but heteroscedastic (Appendix J).

Figure 3

Scatterplot Showcasing the Relationship Between Average Feedback Valence at the Midterm and Peer Feedback Survey Scores at the Final

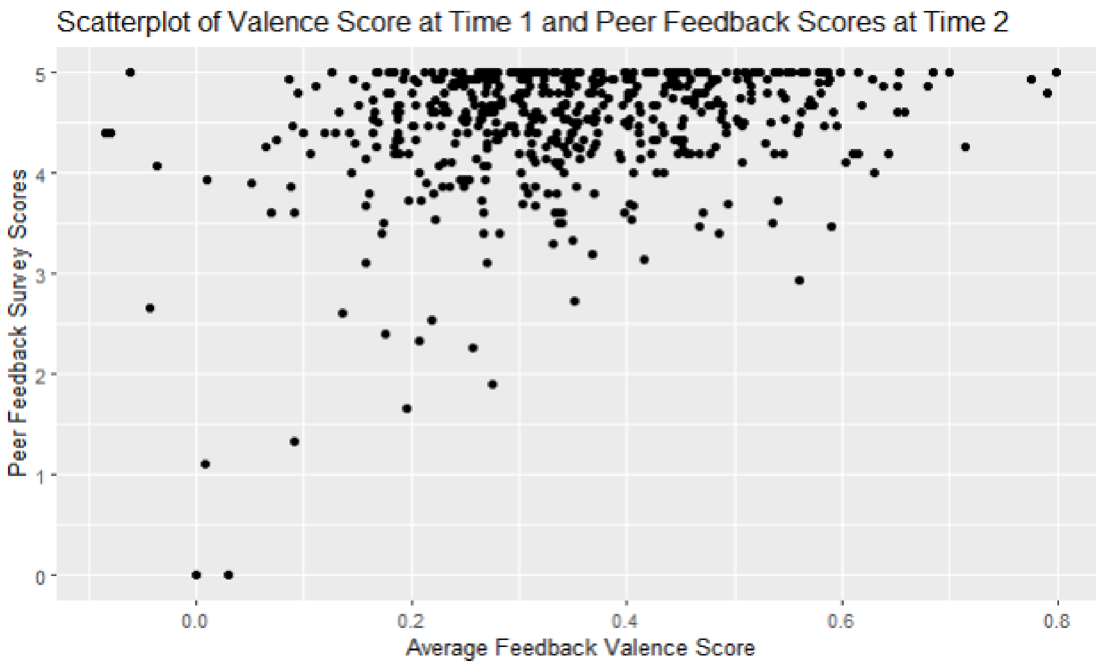
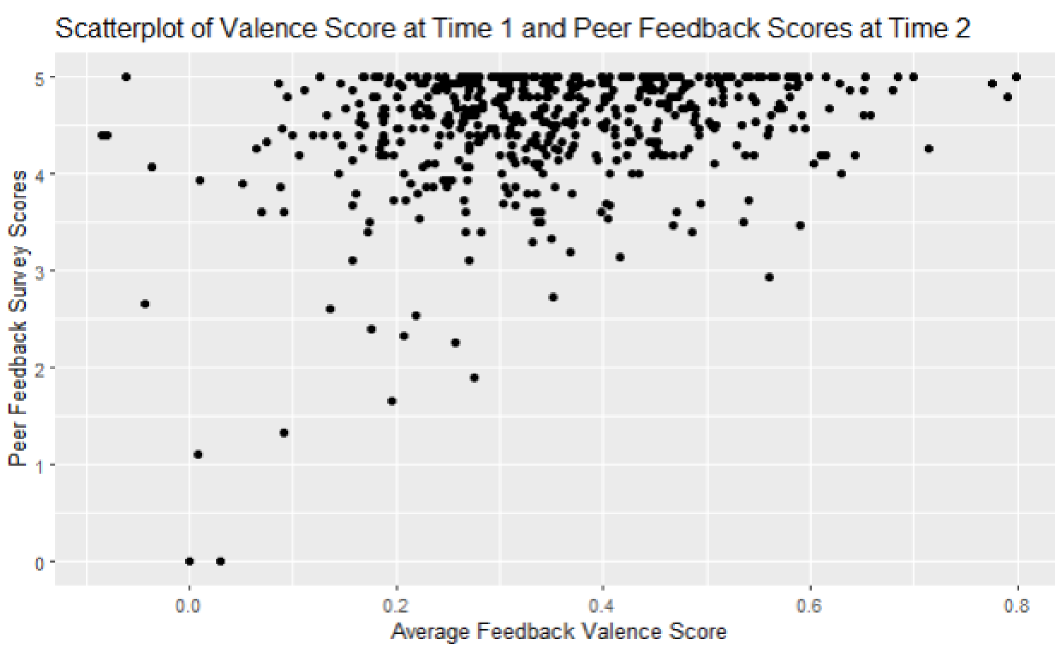


Figure 4

Scatterplot Showcasing the Relationship Between Average Feedback Length at the Midterm and Peer Feedback Survey Scores at the Final



CHAPTER SEVEN

DISCUSSION

This research sought to investigate the capabilities of NLP technology to draw insights about the characteristics of qualitative feedback exchanged using between student team members vis PESs. Furthermore, this research sought to explore the potential of NLP technology in combination with statistical analysis to identify relationships between the characteristics of qualitative peer feedback provided through these PES, and student learning and development outcomes when it comes to teamwork skills and abilities.

Topic Modeling

Research question one investigated the extent to which topic modeling, specifically biterm topic modeling designed for use on short texts, could be used to effectively ascertain topics communicated about through text feedback exchanged between members of student teams. Given that the validated topics in our final, 6-topic BTM demonstrated meaningful, interpretable, and unique labels, it was concluded that topic modeling can be used to extract meaningful topics from open-ended qualitative feedback provided between team members in student teams. This supports previous research that used TM to evaluate student text data using other TM methods such as LDA (Balahadia et al., 2016; Gottipati et al., 2017, 2018; Nanda et al., 2022; Sun & Yan, 2023).

The current findings contribute a new understanding that BTM is, seemingly to a large extent, an effective approach to extracting topics from student feedback data.

Specifically, Timely Communication, Idea Generation, Coordination & Adaptation, Work Quality, Team Support & Focusing, and Work Accountability are consistently, and relevant topics considered by student team members during peer evaluations submitted through PES. Given that our research included text data from across different classes spanning sophomore, junior, and senior years of a civil engineering course program, it may be concluded that these topics are relevant for engineering student teams at a broad level.

It should be noted that the topics identified through the BTM may not have comprehensively captured all the topics represented within the text feedback data. Indeed, the decision to consider models with containing only 5-11 topics, and to retain only the most reliably replicable topics, means we are not currently aware of what other topics might be found when BTMs are allowed to be larger, or permitted to have less restrictive topic reliabilities.

Topic Overlap with Teamwork & Learning Frameworks

Research question two sought comparison between BTM topics and constructs within existing frameworks of teamwork and learning and feedback. First, topics were compared with O'Neill et al.'s (2019) quantitative peer feedback ratings that would have been familiar to students through the PES. Here it was found that, perhaps unsurprisingly, students communicated about topics that they were also prompted to use in quantitative evaluations. This is similar to findings from other studies comparing qualitative topics and quantitative ratings from PES (e.g., Nanda et al., 2022). However, the overlap, there is not a one-to-one overlap between the BTM topics and ITP metrics ratings, with topics instead encompass several concepts and sub-concepts within the O'Neill Taxonomy.

Similar results are seen for the second framework under consideration, that is the Marks et al. (2001) taxonomy of team transition, action, and interpersonal processes, which showcase that topics often transcend behaviors germane to these phases. Some exceptions were found in the Timely Communication and Work Quality topics, and which only demonstrated strongly association with, respectively, *action – coordination* and *transition – goal specification*. Finally, topics were compared with Hattie and Timperley's (2007) framework which distinguishes the temporal orientation of feedback, and for levels at which feedback can be focused. Here, topic descriptions themselves did not clearly indicate whether feedback was focused on past or future actions. However, topic overlap was found with Hattie and Timperley focus of feedback, (e.g., task, process, self-regulation, and self level), but again overlapped with at least 2 of these levels.

Taking the results of this research question inquiry together (see Table 4), it appears that most of the BTM topics amalgamate concepts both within singular frameworks, and across frameworks. It appears that TM is not necessarily effective at generating topics that map seamlessly on to any one framework. This is perhaps unsurprising, given that qualitative data is inherently complex and nuanced**,** which has made qualitative feedback, at times, more impactful for learning than quantitative feedback alone (Koenka et al., 2021)**.** Thus, the amalgamation of teamwork and feedback constructs presented within these topics might reflect the thought processes that students experience when providing feedback – not necessarily focusing on behaviors that these models have neatly categorized, but instead commenting about multiple constructs which transcend team process phases, and levels of task, process, self-regulation and self.

An alternative perspective may be that the extent of overlap with varying concepts points to the need for smaller, more distinct topics in the BTM. Perhaps, a BTM with larger *K* could have more effectively dis-entangled these constructs. Furthermore, it is interesting to note that while many of the constructs across all three comparison frameworks were touched on, at least to some degree, by some of the topics in the model, certain constructs in the O’Neill et al.'s (2019) and Marks et al. (2001) frameworks were left out. Potentially, this means that peer feedback did not address these constructs to a large extent or in a way that was directly ascertainable from the topic description.

It may be considered a limitation of this qualitative portion of the research that a more rigorous approach to mapping TM topics to existing frameworks was not adopted. Although subject matters within the field of team science and teaching/learning discussed topic, this did not involve systematic qualitative coding, or inter-coder reliability checks. This could result in a less thorough or nuanced understanding of how the topics in the model overlapped with existing frameworks. Thus, future research may wish to exploit more rigorous qualitative methods.

Topic Relationship with Learning Outcomes

The third research question sought to understand whether receiving feedback on topics identified by the BTM resulted in significant changes to student teamwork skills at a later time-point in the semester, as measured quantitatively through peer evaluations. The findings suggested that there was not significant relationship between any of the 6 topics and student outcomes, discounting the possibility that receiving feedback on these topics' boosts student performance, at least in terms of peer-rated teamwork abilities. Thus, we may conclude that receiving feedback on one's Timely Communication, Idea Generation, Coordination & Adaptation, Work Quality, Team Support & Focusing, or Work Accountability, has a substantial impact on teamwork skills development. It is possible that, as opposed to the topics themselves, it is the manner in which these topics are communicated about that has a greater impact on learning outcomes. Previous research has demonstrated that peer feedback through PES can, indeed, have measurable positive impact on student learning (Donia et al., 2018; Petkova et al., 2021), yet research of feedback is also known to have negative, or non-significant relationships with performance (Kluger & DeNisi's, 1996; Krenn et al., 2013). This may relate to the quality and focus of feedback, since feedback that is more self-oriented rather than process or self-regulation oriented has been suggested to be of poorer quality (Hattie et al., 2021; Kluger & DeNisi, 1996; Mandouit & Hattie, 2023), and feedback that is overly positive with little future-oriented, constructive criticism may garner less helpful results (Fong et al., 2019).

As previously discussed, topics in the BTM could target several levels of Hattie 's feedback framework, suggesting that peers' feedback could focus different levels while communicating about the same overall topic, thus not guaranteeing feedback quality. Furthermore, the results of the sentiment analysis suggest that feedback was overwhelmingly within the neutral to positive range, suggesting a dearth of constructive criticism which might be needed to boost performance. Finally, our NLP analyses could not provide measures of the extent of future-oriented vs past-oriented commentary, a high-level analysis of example documents suggests that future-oriented feedback was likely scant in comparison to past-oriented feedback (see Table 3 and Table 4). These uncontrolled characteristics of feedback, namely the level at which feedback is focused and whether feedback is past or future oriented, might have impacted the outcome of our analysis.

Another plausible contribution to our null findings has to do with the overly positive ratings provided by peers at both time 1 midterms, and time 2 finals. A paired samples *t*-test found no significant differences between scores provided at these time-periods, and scores in both cases approached the very upper limits of the Peer Feedback Survey's 1-5 Likert scale. Since peers are providing team members with extremely high ratings at the mid-term, it leaves no room for growth and development to be appropriately showcased. This points to these measures being, perhaps, insufficient operationalization of students' learning of teamwork skills over a semester.

Impact of Feedback Length and Valence

This question posed two hypotheses about how the valence (e.g. positivity, neutrality, or negativity) of peer feedback, and the length of peer feedback would impact students learning outcomes. Ultimately, hypothesis testing in this research did not support a curvilinear relationship between valence of students' feedback and learning outcomes, nor a positive relationship between the amount of feedback and learning outcomes. The findings regarding feedback valence contribute to the body of literature that has found mixed effects of valence on student impact, (e.g., Anson & Anson, 2017; Wind & Jensen,

2017) which includes non-significance at times (Fong et al., 2019). Regarding feedback length, the current research generally stands in opposition to previous findings that have found a significant positive impact for length of feedback (e.g. Häkkinen & Ramadan, 2023; Wessa & De Rycker, 2010; Zong et al., 2021).

It is important to note several limitations to the current research when considering these findings. It is entirely possible that non-significant results were found due to a restricted range of both feedback valence and feedback length. The results demonstrate that the majority of the peer feedback fell within the range of neutral-positive valence (see Figure 1). Previous research notes that students often find it challenging to provide negative feedback to peers, due to anxieties about being unkind or lacking competencies to provide effective feedback (Cushing et al., 2011; Mulder et al., 2014). Research of valence of students' peer feedback often shows a skew towards the positive rather than neutral or negative (Burgess et al., 2021; Sridharan & Boud, 2019). Perhaps this is related to the finding that student feedback is often "mealy-mouthed", as in lacking directness or open criticism (Boud et al., 2018). This is evident even in the quantitative ratings provided by peers, who, already at the mid-term, appeared to be providing the maximum ratings using the Peer Feedback Survey.

Another possibility concern is with the quality of the valence analysis output. Validation of the valence scores against original sentences suggested that the *sentimentr* algorithm did not always accurately identify negative statements (see Table 5). In combination with a restricted range, a lack of confidence in the accuracy of the valance scores could be one reason that an alternative hypothesis was not supported in this case.

These are limitations that should be addressed with future research before re-examining potential relationship between feedback valence and learning outcomes.

Theoretical Implications

The current research suggests that, at a broad level, the topics described in our model (see Table 4) are most discussed and relevant to student engineers providing feedback to their teammates. Although the findings of our significance testing of the relationship between topics and student outcomes was negative, the emergence of these topics suggests that students perceive the associated actions and processes as critical to individual and/or team performance and outcomes. The fact that our BTM has conceptual overlap with several frameworks of teamwork and feedback highlights qualitative peer evaluations can provide nuanced insights into what aspects of team members' behaviors and characteristics are seen as important for team success. Perhaps, the constructs that appear more frequently across topics (such as "commitment to the team's work" and "communicating with team members" from the ITP framework) are particularly important for engineering student team outcomes. Also perhaps, the appearance of more numerous future-oriented feedback within some of the topics in the BTM than others (namely Timely Communication, Work Quality, and Work Accountability) might suggest that when an individual's performance in these areas is lacking, the impact to team outcomes is problematic enough to prompt constructive feedback from peers.

This research demonstrates that BTM is capable of capturing constructs of interest to the science of teamwork, learning and pedagogy, just as it has been relevant to capturing theories in other fields (e.g., Debortoli et al., 2016). The overlap that BTM topics demonstrate with existing theories and frameworks, provides evidence that BTM approaches do indeed produce topics and themes we might expect to see given the context in which the qualitative text was generated (a team, learning context). Beyond that, this makes TM a valuable tool in the arsenal of researchers learning to theories and frameworks through develop data-driven approaches harnessing written communications. Continuing to use TM might result in new and emergent theories or taxonomies of effective team member behavior that, perhaps, capture more complex constructs than quantitative rating scales have typically achieved alone. Continued use of TM in these contexts might aid in the development or refined frameworks that are more tailored to specific learning contexts (such as student engineer teams) than other frameworks (e.g. (Hattie & Timperley, 2007; Marks et al., 2001; O'Neill et al., 2019).

Practical Implications

The outcome of our research highlights BTM as a sensible option for educators withing to draw insights from student text feedback. Educators will likely lack the time to read and draw high-level conclusions from thousands of comments but can invest in a BTM approaches to automate this process. The practical take-aways of this research involve informing educators as to which topics, and related behaviors and actions appear to matter most to students working in teams. This may inform how instructors provide support, guidance, or advice to students who need additional support by elucidating areas that students could focus on in order to make improvements (e.g., timely communication, work accountability, etc.) Instructors of this program now have high-level topics that they can communicate about to students, along with descriptions of these topics and how they relate to teamwork competencies and behaviors as described by other frameworks. Students might even use these topics and descriptions as food for group discussions, debriefs, and reflection exercises.

Educators may also draw students' attention to the fact that feedback valence tends to be highly skewed in the positive direction, not only in qualitative feedback as indicated by the valence analyses, but also in quantitative feedback as indicated by high Peer Feedback Survey scores provided at time 1 and time 2. Furthermore, educators might highlight examples of future-oriented feedback, and feedback targeting the process, and self-regulation levels as a means to encourage more feedback with these characteristics. It might be necessary for instructors to provide guidance and/or regulations in order to limit the amount of overall-positive peer feedback and ratings. One such strategy might be to include BARS rating scales (i.e., "Behaviorally Anchored Rating Scales.") , as seen in other surveys used by Peer Evaluation Systems (see the CATME scale; Loughry et al., 2007), in order to provide students with examples and reference points that might guide them to provide more accurate ratings. Another option might be to use the outcomes of this study to showcase examples of student feedback that have qualities of being future-oriented and constructively critical feedback.

Future Research

An interesting use-case for LNPs like TM and valence analysis would be to evaluate the impact of interventions. If educators were to provide guidance and instruction for students on what kind of feedback to emphasize, (e.g., future-oriented work quality), or where to focus the topic (process as opposed to self-level), topic modeling could potentially pick up on whether such interventions produced discernible shifts in text topic communication, or the range of text valence. Hence, BTM could serve as a measure to assess the effectiveness of interventions designed to increase high-quality feedback and might result in different topics and valence analysis outcomes.

Future research might compare BTM with other TM approaches such LDA or STM, on the same short-text peer evaluation data, to compare advantages and disadvantages or similarities and differences in topic identification and topic model quality in these contexts. Future research may also experiment in generating BTM with larger *Ks* in combination with different model parameters such as word co-occurrence windows. By increasing or downsizing the word cooccurrence window, BTMs would capture larger or smaller contexts around each focal word, focusing on closer or more distant semantic relationships. Reducing the size may lead to more focused and specific topics, a broader context would potentially incorporating more diverse topics. Also, searching a wider range of *Ks* to include in the BTM might reveal more than only 6 reliably reproducible topics.

It would be interesting to study whether larger models with more topics, or models with different word-cooccurrence windows demonstrate cleaner overlap with frameworks of teamwork and learning. Perhaps, the extent of overlap seen in the current research is indicative of room to semantically consolidate topics or introduce more topics into a model, thus reducing the amount of overlap of topics in our model with existing frameworks.

For future research interested in continuing to investigate whether topics from BTM have predictive relationships to student or team outcomes, consider using alternative performance metrics, such as final grades or scores on projects, or comparing outcomes to matched controls that did not receive qualitative peer feedback. Perhaps, such research designs could offer alternative angles through which to investigate potential relationships between feedback topics and students' teamwork skills learning and development. Another interesting value would be to test models with different parameters, such as smaller or larger co-occurrence windows. Models with more topics, or more or less restrictive cooccurrence windows might find topics that, in fact, do have significant relationships to student learning outcomes. This might produce different outcomes in significance testing analyses.

Given that our research included text data from across different seniorities and types of classes, we may conclude that these topics are relevant for engineering student teams at a broad level. However, past research has demonstrated that shifts in setting and time period can yield different topics in TMs (Nanda et al., 2022). Thus, more specific, tailored insights could be drawn by introducing covariates related to the different classes students and teams belong to, student seniority, whether students or teams ultimately experience good or poor outcomes.

The current research found numerous flaws with the valence analysis, specifically that algorithms, at times, miss-identified positive peer feedback as having a negative valence (see Table 5) Thus, it appears additional modifications to the sentiment analysis approach and algorithms may be needed. Future research might investigate why such error occurred and incorporate additional measures (e.g., introducing custom lexicons) to improve outcomes. This would make future significance testing analyses more accurate, particularly if research also insured that textual data contained a broad range of positive and negative commentary.

APPENDICES

Appendix A

Sample of ITP Metric's Open-Ended Qualitative Feedback Platform

Peer Feedback

Please provide feedback to each of your team members. Remember that your feedback to your teammates will be anonymous, i.e., they will not know who provided it. Only the assessment administrator can determine who provided each comment.

Feedback may be on positive aspects of working with a team member or areas for improvement or both. Your comments should be:

- Specific - rather than "you did an excellent job", explain "you did an excellent job in responding to emails in a timely manner".
- Constructive (even when providing negative feedback) - rather than "you were lazy and didn't pull your weight in the group", explain "at times it felt like you didn't contribute as much to the team as the rest of us, the entire team would benefit from your participation".
- Respectful - never use abusive language and always address your team members as you would in a professional work setting.

Peer B

☐ Not familiar with team member's behavior

Peer C

☐ Not familiar with team member's behavior

Appendix B

Constructs and Definitions Measured by the ITP Metrics Peer Feedback Survey

ITP Metrics

Peer Feedback

The following questions will ask for your personal opinions regarding each team member. This section is expected to take less than 5 minutes.

Communicating with team members: Please rate the extent to which each team member engages in the following set of behaviors.

Effectively communicates and openly shares information.
Exchanges information with teammates in a timely manner.
Requests feedback regularly and incorporates feedback from team members
Seeks appropriate team input before taking action.

Capabilities: Please rate the extent to which each team member engages in the following set of behaviors.

Seeks to gain the knowledge, skills, and abilities needed by the team.
Learns about other teammates' tasks and roles.
Demonstrates the capabilities needed for the team to perform.

Commitment to the team's work: Please rate the extent to which each team member engages in the following set of behaviors.

- Contributes appropriately to the team's work.
- Demonstrates commitment to the team's work.
- Prepared for team meetings.
- Keeps deadlines and delivering complete, accurate work.

Emphasizing high standards: Please rate the extent to which each team member engages in the following set of behaviors.

- Encourages and motivates the team.
- Shows confidence in the team's ability to perform.
- Believes that the team will achieve high standards.
- Cares about the quality of team's work.

Focus: Please rate the extent to which each team member engages in the following set of behaviors.

Monitors issues that may affect the team and notices problems.
Provides meaningful, growth-oriented, and regular feedback to members.
Helps the team plan and organize work and anticipates issues.

Appendix C

Marks et al.'s (2001) Taxonomy of Team Processes

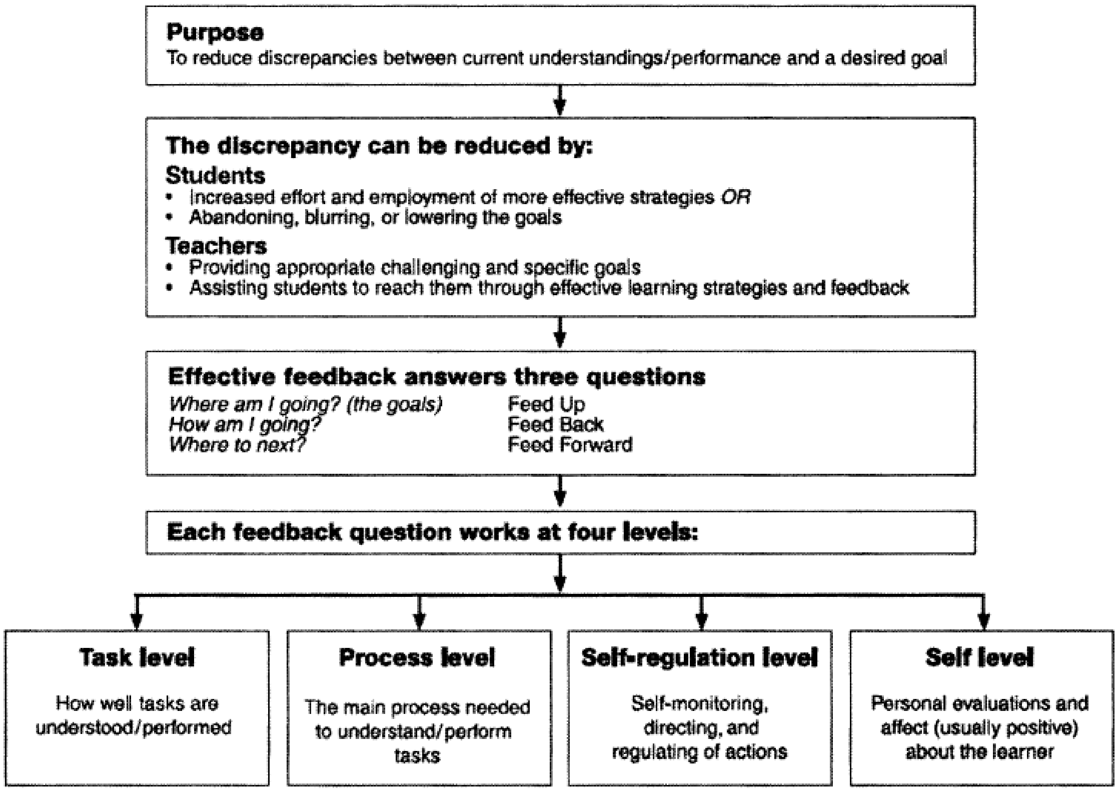
Process Dimensions	Definition
Transition processes	
Mission analysis formulation and planning	Interpretation and evaluation of the team's mission, including identification of its main tasks as well as the operative environmental conditions and team resources available for mission execution
Goal specification	Identification and prioritization of goals and subgoals for mission accomplishment
Strategy formulation	Development of alternative courses of action for mission accomplishment
Action processes	
Monitoring progress toward goals	Tracking task and progress toward mission accomplishment, interpreting system information in terms of what needs to be accomplished for goal attainment, and transmitting progress to team members
Systems monitoring	Tracking team resources and environmental conditions as they relate to mission accomplishment, which involves (1) internal systems monitoring (tracking team resources such as personnel, equipment, and other information that is generated or contained within the team), and (2) environmental monitoring (tracking the environmental conditions relevant to the team)
Team monitoring and backup behavior	Assisting team members to perform their tasks. Assistance may occur by (1) providing a teammate verbal feedback or coaching, (2) helping a teammate behaviorally in carrying out actions, or (3) assuming and completing a task for a teammate
Coordination	Orchestrating the sequence and timing of interdependent actions

Appendix C (continued)

Interpersonal processes	
Conflict management	Preemptive conflict management involves establishing conditions to prevent, control, or guide team conflict before it occurs. Reactive conflict management involves working through task and interpersonal disagreements among team members
Motivation and confidence building	Generating and preserving a sense of collective confidence, motivation, and task-based cohesion with regard to mission accomplishment
Affect management	Regulating member emotions during mission accomplishment, including (but not limited to) social cohesion, frustration, and excitement

Appendix D

Hattie and Timperley's (2007) Model of Effective Feedback



Appendix E

Mixed-Effects Multiple Regression Model Output: Summary of Topic and Teamwork Skills Learning Significance Testing Output

MODEL INFO:
Observations: 515
Dependent Variable: Time 1 Average Peer Feedback Survey (PFS) Score
Type: Mixed effects linear regression

MODEL FIT:
$AIC = 1045.561$, $BIC = 1088.003$
Pseudo-R^2 (fixed effects) = 0.553
Pseudo-R^2 (total) = 0.698

FIXED EFFECTS:

	Est.	S.E.	t val.	d.f.	p	F value
(Intercept)	-0.003	0.041	-0.068	131.138	0.946	
Timely Communication	0.029	0.031	0.936	470.904	0.350	6.613
Idea Generation	-0.016	0.038	-0.425	505.855	0.671	3.225
Coordination & Adaptation	-0.023	0.034	-0.663	480.109	0.508	0.308
Work Quality	-0.040	0.033	-1.198	498.545	0.231	29.267
Team Support & Focusing	-0.021	0.035	-0.593	490.851	0.554	0.293
Work Accountability	-0.018	0.032	-0.569	484.068	0.570	1.226
T2 Averaged PFS Score	0.741	0.030	24.456	506.462	0.000	602.476

p values calculated using Kenward-Roger standard errors and d.f.

RANDOM EFFECTS:

Group	Parameter	Std. Dev.
Group ID	(Intercept)	0.383
Residual		0.554

Grouping variables:

Group	# groups	ICC
Group ID	135	0.324

Appendix F

Testing Assumptions of Residuals Normal Distribution for Mixed-Effects Multiple Linear Regression Model

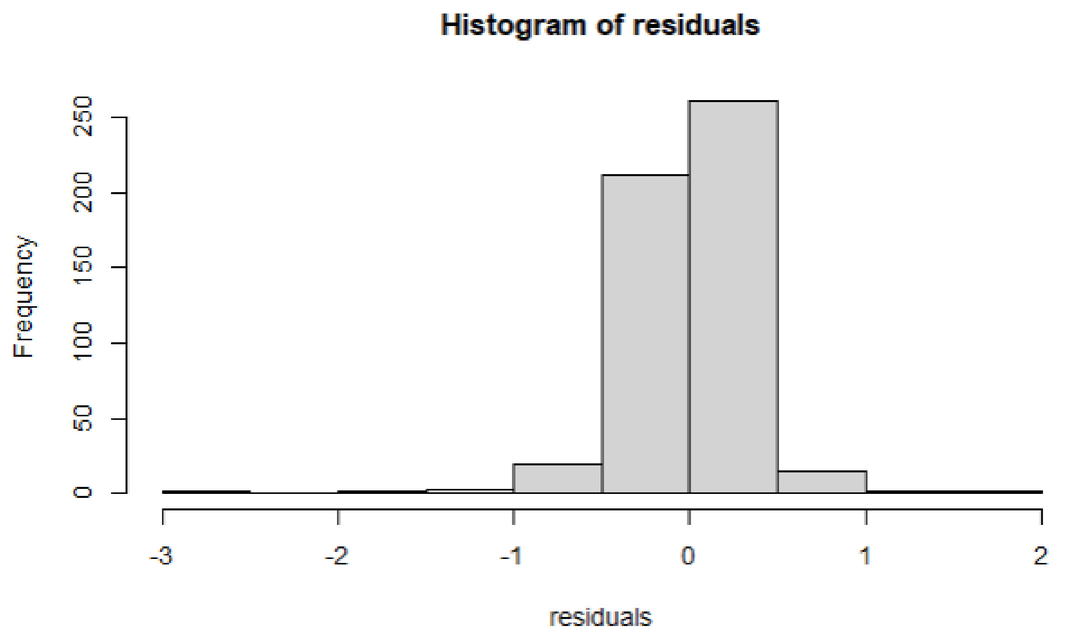


Figure A: Histogram of Residual Distribution

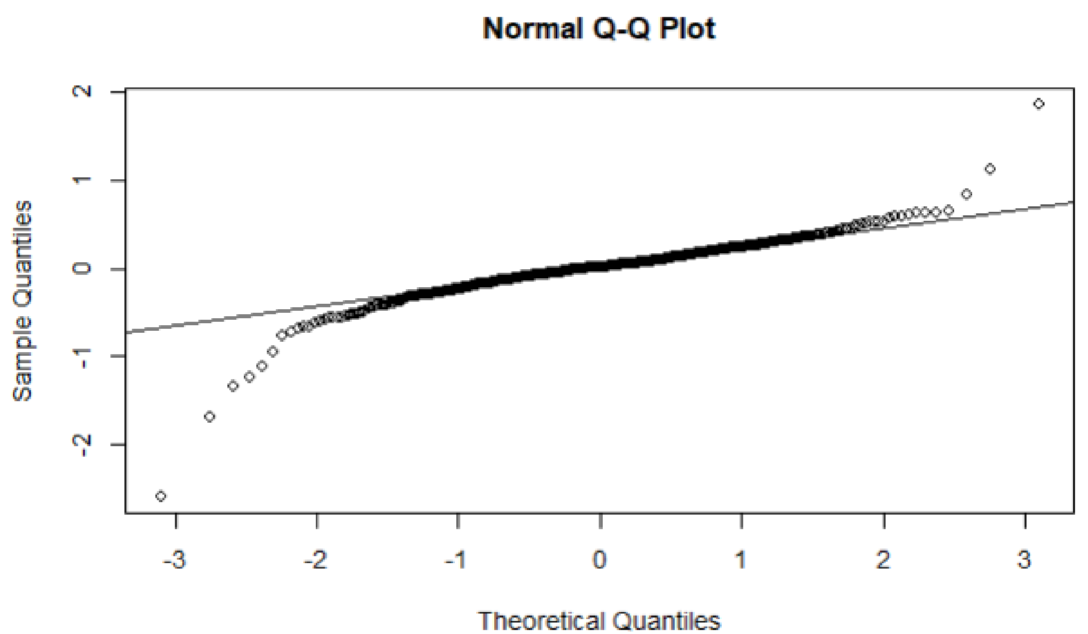
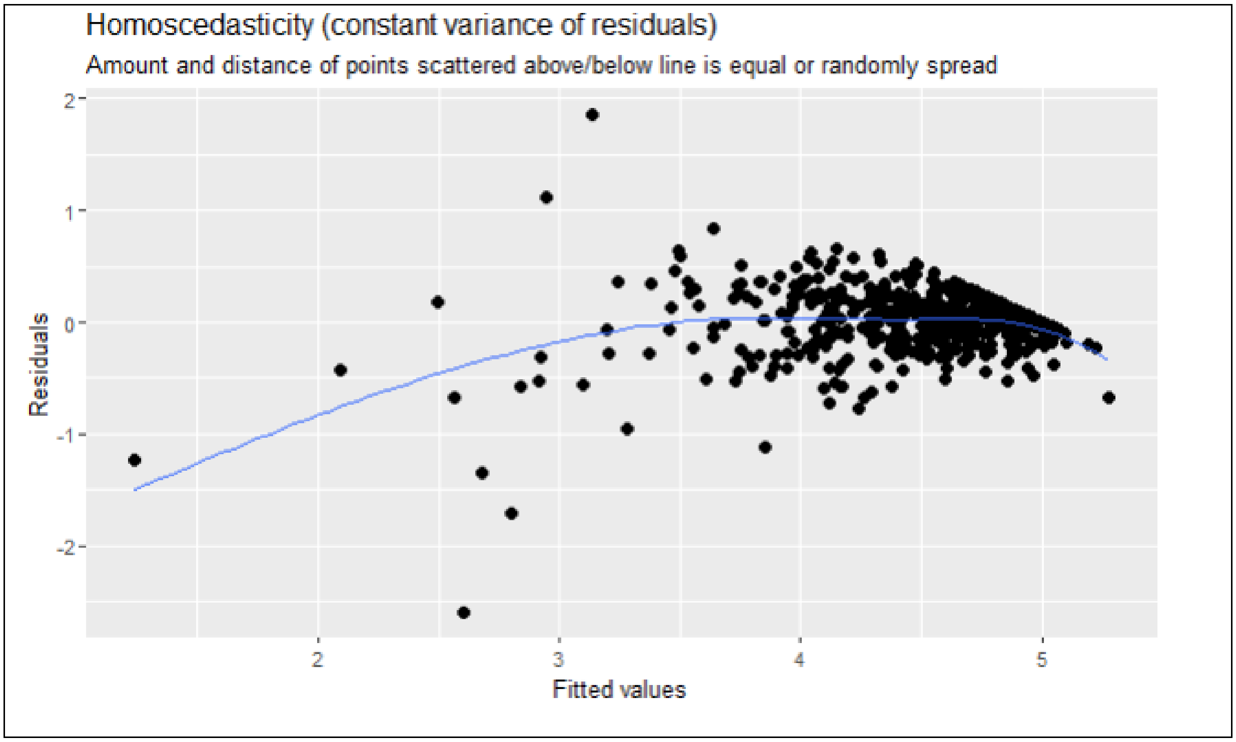


Figure B: QQ Plots of Residuals

Appendix G

Assumption Testing of Homoscedasticity of Residuals of Mixed-Effects Multiple Linear Regression Model: Plot of Model Residuals Against Fitted Values



Appendix H

Mixed-Effects Polynomial Regression Model Output: Summary of Feedback Valence and Feedback Length Significance Testing Output

MODEL INFO:
Observations: 515
Dependent Variable: Time 1 Average Peer Feedback Survey (PFS) Score
Type: Mixed effects linear regression

MODEL FIT:
$AIC = 1026.199$, $BIC = 1055.908$

Pseudo-R^2 (fixed effects) = 0.552
Pseudo-R^2 (total) = 0.693

FIXED EFFECTS:

	Est.	S.E.	t val.	d.f.	p	F value
(Intercept)	-0.002	0.041	-0.052	131.849	0.959	
Quadratic term for Average Valence Score	0.031	0.030	1.027	499.403	0.305	
Average Valence Score	0.026	0.028	-0.916	483.857	0.360	51.480
Feedback Length	-0.015	0.035	-0.428	364.367	0.669	10.979
T2 Averaged PFS Score	0.729	0.032	22.895	509.619	0.000	528.328

p values calculated using Kenward-Roger standard errors and d.f.

RANDOM EFFECTS:

Group	Parameter	Std. Dev.
GroupID	(Intercept)	0.377
Residual		0.555

Grouping variables:

Group	# groups	ICC
GroupID	135	0.315

Appendix I

Testing Assumptions of Residuals Normal Distribution for Mixed-Effects Polynomial Regression

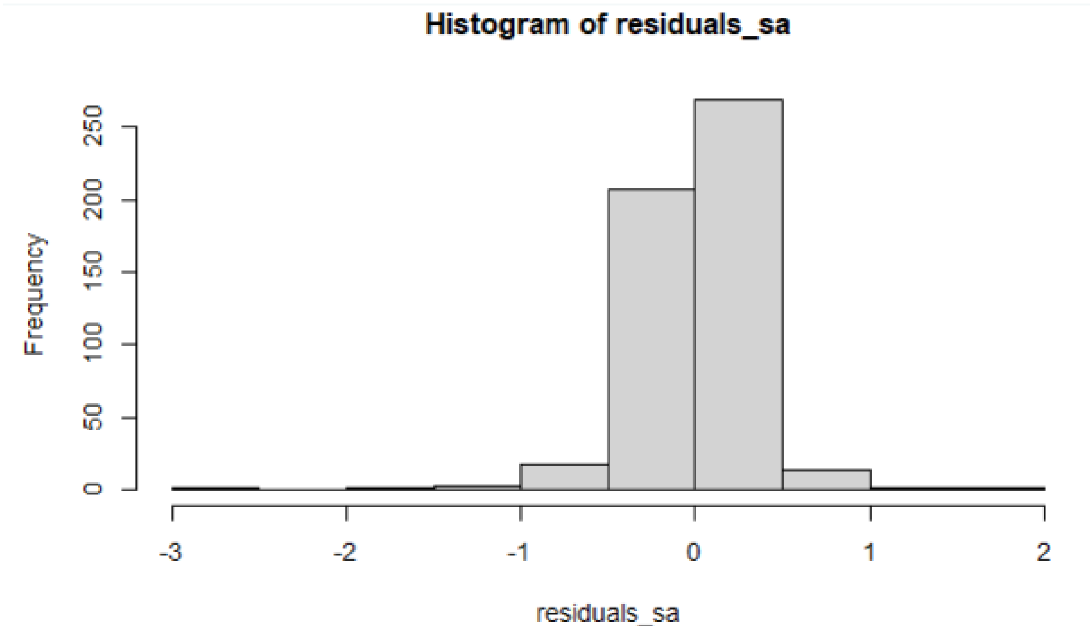


Figure C: Histogram of Residual Distribution

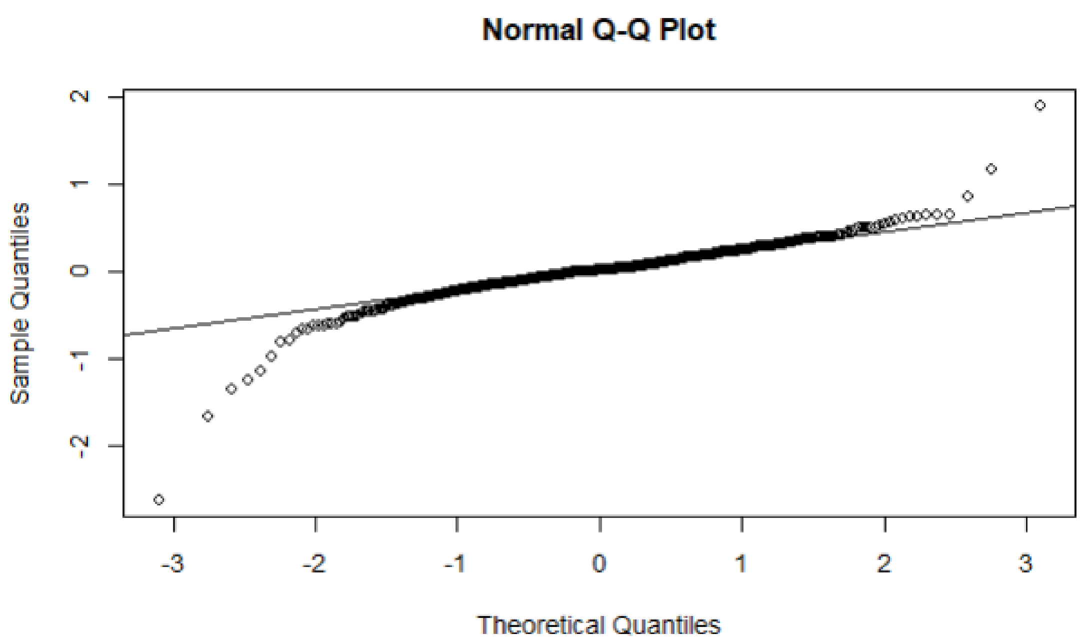
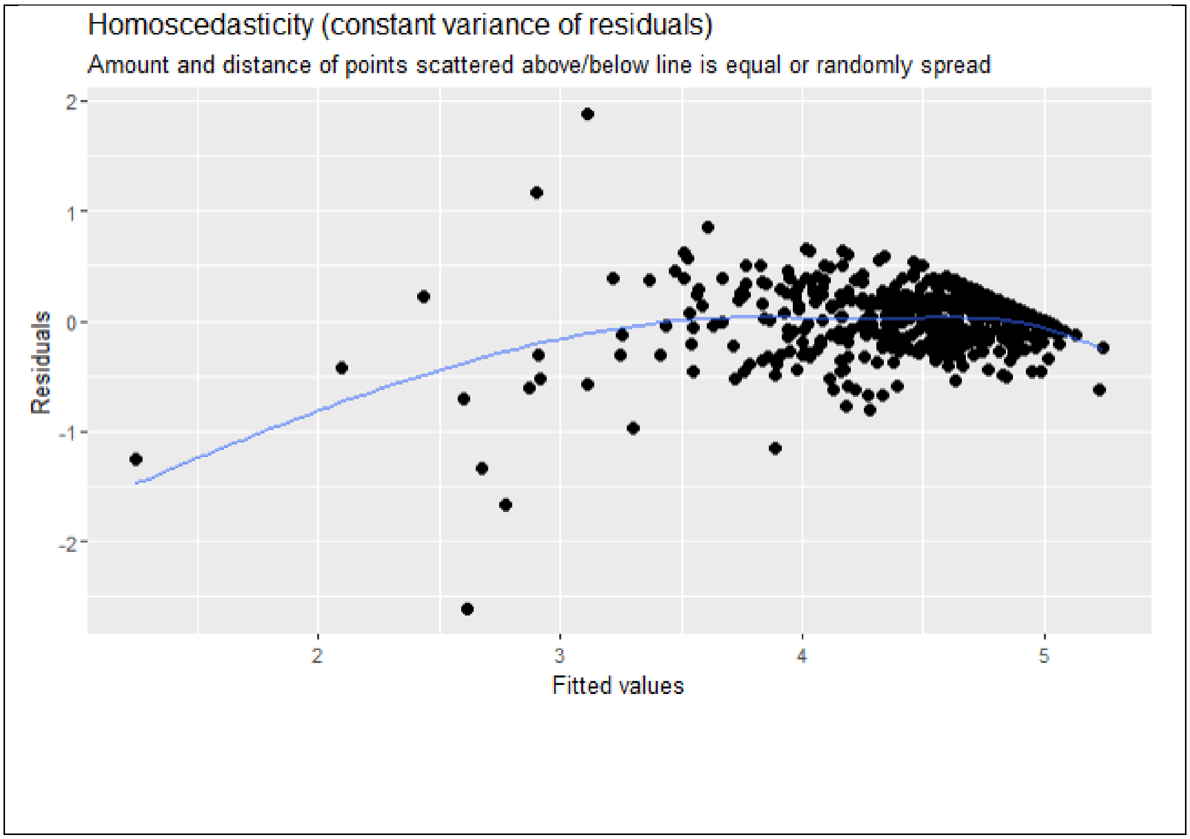


Figure D: QQ Plots of Residuals

Appendix J

Assumption Testing of Homoscedasticity of Residuals of Mixed-Effects Polynomial Regression Model: Plot of Model Residuals Against Fitted Values



Appendix K

R Code for Text Pre-Processing, Biterm Topic Modeling, Valence Analysis, and Regression Analysis

```
---
title: "BTM Model final"
author: "Anna Wolf"
date: "`r Sys.Date()`"
output:
  pdf_document: default
  word_document: default
toc_float: yes
---

## TOPIC MODELING

```{r}
library(readxl)
library(tm)
library(textstem)
library(udpipe)
library(dplyr)
library(BTM)
library(data.table)
library(ggraph)
library(concaveman)
library(ggforce)
library(textplot)
library(lda)
library(stm)
library(quanteda)
library(ggplot2)
library(LDAvis)
library(topicmodels)
library(ldatuning)
library(text2vec)
library(Matrix)
library(shiny)
library(RColorBrewer)
library(wordcloud)
library(stringr)

```

Importing Data

Set the proper working directory. Import the data, which contains a Group ID, Person ID, and the qualitative feedback given to individual team members at the midterm (Time 1/T1) from each of their peers, which have been merged into one column.

```{r}
```
```

Appendix K (continued)

````
#import data
feedback_full <- read.csv("feedback_full")

View(feedback_full)

#add document ids to the data
feedback_full$doc_id <- seq_len(nrow(feedback_full))

dim(feedback_full)
head(feedback_full)


```

## Text Preprocessing

Preprocess the text data, create corpus, create document frequency matrix and document term matrix

```{r}

cleaned_text <- feedback_full$Feedback %>%
  str_replace_all("[[:punct:]]", "") %>%              # Remove punctuation
  str_replace_all("\\s+", " ") %>%                   # Replace multiple whitespaces with a single space
  str_trim() %>%                                     # Trim leading and trailing spaces
  tolower()

# Create a corpus from the cleaned text
corpus <- Corpus(VectorSource(cleaned_text))


# Remove custom stop words
custom_stop <- c("caleb", "henry", "annamarie", "aidan", "brian", "luke", "elizabeth", "isaiah", "james",
"levi", "marie", "britney", "carrie", "connor", "adam", "alexandria", "allie", "josh", "bob", "chris", "jordan",
"courtney", "alice", "mary", "saxton",  "zach", "jackie", "diana", "addison", "jerrod", "kyle", "marielle",
"michelle", "samantha", "megan", "sam", "emily", "joshua", "camilo", "tyler", "brendan", "caroline",
"david", "lexi", "jimmy", "abby", "matt", "jonah",  "addie", "ali", "andrew", "ben", "alyssa", "na","aaron",
"jamal", "matthew", "rob", "supreet", "shreya", "sabbir", "john",  "truett", "coy", "bayden", "alec",
"jeremy", "taylor", "todd", "alex", "kylee", "blake", "christian", "britney", "sam", "emily", "joshua",
"camilo", "tyler", "brendan", "caroline", "david", "lexi", "jimmy", "abby", "matt", "jackie", "jonah",
"addie", "ali", "andrew", "ben", "alyssa", "darryl",  "blake", "brantley", "charlie", "cameron", "kylee",
"rob", "arron", "trent", "brett", "gabel", "olivia", "na", "joey","julian", "deirdre", "michael", "moe",
"brandon", "vincent", "garret", "dylan", "paul", "kevin", "gage", "stephanie", "mathew", "josue", "daze",
"erik", "nathan", "trey", "zary","paul", "grant" ,"young","victor","corey", "maggie", "tommy","van","elina,"
,"zary", "cooper", "farhad", "hanzhi", "ross", "mikayla", "grant", "karl", "josh", "tim", "victor", "maggie",
"josh", "NA", "na","work", "job", "great", "team", "group", "good")

corpus <- tm_map(corpus, removeWords, custom_stop)
corpus <- tm_filter(corpus, FUN = function(x) nchar(as.character(x)) > 0)

# Get the cleaned text back from the corpus
````

Appendix K (continued)

```
cleaned_text <- sapply(corpus, as.character)


#convert cleaned text to data frame
cleaned_df <- as.data.frame(cleaned_text, stringsAsFactors = FALSE)

# Remove empty rows from the data frame
cleaned_df <- cleaned_df[apply(cleaned_df, 1, function(x) any(nchar(trimws(x)) > 0)), , drop = FALSE]
dim(cleaned_df)


# Create Quanteda Corpus

quanteda_options("threads" = 7)
corpus_quanteda <- corpus(cleaned_text)

# Forgood measure perform DFM Preprocessing

DFM <- corpus_quanteda %>%
  quanteda::tokens(remove_numbers = TRUE,
                   remove_punct = TRUE,
                   remove_symbols = TRUE,
                   remove_url = TRUE) %>%
  tokens_tolower() %>%
  tokens_remove(pattern = stopwords("en")) %>%  # Remove common English stopwords
  tokens_remove(pattern = "\\s+") %>%  # Remove white spaces
  tokens_remove(pattern = "[^a-zA-Z0-9\\s]") %>%  # Remove symbols except spaces
  tokens_remove(custom_stop) %>%  # Remove custom stopwords
  dfm()

# View the top words in the DFM
topfeatures(DFM, n = 10)
dim(DFM)


# Create a  DTM
DTM <- convert(DFM, to = "topicmodels")
dim(DTM)


```
```
```

Appendix K (continued)

Lemmitization of the DTM using UDPipe

Lemmitization transforms words into "lemmas", which as their base or dictionary forms. For example, the word "running" is lemmatized to "run." Unlike stemming, it considers word meanings and context to ensure the lemmas are valid words, and is therefore considered more accurate than stemming which tends to chop off the ends of words in a crude manner. This can result in nonsense words when, for instance the word 'communicate' is stemmed to 'communic'. Lemmatization may enhance text analysis by providing precise and interpretable word forms, benefiting tasks like information retrieval and sentiment analysis. Therefore, the tokens in the DFM were replaced with their lemmatized forms using additional preprocessing

```{r}

# Load the pre-trained UDPipe model from the GitHub URL
eng <- udpipe_download_model(language = "english-ewt")

eng$file_model

m_eng_ewt_path <- eng$file_model

m_eng_ewt_loaded <- udpipe_load_model(file = m_eng_ewt_path)

# create tokens of lemmas
udpipe_annotate <-  udpipe_annotate(m_eng_ewt_loaded, x = cleaned_text) %>%
      as.data.frame()

View(udpipe_annotate)

# Create a mapping between tokens and lemmas
lemma_map <- setNames(tolower(udpipe_annotate$lemma), udpipe_annotate$token)
patterns <- names(lemma_map)
replacements <- unname(lemma_map)

# Replace tokens in the DFM with lemmas

DFM_lemma <- dfm_replace(DFM, pattern = patterns, replacement = replacements, case_insensitive =
TRUE)

# See top lemmas in our data
top_lemmas <- topfeatures(DFM_lemma, n = 10)  # Get top 10 lemmas
print(top_lemmas)


# Create lemmatized DTM
DTM_lemma <- convert(DFM_lemma, to = "topicmodels")
dim(DTM_lemma)

```

Appendix K (continued)

Biterm Specification

Tagging Parts of Speach
Biterm topic modeling allows us to specify which biterms we are interested in paying attention to in our data. This has be benefit of instructing the algorithm to ignore parts of speech that are unlikely to add much meaning to our topics (e.g. conjucation words like "and", adposition words like "on") irreverent (i.e., stop-words like). We can specify which words are believed to create relevant biterms, in this case specifying "Nouns", "adjectives" and "verbs".

Specifying Word Cooccurence Window

BTM also allows us to specify how far apart words in biterm pairs are allowed to be by adjusting the skipgram parameter in the cooccurrence function. A smaller value like 3 captures more immediate word relationships, whereas a larger value like 10 captures broader context but might include less directly related word pairs.

The average length of a student's feedback was 26 words, setting skipgram to 26 would theoretically capture cooccuring biterms across that might occur within the average feedback provided. Previous research suggests that this is around the number of words that it takes to express singular thoughts in oral or written form (i.e. between 20 - 29 words) (Einhorn, 1978).

```{r}

anno  <- udpipe(cleaned_text, "english", trace = F)

# create biterms
biterms <- as.data.table(anno)
biterms <- biterms[, cooccurrence(x = lemma,
                          relevant = upos %in% c("NOUN", "ADJ", "VERB") &
                                 nchar(lemma) > 2 & !lemma %in% stopwords("en"),
                          skipgram = 26), # adjust skipgram = to choose desired word cooccurence
             by = list(doc_id)]


```

Model Selection & Evaluation

Create x_token object, which actually contains a list of lemmas

```{r}
# BTM preprocessing and tokenization
x_token <- as.data.frame(anno$lemma)
x_token <- as.data.frame(x_token[x_token$`anno$lemma` != "",]) # drop empty docs
x_token$doc <- rownames(x_token)
x_token <- x_token[,c(2,1)]
names(x_token) <- c("doc_id", "text") # names as expected from udpipe()
x_token$doc_id <- anno$doc_id # add a 'doc_id' to the list


```

Semantic coherence and exclusivity search

```{r}
# import scripts developed by from Bitterman et al. (2021)
source("BTM_grid_new.R")
source("exclusivityBTM_new.R")
source("coherenceBTM_new.R")

# Specify the grid search aka K range
klist <-5:11 # range of k based on preliminary examinations


# For each k, BTM_grid returns best model (coherence * exclusivity) w.r.t. alpha and seed
## this could take a while ##

btm_bestk <- BTM_grid(x_token, biterms, DTM_lemma, klist)

save(btm_bestk, file = "btm_bestk_s26.RData")

# Evaluation metrics

EX_best <- list()
CO_best <- list()

for (i in 1:length(klist)){
  EX_best[[i]] <- exclusivityBTM(btm_bestk[[i]])
  CO_best[[i]] <- mean(coherenceBTM(btm_bestk[[i]], DTM_lemma, 5))
}

plot(scale(unlist(CO_best)), type = "l", col = "black", xaxt = "n", ylim = c(-2, 2), lwd = 2,
     main = "BTM Evaluation Metrics", ylab = paste("Scaled Score"),
     xlab = "Number of Topics")
lines(scale(unlist(EX_best)), type = "l", col = "blue", lwd = 2)
lines(rowMeans(cbind(scale(unlist(EX_best)), scale(unlist(CO_best)))), type = "l", col = "red", lwd = 2)
axis(1, at = 1:length(klist), labels = klist)
legend("bottomright", c("Semantic Coherence", "Exclusivity", "Mean Coherence & Exclusivity"),
       col = c("black", "blue", "red"), lty = "solid", lwd = 2)


# table for mean Coherence & Exclusivity
coex <- rowMeans(cbind(scale(unlist(EX_best)), scale(unlist(CO_best))))
names(coex) <- klist
print(coex)

# k values of top 5 coex
coex_top <- as.numeric(names(sort(coex, decreasing = TRUE)[1:3]))
print(coex_top)

## inspect candidate models
```

Appendix K (continued)

````
library(textplot)
library(ggraph)

```

### Final Model
Choosing the model with the best semantic coherence and exclusivity trade-off, which is model 9
```{r}
k <- 9
model <- btm_bestk[[which(klist == k)]]
alpha <- model$alpha
terms <- terms(model, top_n = 15)

# save model and model plot
save(model, file = "model_s26.RData")

model_plot_full <- plot(model, top_n = 15,
     title = "BTM Model with Skipgram 26")

ggsave("BTM Model with Skipgram 26.png", plot = model_plot_full)


# create data frame with topics, theta, and labels
topic_all <- data.frame(
  TopicNumber = 1:k,
  Prevalence = model$theta)
for (i in 1:k){
  topic_all$Terms[i] <- paste(terms[[i]][,1], collapse = " ")
}
```



## Topic Reliability
This code is part of a process used for evaluating the stability or reliability of topics generated by topic
modeling algorithms. Specifically, it uses cosine similarity to measure the similarity between topics
generated by different runs of the same topic modeling algorithm.
```{r}
# for same k and alpha, 10 other models with different seeds are calculated.
# check, which topics are stable across different inference runs (different seeds)

# different seeds
# do not use same seeds as in BTM_grid() (12, 4321, 28, 3323)
seedlist_rel <- c(5412, 441, 1128, 30323, 3735, 2721, 127722, 97039, 173, 317231)


# btm_temp: same model for 10 different seeds
cores <- 5
cluster <- parallel::makeCluster(cores)
parallel::setDefaultCluster(cluster)
````

Appendix K (continued)

```
parallel::clusterExport(varlist = c("x_token", "biterms", "k", "alpha"), envir = environment())
btm_temp <- parallel::parLapply(X = seedlist_rel, fun = function(x) {
  set.seed(x)
  BTM::BTM(x_token, biterms = biterms, k = k, alpha = alpha, beta = 1/k, iter = 2000, trace = 100, detailed
= TRUE)
})
parallel::stopCluster(cluster)

save(btm_temp, file = "btm_temp_s10_k15_clean.RData")


source("reliabilityBTM_new.R") # modified from LDA version by A. Niekler & G. Wiedemann
# Niekler, A.: Automatisierte Verfahren für die Themenanalyse nachrichtenorientierter Textquellen:
# Dissertation zur Erlangung des akademischen Grades Doktor-Ingenieur (Dr.-Ing.) im Fachgebiet
Informatik,
# http://asv.informatik.uni-leipzig.de/publication/file/350/Niekler_Diss.pdf, (2016).

topic_rel <- topic_all # df with topic numbers and terms

for (i in 1:length(btm_temp)){
  tm_fit1 <- from_BTM_package(model)
  tm_fit2 <- from_BTM_package(btm_temp[[1]])

  relCosine <- reliability_cosine(tm_fit1, tm_fit2, topWordsToMatch = 10, threshold = 0.8)
  #print(relCosine$reliability)
  topicmatches <- relCosine$ids

  rel_df <- data.frame(topicmatches[topicmatches[,1] > 0, 1], rep(1, length(topicmatches[topicmatches[,1] >
0, 1])))
  names(rel_df) <- c("TopicNumber", paste0("rel_",i))
  topic_rel <- dplyr::left_join(topic_rel, rel_df, by = "TopicNumber")
}

topic_rel$rel_sum <- rowSums(topic_rel[,4:length(topic_rel)], na.rm = TRUE)
table(topic_rel$rel_sum)

keep_rel <- topic_rel[topic_rel$rel_sum >= 8, 1] # numbers of topics that occurred in at least 80 % of other
inference runs
length(keep_rel)

View(topic_all[keep_rel,]) # View reliable topics
```
```` ``` ````
````

Appendix K (continued)

````
### Drop Excluded Topics

```{r}
# TopicNumbers of excluded topics
drop <- c(4,6,7)
keep <- keep_rel[!(keep_rel %in% drop)] # check which topic numbers of keep_rel are not included in drop
k_incl <- length(keep) # number of included topics

# data frame with topic information of included topics and NEW TOPIC NUMBERS
topics <- topic_all[keep,]
topics$OriginalNumber <- topics$TopicNumber
topics$TopicNumber <- 1:k_incl
rownames(topics) <- NULL


# terms and beta probabilities of included topics
terms_incl <- terms[keep]
View(terms_incl)

```

### Final Reliable Model

```{r}
library(textplot)
library(ggraph)
set.seed(1128)


 model_keep <- plot(model, top_n = 15, which = c(1, 2, 3, 5, 8, 9), # only keep reliable topics
    title = "Biterm Topic Model: Topics Emergent in Engineering Students' Team-Member Feedback",
    labels = c('1. Timely Communication',
              '2. Idea Generation',
              '3. Coordination and Adaptation',
              '', # put '' as placeholders for removed topics
              '4. Work Quality',
              '',
              '',
              '5. Team Support & Focusing',
              '6. Work Accountability'))


ggsave("plot_model_s26_3.png", plot = model_keep)
```
## Theta Scores

The theta values represent the distribution of topics for a specific document or a set of documents. These values indicate the proportion of each topic in the given text.

```{r}
````
````

Appendix K (continued)

```
#full model theta
theta_full <- predict(model, newdata = x_token)
dim(theta_full)

# keep only included topics
theta <- theta_full[,keep] # note that docs are ordered differently in theta.
dim(theta)
View(theta)

theta <- as.data.frame(theta)
names(theta) <- 1:k_incl # assign New Topic Numbers
colnames(theta)[1:6] <- c("Topic1", "Topic2", "Topic3", "Topic4","Topic5","Topic6") # name the topics



```

Merge theta with feedback information

```{r}


# Create a common identifier in the 'original_text' dataset
anno$doc_id <- gsub("doc", "", anno$doc_id)

# Create a common identifier in the 'theta' dataset
theta$doc_id <- gsub("doc", "", rownames(theta))

# Merge the 'original_text' and 'anno' datasets based on the 'doc_id' column
theta_merged <- merge(theta, anno, by = "doc_id")

View(theta_merged)
# Merge the 'merged_data' with the 'theta' dataset based on the 'doc_id' column
theta_merged <- theta_merged[, c(1:7, 10)]

# retain only unique rows
theta_merged <- unique(theta_merged)

dim(theta_merged)

#save theta tables

# Save theta_merged as a CSV file with a specific filename
write.csv(theta_merged, file = "theta_merged_s26.csv", row.names = FALSE)

```


TOPIC MODEL REGRESSION ANALYSIS
```
```

Appendix K (continued)

Preparing the data
```{r}

# Merge thetas of topic model feedback and performance values for each participant

merged_data <- merge(theta_merged, feedback_full, by = "doc_id")

colnames(merged_data)[2:7] <- c("Topic1", "Topic2", "Topic3", "Topic4","Topic5","Topic6")

# Save theta_merged as a CSV file with a specific filename

write.csv(merged_data, file = "reg_full_s26.csv", row.names = FALSE)

# Average the thetas for each participant

avrg_merged_data <- merged_data %>%
  group_by(ID, `GroupID`) %>%
  summarise(
    Topic1 = mean(Topic1, na.rm = TRUE),
    Topic2 = mean(Topic2, na.rm = TRUE),
    Topic3 = mean(Topic3, na.rm = TRUE),
    Topic4 = mean(Topic4, na.rm = TRUE),
    Topic5 = mean(Topic5, na.rm = TRUE),
    Topic6 = mean(Topic6, na.rm = TRUE),
    `T1 PeerFeedAvg` = first(`T1 PeerFeedAvg`),
    `T1 CommunicationAvg` =first(`T1 CommunicationAvg`),
    `T1 KSAAvg` =first(`T1 KSAAvg`),
    `T1 CommitmentAvg`=first(`T1 CommitmentAvg`),
    `T1 StandardsAvg`=first(`T1 StandardsAvg`),
    `T1 FocusAvg`=first(`T1 FocusAvg`) ,
    `T2 PeerFeedAvg` = first(`T2 PeerFeedAvg`),
    `T2 CommunicationAvg` =first(`T1 CommunicationAvg`),
    `T2 KSAAvg` =first(`T1 KSAAvg`),
    `T2 CommitmentAvg`=first(`T1 CommitmentAvg`),
    `T2 StandardsAvg`=first(`T1 StandardsAvg`),
    `T2 FocusAvg`=first(`T1 FocusAvg`)
  )

# Save theta_merged as a CSV file with a specific filename
write.csv(avrg_merged_data, file = "reg_avrgs_s26.csv", row.names = FALSE)
```

Appendix K (continued)

````
```{r}

# Load the packages
library(olsrr)
library(jtools)
library(moments)
library(lmtest)
library(car)
library(betareg)
library(performance)
library(lme4)
library(sjPlot)
library(glmmTMB)
library(apaTables)
```

## Linear Mixed-effects Regression Model

 Running the multiple regression to test the impact of getting feedback on topics in the topic model at time 1 and the outcome of peer-rating scores of teamwork abilities at time 2, while controlling for similar ratings at time 1
```{r}
# Run a mixed-effects regression model
mixed_model <- lmer(`T2 PeerFeedAvg` ~ Topic1 + Topic2 + Topic3 + Topic4 +Topic5 + Topic6 + `T1 PeerFeedAvg` + (1 | `Group ID`), data = avrg_merged_data)

my_confidence <- 0.95 # confidence level
my_digits <- 3 # number of decimals


# Model summary
summ(mixed_model) # unstandardized results
summ(mixed_model, scale = T, transform.response = T,  digits = my_digits) # standardized results

# F values
anova(mixed_model)

# Effect Size
prop_varience <-  r2(mixed_model)
print(prop_varience) # so the fixed effect explains 69..8%  percent of the variance in the outcome variable
                  # and the whole model including both the fixed and random effect explains 55.3% of variance

# AIC, BIC, ICC
performance(mixed_model)

# Save the model
saveRDS(mixed_model, file = "reg_results_s26.rds")


```
````

Appendix K (continued)

````
## Checking Assumptions

```{r}

# Homogeneity of Variance of Residuals (Homoscedasticity).
## when plotting the residuals we want to see randomness. If there appears to be non-random patterns this
may be one reason we don't find effects.

residuals <- residuals(mixed_model)
plot(residuals) # Assumption of Homoscedasticity appears to be violated
plot_model(mixed_model, type='diag')


# Normal Distribution of Residuals using qq plot and histogram
## As sample size goes up we might expect more normality

qqnorm(residuals)
qqline(residuals)
hist(residuals) # QQ plot and histograms of the residual data suggest a normal distribution, but with
outlines.
# future research could remove the outlines to see if the model improves.

# Multicollinearity

 car::vif(mixed_model) # Assumption met, no multicollinearity
# A score of above 10 indicates high multicollinearity, suggesting that the predictor variable is highly
correlated with other predictors
```

## Checking Changes in Peer Feedback Survey Score Between T1 & T2

```{r}
# Paired sample t-test
peersurvey_ttest <- t.test(avrg_merged_data$`T1 PeerFeedAvg`, avrg_merged_data$`T2 PeerFeedAvg`,
paired = TRUE, na.rm = TRUE)

print(peersurvey_ttest)


mean_time1 <- mean(avrg_merged_data$`T1 PeerFeedAvg`)
mean_time2 <- mean(avrg_merged_data$`T2 PeerFeedAvg`)

# Print the mean scores
cat("Mean at Time 1:", mean_time1)
cat("Mean at Time 2:", mean_time2)

#The paired samples t-test comparing average peer feedback scores at time 1 and time 2, revealed no
statistically significant difference, t(514) = -1.4747, p = 0.1409. The 95% confidence interval for the mean
difference ranged from -0.0641 to 0.0091. The sample estimate for the mean difference was -0.0275. The
mean score at Time 1 was 4.43831, and the mean score at Time 2 was 4.474427.
````
````

Appendix K (continued)

```
# scatterplot
ggplot(avrg_merged_data, aes(x = avrg_merged_data$`T1 PeerFeedAvg`, y = avrg_merged_data$`T2
PeerFeedAvg`)) +
  geom_point() +
  labs(title = "Scatterplot of Scores at Time 1 and Time 2",
       x = "Time 1 Scores",
       y = "Time 2 Scores")

```

SENTIMENT ANALYSIS

Calculate Polarity Scores
```{r}
library(pacman)
library(sentimentr)
library(syuzhet)
library(SentimentAnalysis)

# split the sentences from each feedback entry, keep the relevant doc_id
sentences <- map2(merged_data$Feedback, merged_data$doc_id, ~ data.frame(sentence =
get_sentences(.x), doc_id = .y))
sentences_df <- do.call(rbind, sentences)

# remove empty cells
sentences_df <- sentences_df %>%
  filter(grepl("[a-zA-Z]", sentence))

# calculate sentiment score for each sentence

SA_doc <- sentences_df %>%
  mutate(sentiment_score = sentiment(sentence)$sentiment)
SA_doc1 <- sentences_df %>%
  mutate(sentiment_score = sentiment(sentence, neutral.nonverb.like = T,valence_shifters_dt =
T,adversative.weight = T,n.before = Inf,n.after = Inf,amplifier.weight = T)$sentiment)

# now calculate the average scores for each document using doc_id as a grouping variable
SA_doc_id1  <- sentences_df$sentence %>% sentiment_by(sentences_df$doc_id)

# merge back with data

SA_merged_data1 <- merge(merged_data, SA_doc_id, by = "doc_id")

# save data
write.csv(SA_merged_data, file = "SA_data_scaled.csv", row.names = FALSE)


# get highlights of data
```

Appendix K (continued)

```
SA_doc1$sentence %>%
  get_sentences() %>%
  sentiment_by() %>%
  highlight()

# sentiment plot
SA_doc1 %>%
  ggplot()+geom_density(aes(sentiment_score))+
  ggtitle("Distribution of Sentiment Scores of Sentences in Peer Feedback")

SA_doc_id1 %>%
  ggplot(aes(x = ave_sentiment)) +
  geom_density() +
  geom_vline(xintercept = quantile(SA_doc_id1$ave_sentiment, 0.25), color = "red", linetype = "dashed",
size = 0.5) + # Q1 (25th percentile)
  geom_vline(xintercept = quantile(SA_doc_id1$ave_sentiment, 0.5), color = "green", linetype = "dashed",
size = 0.5) + # Q2 (50th percentile, median)
  geom_vline(xintercept = quantile(SA_doc_id1$ave_sentiment, 0.75), color = "blue", linetype = "dashed",
size = 0.5) + # Q3 (75th percentile)
  geom_text(aes(x = quantile(SA_doc_id1$ave_sentiment, 0.25), y = 0.5, label = paste("Q1: ",
round(quantile(SA_doc_id1$ave_sentiment, 0.25), 2))), vjust = -0.5, color = "red") +
  geom_text(aes(x = quantile(SA_doc_id1$ave_sentiment, 0.5), y = 0.7, label = paste("Q2: ",
round(quantile(SA_doc_id1$ave_sentiment, 0.5), 2))), vjust = -0.5, color = "green") +
  geom_text(aes(x = quantile(SA_doc_id1$ave_sentiment, 0.75), y = 0.3, label = paste("Q3: ",
round(quantile(SA_doc_id1$ave_sentiment, 0.75), 2))), vjust = -0.5, color = "blue") +
  ggtitle("Distribution of Averaged Valence Scores for Each (N = 1,385) Document") +
  xlab("Average Valence Score")+
  ylab("Density")

SA_doc_id1 %>%
  ggplot(aes(x = word_count)) +
  geom_density() +
  geom_vline(xintercept = quantile(SA_doc_id1$word_count, 0.25), color = "red", linetype = "dashed", size
= 0.5) + # Q1 (25th percentile)
  geom_vline(xintercept = quantile(SA_doc_id1$word_count, 0.5), color = "green", linetype = "dashed",
size = 0.5) + # Q2 (50th percentile, median)
  geom_vline(xintercept = quantile(SA_doc_id1$word_count, 0.75), color = "blue", linetype = "dashed",
size = 0.5) + # Q3 (75th percentile)
  geom_text(aes(x = quantile(SA_doc_id1$word_count, 0.25), y =  0.027, label = paste("Q1: ",
round(quantile(SA_doc_id1$word_count, 0.25), 2))), vjust = -1, color = "red") +
  geom_text(aes(x = quantile(SA_doc_id1$word_count, 0.5), y =  0.026, label = paste("Q2: ",
round(quantile(SA_doc_id1$word_count, 0.5), 2))), vjust = -1, color = "green") +
  geom_text(aes(x = quantile(SA_doc_id1$word_count, 0.75), y = 0.02, label = paste("Q3: ",
round(quantile(SA_doc_id1$word_count, 0.75), 2))), vjust = -1, color = "blue") +
  ggtitle("Distribution of Word Count for Each (N = 1,385) Document") +
  xlab("Average Word Count") +
  ylab("Density")

  ```
```
```

Appendix K (continued)

```
# POLARITY SCORE REGRESSION ANALYSIS

## Date Preperation
```{r}

SA_reg_avrg1 <- SA_merged_data1 %>%
  group_by(ID, `GroupID`) %>%
  summarise(
    ave_sentiment = mean(ave_sentiment, na.rm = TRUE),
    word_count = mean(word_count, na.rm = TRUE),
    Topic1 = mean(Topic1, na.rm = TRUE),
    Topic2 = mean(Topic2, na.rm = TRUE),
    Topic3 = mean(Topic3, na.rm = TRUE),
    Topic4 = mean(Topic4, na.rm = TRUE),
    Topic5 = mean(Topic5, na.rm = TRUE),
    Topic6 = mean(Topic6, na.rm = TRUE),
    `T1 PeerFeedAvg` = first(`T1 PeerFeedAvg`),
    `T1 CommunicationAvg` =first(`T1 CommunicationAvg`),
    `T1 KSAAvg` =first(`T1 KSAAvg`),
    `T1 CommitmentAvg`=first(`T1 CommitmentAvg`),
    `T1 StandardsAvg`=first(`T1 StandardsAvg`),
    `T1 FocusAvg`=first(`T1 FocusAvg`) ,
    `T2 PeerFeedAvg` = first(`T2 PeerFeedAvg`),
    `T2 CommunicationAvg` =first(`T1 CommunicationAvg`),
    `T2 KSAAvg` =first(`T1 KSAAvg`),
    `T2 CommitmentAvg`=first(`T1 CommitmentAvg`),
    `T2 StandardsAvg`=first(`T1 StandardsAvg`),
    `T2 FocusAvg`=first(`T1 FocusAvg`)
  )

write.csv(SA_reg_avrg, file = "SA_reg_avrg_scaled.csv", row.names = FALSE)
```

Mixed-Effects Polynomial Regression
```{r}
# Fit the mixed-effects polynomial regression model

model_SA <- lmer(`T2 PeerFeedAvg` ~ poly(ave_sentiment, 2)+ word_count+`T1 PeerFeedAvg`+ (1 |
`GroupID`),
            data = SA_reg_avrg)


my_confidence <- 0.95 # confidence level
my_digits <- 3 # number of decimals

# Model Summary
   # unstandardized results
summ(model_SA, confint = T,  ci.width = my_confidence, digits = my_digits)

 # standardized results
```
```

Appendix K (continued)

```
summm(model_SA, scale = T, transform.response = T,  digits = my_digits)

# F values
anova(model_SA)

# Effect Size
prop_varience_sa <-  r2(model_SA)
print(prop_varience_sa) # so the fixed effect explains 69.3%  percent of the variance in the outcome
variable
                 # and the whole model including both the fixed and random effect explains 55.2% of variance

# AIC, BIC, ICC
performance(model_SA)


# Save the model
saveRDS(model_SA, file = "SA_reg_results_s26.rds")


# scatterplot for valence
ggplot(SA_reg_avrg, aes(x = SA_reg_avrg$ave_sentiment, y = SA_reg_avrg$`T2 PeerFeedAvg`)) +
  geom_point() +
  labs(title = "Scatterplot of Valence Score at Time 1 and Peer Feedback Scores at Time 2",
       x = "Average Feedback Valence Score",
       y = "Peer Feedback Survey Scores")

# scatterplot for word count
ggplot(SA_reg_avrg, aes(x = SA_reg_avrg$word_count, y = SA_reg_avrg$`T2 PeerFeedAvg`)) +
  geom_point() +
  labs(title = "Scatterplot of Feedback Length at Time 1 and Peer Feedback Score at Time 2",
       x = "Average Feedback Length",
       y = "Peer Feedback Survey Scores")
```

Checking Assumptions

```{r}

# this should all be the same as for assumptions checked during topic model regression analysis.


# Homogeneity of Variance of Residuals (Homoscedasticity).
## when plotting the residuals we want to see randomness. If there appears to be non-random patterns this
may be one reason we don't find effects.

residuals_sa <- residuals(model_SA)
plot(residuals_sa) # Assumption of Homoscedasticity appears to be violated
plot_model(model_SA, type='diag')
```

Appendix K (continued)

```
# Normal Distribution of Residuals using qq plot and histogram
## As sample size goes up we might expect more normality

qqnorm(residuals_sa)
qqline(residuals_sa)
hist(residuals_sa) # QQ plot and histograms of the residual data suggest a normal distribution, but with outlines.
# future research could remove the outlines to see if the model improves.

# Multicollinearity

 car::vif(model_SA) # Assumption met, no multicollinearity
# A score of above 10 indicates high multicollinearity, suggesting that the predictor variable is highly correlated with other predictors
```

WORKS CITED

ABET. (2021). *Criteria for Accrediting Engineering Programs, 2022 – 2023*. https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2022-2023/

Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis. *Frontiers in Artificial Intelligence*, *3*. https://doi.org/10.3389/frai.2020.00042

Alsharif, A., Katz, A., Knight, D., & Alatwah, S. (2022). *Using Sentiment Analysis to Evaluate First-year Engineering Students Teamwork Textual Feedback*. 2022 ASEE Annual Conference & Exposition.

Anson, I. G., & Anson, C. M. (2017). Assessing peer and instructor response to writing: A corpus analysis from an expert survey. *Assessing Writing*, *33*, 12–24. https://doi.org/10.1016/j.asw.2017.03.001

Anson, R., & Goodman, J. A. (2014). A Peer Assessment System to Improve Student Team Experiences. *Journal of Education for Business*, *89*(1), 27–34. https://doi.org/10.1080/08832323.2012.754735

Bacon, D. R., Stewart, K. A., & Silver, W. S. (1999). Lessons from the Best and Worst Student Team Experiences: How a Teacher can make the Difference. *Journal of Management Education*, *23*(5), 467–488. https://doi.org/10.1177/105256299902300503

Baker, D. F. (2008). Peer Assessment in Small Groups: A Comparison of Methods. *Journal of Management Education*, *32*(2), 183–209. https://doi.org/10.1177/1052562907310489

Balahadia, F. F., Fernando, M. C. G., & Juanatas, I. C. (2016). *Teacher's performance evaluation tool using opinion mining with sentiment analysis*. 95–98.

Bell, B. S., & Kozlowski, S. W. J. (2011). Collective failure: The emergence, consequences, and management of errors in teams. In *Errors in organizations.* (pp. 113–141). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9780203817827

Bittermann, A., Batzdorfer, V., Müller, S. M., & Steinmetz, H. (2021). Mining Twitter to Detect Hotspots in Psychology. *Zeitschrift Für Psychologie*, *229*(1), 3–14. https://doi.org/10.1027/2151-2604/a000437

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Bloom, A. J., & Hautaluoma, J. E. (1987). Effects of Message Valence, Communicator Credibility, and Source Anonymity on Reactions to Peer Feedback. *The Journal of Social Psychology*, *127*(4), 329–338. https://doi.org/10.1080/00224545.1987.9713712

Boud, D., Ajjawi, R., Dawson, P., & Tai, J. (2018). *Developing evaluative judgement in higher education: Assessment for knowing and producing quality work*. Routledge.

Bracken, D. W., Timmreck, C. W., Fleenor, J. W., & Summers, L. (2001). 360 Feedback from Another Angle. *Human Resource Management*, *40*(1), 3–20. https://doi.org/10.1002/hrm.4012

Brockx, B., Van Roy, K., & Mortelmans, D. (2012). The Student as a Commentator: Students' Comments in Student Evaluations of Teaching. *International Conference on Education & Educational Psychology (ICEEPSY 2012)*, *69*, 1122–1133. https://doi.org/10.1016/j.sbspro.2012.12.042

Brutus, S., & Donia, M. B. L. (2010). Improving the effectiveness of students in groups with a centralized peer evaluation system. *Academy of Management Learning & Education*, *9*(4), 652–662. https://doi.org/10.5465/AMLE.2010.56659882

Brutus, S., Donia, M. B. L., & Ronen, S. (2013). Can business students learn to evaluate better? Evidence from repeated exposure to a peer-evaluation system. *Academy of Management Learning & Education*, *12*(1), 18–31. https://doi.org/10.5465/amle.2010.0204

Burgess, A., Roberts, C., Lane, A. S., Haq, I., Clark, T., Kalman, E., Pappalardo, N., & Bleasel, J. (2021). Peer review in team-based learning: Influencing feedback literacy. *BMC Medical Education*, *21*(1), 426. https://doi.org/10.1186/s12909-021-02821-6

Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (2017). *A practical guide to sentiment analysis* (Vol. 5). Springer.

Chen, G., Donahue, L. M., & Klimoski, R. J. (2004). Training Undergraduates to Work in Organizational Teams. *Academy of Management Learning & Education*, *3*(1), 27–40. https://doi.org/10.5465/AMLE.2004.12436817

Crawford, E. R., & LePine, J. A. (2013). A Configural Theory of Team Processes: Accounting for the Structure of Taskwork and Teamwork. *Academy of Management Review*, *38*(1), 32–48. https://doi.org/10.5465/amr.2011.0206

Creswell, J. W. (2008). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications, Incorporated.

Croy, G., & Eva, N. (2018). Student success in teams: Intervention, cohesion and performance. *Education + Training*, *60*(9), 1041–1056. https://doi.org/10.1108/ET-11-2017-0174

Cushing, A., Abbott, S., Lothian, D., Hall, A., & Westwood, O. M. R. (2011). Peer feedback as an aid to learning – What do we want? Feedback. When do we want it? Now! *Medical Teacher*, *33*(2), e105–e112. https://doi.org/10.3109/0142159X.2011.542522

David, E. M. (2013). Examining the Role of Narrative Performance Appraisal Comments on Performance. *Human Performance*, *26*(5), 430–450. https://doi.org/10.1080/08959285.2013.836197

Debortoli, S., Müller, O., Junglas, I., & Vom Brocke, J. (2016). Text mining for information systems researchers: An annotated topic modeling tutorial. *Communications of the Association for Information Systems (CAIS)*, *39*(1), 7.

Denny, M. J., & Spirling, A. (2018). Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. *Political Analysis*, *26*(2), 168–189. Cambridge Core. https://doi.org/10.1017/pan.2017.44

DeShon, R. P., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004). A Multiple-Goal, Multilevel Model of Feedback Effects on the Regulation of Individual and Team Performance. *Journal of Applied Psychology*, *89*(6), 1035–1056. https://doi.org/10.1037/0021-9010.89.6.1035

Dominick, P. G., Reilly, R. R., & Mcgourty, J. W. (1997). The Effects of Peer Feedback on Team Member Behavior. *Group & Organization Management*, *22*(4), 508–520. https://doi.org/10.1177/1059601197224006

Donia, M. B. L., Mach, M., O'Neill, T. A., & Brutus, S. (2022). Student satisfaction with use of an online peer feedback system. *Assessment & Evaluation in Higher Education*, *47*(2), 269–283. https://doi.org/10.1080/02602938.2021.1912286

Donia, M. B. L., O'Neill, T. A., & Brutus, S. (2018). The longitudinal effects of peer feedback in the development and transfer of student teamwork skills. *Learning and Individual Differences*, *61*, 87–98. https://doi.org/10.1016/j.lindif.2017.11.012

Driskell, T., Driskell, J. E., Burke, C. S., & Salas, E. (2017). Team Roles: A Review and Integration. *Small Group Research*, *48*(4), 482–511. https://doi.org/10.1177/1046496417711529

Dyer, W. G., Dyer Jr., W. G., & Dyer, J. H. (2007). *Team building: Proven strategies for improving team performance, 4th ed.* (pp. xviii, 240). Jossey-Bass/Wiley.

Einhorn, L. (1978). Oral and written style: An examination of differences. *Southern Speech Communication Journal*, *43*(3), 302–311. https://doi.org/10.1080/10417947809372388

Ellis, S., & Davidi, I. (2005). After-Event Reviews: Drawing Lessons From Successful and Failed Experience. *Journal of Applied Psychology*, *90*(5), 857–871. https://doi.org/10.1037/0021-9010.90.5.857

Falchikov, N., & Goldfinch, J. (2000). Student Peer Assessment in Higher Education: A Meta-Analysis Comparing Peer and Teacher Marks. *Review of Educational Research*, *70*(3), 287–322. https://doi.org/10.3102/00346543070003287

Ferris, D. R. (2014). Responding to student writing: Teachers' philosophies and practices. *Feedback in Writing: Issues and Challenges*, *19*, 6–23. https://doi.org/10.1016/j.asw.2013.09.004

Fong, C. J., Patall, E. A., Vasquez, A. C., & Stautberg, S. (2019). A Meta-Analysis of Negative Feedback on Intrinsic Motivation. *Educational Psychology Review*, *31*(1), 121–162. https://doi.org/10.1007/s10648-018-9446-6

Gabelica, C., Bossche, P. V. den, Segers, M., & Gijselaers, W. (2012). Feedback, a powerful lever in teams: A review. *Educational Research Review*, *7*(2), 123–144. https://doi.org/10.1016/j.edurev.2011.11.003

Gottipati, S., Shankararaman, V., & Gan, S. (2017). A Conceptual Framework for Analyzing Students' Feedback. *2017 IEEE Frontiers in Education Conference (FIE)*, 1–8. https://doi.org/10.1109/FIE.2017.8190703

Gottipati, S., Shankararaman, V., & Lin, J. R. (2018). Text analytics approach to extract course improvement suggestions from students' feedback. *Research and Practice in Technology Enhanced Learning*, *13*(1), 6. https://doi.org/10.1186/s41039-018-0073-0

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *101*(suppl_1), 5228–5235. https://doi.org/10.1073/pnas.0307752101

Häkkinen, J., & Ramadan, Z. (2023). *A Study on the Perception of Feedback with Varying Sentiment Generated Using a Large Language Model*.

Handke, L., Klonek, F., O'Neill, T. A., & Kerschreiter, R. (2022). Unpacking the Role of Feedback in Virtual Team Effectiveness. *Small Group Research*, *53*(1), 41–87. https://doi.org/10.1177/10464964211057116

Hattie, J., Crivelli, J., Van Gompel, K., West-Smith, P., & Wike, K. (2021). *Feedback that leads to improvement in student essays: Testing the hypothesis that "Where to Next" feedback is most powerful*. 182.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hodges, L. C., & Stanton, K. (2007). Translating Comments on Student Evaluations into the Language of Learning. *Innovative Higher Education*, *31*(5), 279–286. https://doi.org/10.1007/s10755-006-9027-3

Hughes, R. L., & Jones, S. K. (2011). Developing and assessing college student teamwork skills. *New Directions for Institutional Research*, *2011*(149), 53–64. https://doi.org/10.1002/ir.380

Kapp, E. (2009). Improving Student Teamwork in a Collaborative Project-Based Course. *College Teaching*, *57*(3), 139–143. https://doi.org/10.3200/CTCH.57.3.139-143

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Koenka, A. C., Linnenbrink-Garcia, L., Moshontz, H., Atkinson, K. M., Sanchez, C. E., & Cooper, H. (2021). A meta-analysis on the impact of grades and comments on academic motivation and achievement: A case for written feedback. *Educational Psychology*, *41*(7), 922–947. https://doi.org/10.1080/01443410.2019.1659939

Kommission, E. (2011). *Flash Eurobarometer 304 (Employers' Perception of Graduate Employability)*. https://doi.org/10.4232/1.10241

Kozlowski, S. W. J., & Bell, B. S. (2003). Work groups and teams in organizations. In *Handbook of psychology: Industrial and organizational psychology, Vol. 12.* (pp. 333–375). John Wiley & Sons Inc.

Krenn, B., Würth, S., & Hergovich, A. (2013). The impact of feedback on goal setting and task performance: Testing the feedback intervention theory. *Swiss Journal of Psychology*, *72*(2), 79–89. https://doi.org/10.1024/1421-0185/a000101

Leshed, G., Hancock, J. T., Cosley, D., McLeod, P. L., & Gay, G. (2007). Feedback for Guiding Reflection on Teamwork Practices. *Proceedings of the 2007 ACM International Conference on Supporting Group Work*, 217–220. https://doi.org/10.1145/1316624.1316655

London, M., & Sessa, V. I. (2006). Group Feedback for Continuous Learning. *Human Resource Development Review*, *5*(3), 303–329. https://doi.org/10.1177/1534484306290226

London, M., & Smither, J. W. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory-based applications and directions for research. *Personnel Psychology*, *48*(4), 803–839. https://doi.org/10.1111/j.1744-6570.1995.tb01782.x

Loughry, M. L., Ohland, M. W., & DeWayne Moore, D. (2007). Development of a Theory-Based Assessment of Team Member Effectiveness. *Educational and Psychological Measurement*, *67*(3), 505–524. https://doi.org/10.1177/0013164406292085

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, *12*(2–3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Mandouit, L., & Hattie, J. (2023). Revisiting “The Power of Feedback” from the perspective of the learner. *Learning and Instruction*, *84*, 101718. https://doi.org/10.1016/j.learninstruc.2022.101718

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A Temporally Based Framework and Taxonomy of Team Processes. *The Academy of Management Review*, *26*(3), 356–376. JSTOR. https://doi.org/10.2307/259182

Mathieu, J. E., Gallagher, P. T., Domingo, M. A., & Klock, E. A. (2019). Embracing Complexity: Reviewing the Past Decade of Team Effectiveness Research. *Annual Review of Organizational Psychology and Organizational Behavior*, *6*(1), 17–46. https://doi.org/10.1146/annurev-orgpsych-012218-015106

Mathieu, J. E., Tannenbaum, S. I., Donsbach, J. S., & Alliger, G. M. (2014). A Review and Integration of Team Composition Models: Moving Toward a Dynamic and Temporal Framework. *Journal of Management*, *40*(1), 130–160. https://doi.org/10.1177/0149206313503014

Mathieu, J. E., Tannenbaum, S. I., Kukenberger, M. R., Donsbach, J. S., & Alliger, G. M. (2015). Team Role Experience and Orientation: A Measure and Tests of Construct Validity. *Group & Organization Management*, *40*(1), 6–34. https://doi.org/10.1177/1059601114562000

Meddeb, A., & Romdhane, L. B. (2022). Using Topic Modeling and Word Embedding for Topic Extraction in Twitter. *Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 26th International Conference KES2022*, *207*, 790–799. https://doi.org/10.1016/j.procs.2022.09.134

Millis, B. J., & Cottell, P. G., Jr. (1998). *Cooperative Learning for Higher Education Faculty. Series on Higher Education.* Oryx Press.

Misuraca, M., Forciniti, A., Scepi, G., & Spano, M. (2020). Sentiment Analysis for Education with R: packages, methods and practical applications. *arXiv Preprint arXiv:2005.12840*.

Mite-Baidal, K., Delgado-Vera, C., Solís-Avilés, E., Espinoza, A. H., Ortiz-Zambrano, J., & Varela-Tapia, E. (2018). Sentiment Analysis in Education Domain: A Systematic Literature Review. In R. Valencia-García, G. Alcaraz-Mármol, J. Del Cioppo-Morstadt, N. Vera-Lucio, & M. Bucaram-Leverone (Eds.), *Technologies and Innovation* (pp. 285–297). Springer International Publishing.

Mohr, J. W., & Bogdanov, P. (2013). Introduction—Topic models: What they are and why they matter. *Topic Models and the Cultural Sciences*, *41*(6), 545–569. https://doi.org/10.1016/j.poetic.2013.10.001

Mulder, R., Baik, C., Naylor, R., & Pearce, J. (2014). How does student peer review influence perceptions, engagement and academic outcomes? A case study. *Assessment & Evaluation in Higher Education*, *39*(6), 657–677. https://doi.org/10.1080/02602938.2013.860421

Murphy, K. R., & Cleveland, J. N. (1991). *Performance appraisal: An organizational perspective.* Allyn & Bacon.

Naldi, M. (2019). A review of sentiment computation methods with R packages. *arXiv Preprint arXiv:1901.08319*.

Nanda, G., A. Douglas, K., R. Waller, D., E. Merzdorf, H., & Goldwasser, D. (2021). Analyzing Large Collections of Open-Ended Feedback From MOOC Learners Using LDA Topic Modeling and Qualitative Analysis. *IEEE Transactions on Learning Technologies*, *14*(2), 146–160. https://doi.org/10.1109/TLT.2021.3064798

Nanda, G., Wei, S., Katz, A., Brinton, C., & Ohland, M. (2022). *Work-in-Progress: Using Latent Dirichlet Allocation to uncover themes in student comments from peer evaluations of teamwork*. 2022 ASEE Annual Conference & Exposition.

Niekler, A. (2016). *Automatisierte Verfahren für die Themenanalyse nachrichtenorientierter Textquellen*. https://doi.org/10.13140/RG.2.2.28090.39366

Oakley, B., Felder, R. M., Brent, R., & Elhajj, I. (2004). Turning student groups into effective teams. *Journal of Student Centered Learning*, *2*(1), 9–34.

Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Felder, R. M., Finelli, C. J., Layton, R. A., Pomeranz, H. R., & Schmucker, D. G. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self- and peer evaluation. *Academy of Management Learning & Education*, *11*(4), 609–630. https://doi.org/10.5465/amle.2010.0177

ONeill, T. A., Boyce, M., & McLarnon, M. J. (2020). *Team health and project quality are improved when peer evaluation scores affect grades on team projects*. *5*, 49.

O'Neill, T. A., & Salas, E. (2018). Creating high performance teamwork in organizations. *Creating High Performance Teamwork in Organizations*, *28*(4), 325–331. https://doi.org/10.1016/j.hrmr.2017.09.001

O'Neill, T., Larson, N., Smith, J., Donia, M., Deng, C., Rosehart, W., & Brennan, R. (2019). Introducing a scalable peer feedback system for learning teams. *Assessment & Evaluation in Higher Education*, *44*(6), 848–862. https://doi.org/10.1080/02602938.2018.1526256

Passow, H. J. (2012). Which ABET Competencies Do Engineering Graduates Find Most Important in their Work? *Journal of Engineering Education*, *101*(1), 95–118. https://doi.org/10.1002/j.2168-9830.2012.tb00043.x

Petkova, A. P., Domingo, M. A., & Lamm, E. (2021). Let's be frank: Individual and team-level predictors of improvement in student teamwork effectiveness following peer-evaluation feedback. *The International Journal of Management Education*, *19*(3), 100538. https://doi.org/10.1016/j.ijme.2021.100538

Pienaar, J. W. (Cobus), & Nel, P. (2017). A conceptual framework for understanding leader self-schemas and the influence of those self-schemas on the integration of feedback. *South African Journal of Human Resource Management*, *15*. Gale Academic OneFile. https://link.gale.com/apps/doc/A541474855/AONE?u=anon~ef74aac4&sid=googleScholar&xid=4001b0e3

Pietsch, A.-S., & Lessmann, S. (2018). Topic modeling for analyzing open-ended survey responses. *Journal of Business Analytics*, *1*(2), 93–116. https://doi.org/10.1080/2573234X.2019.1590131

Pinargote-Ortega, Bowen-Mendoza, Meza, & Ventura. (2023). Sentiment Analysis Techniques for Peer Feedback: A Review. *2023 Ninth International Conference on eDemocracy & eGovernment (ICEDEG)*, 1–8. https://doi.org/10.1109/ICEDEG58167.2023.10122085

Planas-Lladó, A., Feliu, L., Arbat, G., Pujol, J., Suñol, J. J., Castro, F., & Martí, C. (2021). An analysis of teamwork based on self and peer evaluation in higher education. *Assessment & Evaluation in Higher Education*, *46*(2), 191–207. https://doi.org/10.1080/02602938.2020.1763254

Rapp, T., Maynard, T., Domingo, M., & Klock, E. (2021). Team Emergent States: What Has Emerged in The Literature Over 20 Years. *Small Group Research*, *52*(1), 68–102. https://doi.org/10.1177/1046496420956715

Rinker, T. (2021). *sentimentr: Calculate Text Polarity Sentiment* (2.9.0) [Computer software]. https://cran.r-project.org/web/packages/sentimentr/index.html

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R Package for Structural Topic Models. *Journal of Statistical Software*, *91*(2), 1–40. https://doi.org/10.18637/jss.v091.i02

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology*, *52*, 471–499. https://doi.org/10.1146/annurev.psych.52.1.471

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In *Teams: Their training and performance.* (pp. 3–29). Ablex Publishing.

Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2015). Understanding and Improving Teamwork in Organizations: A Scientifically Based Practical Guide. *Human Resource Management*, *54*(4), 599–622. https://doi.org/10.1002/hrm.21628

Scriven, M. (1967). The Methodology of Evaluation. In *Tyler, R. W., Gagne, R. M., Scriven, M. (Eds), Perspectives of Curriculum Evaluation*. Rand McNally.

Shuffler, M. L., Diazgranados, D., Maynard, M. T., & Salas, E. (2018). Developing, sustaining, and maximizing team effectiveness: An integrative, dynamic perspective of team development interventions. *The Academy of Management Annals*, *12*(2), 688–724. https://doi.org/10.5465/annals.2016.0045

Smither, J. W., & Walker, A. G. (2004). Are the Characteristics of Narrative Comments Related to Improvement in Multirater Feedback Ratings Over Time? *Journal of Applied Psychology*, *89*(3), 575–581. https://doi.org/10.1037/0021-9010.89.3.575

Sridharan, B., & Boud, D. (2019). The effects of peer judgements on teamwork and self-assessment ability in collaborative group work. *Assessment & Evaluation in Higher Education*, *44*(6), 894–909. https://doi.org/10.1080/02602938.2018.1545898

Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012). *Exploring topic coherence over many models and many topics*. 952–961.

Sun, J., & Yan, L. (2023). *Use Topic Modeling to Understand Comments in Student Evaluations of Teaching*. https://doi.org/10.21203/rs.3.rs-2444380/v1

Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do team and individual debriefs enhance performance? A meta-analysis. *Human Factors*, *55*(1), 231–245. https://doi.org/10.1177/0018720812448394

Vajjala, S., Majumder, B., Gupta, A., & Surana, H. (2020). *Practical natural language processing: A comprehensive guide to building real-world NLP systems*. O'Reilly Media.

Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, *94*, 101582. https://doi.org/10.1016/j.is.2020.101582

Veine, S., Anderson, M. K., Andersen, N. H., Espenes, T. C., Søyland, T. B., Wallin, P., & Reams, J. (2020). Reflection as a core student learning activity in higher education—Insights from nearly two decades of academic development. *International Journal for Academic Development*, *25*(2), 147–161. https://doi.org/10.1080/1360144X.2019.1659797

Wessa, P., & De Rycker, A. (2010). *Reviewing peer reviews-A rule-based approach*. 408–418.

Wijffels, J., & Yan, X. (2023). *BTM: Biterm Topic Models for Short Text* (0.3.7) [Computer software]. https://cran.r-project.org/web/packages/BTM/index.html

Wind, D. K., & Jensen, U. A. (2017). Quantifying feedback: Insights into peer assessment data. *Proc. 12th Int. Conf. e-Learning*, 256.

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A Biterm Topic Model for Short Texts. *Proceedings of the 22nd International Conference on World Wide Web*, 1445–1456. https://doi.org/10.1145/2488388.2488514

Yu, Q., & Schunn, C. D. (2023). Understanding the what and when of peer feedback benefits for performance and transfer. *Computers in Human Behavior*, *147*, 107857. https://doi.org/10.1016/j.chb.2023.107857

Yüksel, İ., & Başaran, B. Ç. (2020). Reflective peer feedback in the practicum: Qualitative and quantitative practices. *Turkish Online Journal of Qualitative Inquiry*, *11*(1), 85–109.

Zhao, W., Chen, J. J., Perkins, R., Liu, Z., Ge, W., Ding, Y., & Zou, W. (2015). A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics*, *16*(13), S8. https://doi.org/10.1186/1471-2105-16-S13-S8

Zong, Z., Schunn, C. D., & Wang, Y. (2021). Learning to improve the quality peer feedback through experience with peer feedback. *Assessment & Evaluation in Higher Education*, *46*(6), 973–992. https://doi.org/10.1080/02602938.2020.1833179