Innovating the Study of Self-Regulated Learning: An Exploration through NLP, Generative AI, and LLMs

Yasir Gamieldien

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Engineering Education

Andrew S. Katz, Chair

Jennifer M. Case

Jacob R. Grohs

Rachel E. McCord Ellestad

August 11, 2023

Blacksburg, VA

Keywords: large language models, natural language processing, self-regulated learning, exam wrappers, qualitative analysis

Innovating the Study of Self-Regulated Learning: An Exploration through NLP, Generative AI, and LLMs

Yasir Gamieldien

ABSTRACT

This dissertation explores the use of natural language processing (NLP) and large language models (LLMs) to analyze student self-regulated learning (SRL) strategies in response to exam wrappers. Exam wrappers are structured reflection activities that prompt students to practice SRL after they get their graded exams back. The dissertation consists of three manuscripts that compare traditional qualitative analysis with NLP-assisted approaches using transformer-based models including GPT-3.5, a state-of-the-art LLM. The data set comprises 3,800 student responses from an engineering physics course. The first manuscript develops two NLP-assisted codebooks for identifying learning strategies related to SRL in exam wrapper responses and evaluates the agreement between them and traditional qualitative analysis. The second manuscript applies a novel NLP technique called zero-shot learning (ZSL) to classify student responses into the codes developed in the first manuscript and assesses the accuracy of this method by evaluating a subset of the full dataset. The third manuscript identifies the distribution and differences of learning strategies and SRL constructs among students of different exam performance profiles using the results from the second manuscript. The dissertation demonstrates the potential of NLP and LLMs to enhance qualitative research by providing scalable, robust, and efficient methods for analyzing large corpora of textual data. The dissertation also contributes to the understanding of SRL in engineering education by revealing the common learning strategies, impediments, and SRL constructs that students report they use while preparing for exams in a first-year engineering physics course. The dissertation suggests implications, limitations, and directions for future research on NLP, LLMs, and SRL.

Innovating the Study of Self-Regulated Learning: An Exploration through NLP, Generative AI, and LLMs

Yasir Gamieldien

GENERAL AUDIENCE ABSTRACT

This dissertation is about using artificial intelligence (AI) to help researchers and teachers understand how students learn from their exams. Exams are not only a way to measure what students know, but also a chance for students to reflect on how they studied and what they can do better next time. One way that students can reflect is by using exam wrappers, which are short questions that students answer after they get their graded exams back. A type of AI called natural language processing (NLP) is used in this dissertation, which can analyze text and find patterns and meanings in it. This study also uses a powerful AI tool called GPT-3.5, which can generate text and answer questions. The dissertation has three manuscripts that compare the traditional way of analyzing exam wrappers, which is done by hand, with the new way of using NLP and GPT-3.5, evaluate a specific promising NLP method, and use this method to try and gain a deeper understanding in students self-regulated learning (SRL) while preparing for exams. The data comes from 3,800 exam wrappers from a physics course for engineering students. The first manuscript develops a way of using NLP and GPT-3.5 to find out what learning strategies and goals students talk about in their exam wrappers and compares it to more traditional methods of analysis. The second manuscript tests how accurate a specific NLP technique is in finding these strategies and goals. The third manuscript looks at how different students use different strategies and goals depending on how well they did on the exams using the NLP technique in the second manuscript. I found that NLP and GPT-3.5 can aid in analyzing exam wrappers faster and provide nuanced insights when compared with manual approaches. The dissertation also shows what learning strategies and goals are most discussed for engineering students as they prepare for

exams. The dissertation gives some suggestions, challenges, and ideas for future research on AI and learning from exams.

.

Raisa, your support, love, and unwavering confidence in me kept me going.

Mommy, all I am today is the product of your love, care, perseverance, and belief.

Acknowledgments

This work would not have been possible without the people around me. My wife, Raisa, has been my constant cheerleader, accompanying me through each milestone and occasionally dragging me toward them. My mother, Yasmina, has been unconditionally supportive of every decision I have made, even those that have taken me to the farthest and most intriguing places.

I owe a massive thank you to Dr. Andrew Katz. You believed in me, supported me, and went above and beyond to advance an interesting timeline. You will forever remain in my memory as an inspiring mentor, teacher, and advisor.

Dr. Jennifer Case, our journey together - from the Department of Chemical Engineering at UCT to the Department of Engineering Education at VT - was nothing short of remarkable. I will never forget our coffee chat in Radford when I had just arrived in the US and was feeling homesick. I am eternally grateful for the opportunity you have presented me with and for your belief in me.

To the other members of my committee, Dr. Rachel McCord, thank you for your encouraging words, for teaching me about metacognition and self-regulated learning, and for tolerating my constant emails. Dr. Nicole Pitterson, thank you for enlightening me about conceptual change and qualitative research and for being part of my formative phases of research. Dr. Jacob Grohs, your willingness to join the process later and play a pivotal role in the change of direction of my dissertation has been invaluable. You also took the time to discuss my mini-studies before the qualifier and laid the foundations for my methodology in the field.

To my siblings, Ammaar, Nilfa, Naushaad, and Reza: Thank you for being my teachers, my role models, my friends, and my mentors. Living far away from home and you all, as well as the new

family members we keep welcoming, has been challenging. My heartfelt thanks also go to my new in-law's family, who welcomed me as one of their own.

My friends, the "Salt Gatherers," "Winterbrown," my fellow engineering education friends and peers, and the amazing friends from Cape Town and various places I have visited throughout my experiences: you all have a special place in my heart. Despite the distance, our communication never ceases. I am deeply grateful to everyone in Cape Town and Blacksburg who has been my family and support during this adventure. Although I may have inadvertently missed some names, I am forever grateful and humbled by our shared journey.

This material is based on work supported by the National Science Foundation under Grant No. EEC - 2017008. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Attributions

This dissertation is my original work where I am the primary author and researcher. I would like to, however, acknowledge the intellectual contributions of other researchers, who will receive due credit as co-authors when the manuscripts are submitted for publication. Manuscript 1 builds on work conducted by Dr. Andrew Katz who ignited my interest in NLP in qualitative analysis. I would also like to acknowledge Dr. Jennifer Case for her intellectual contribution to the qualitative analysis portion of the study. I would like to acknowledge Dr. Andrew Katz for guidance in Manuscript 2, for assistance with zero-shot analysis, and expertise in NLP. Dr. Rachel McCord provided insights into SRL in terms of the background and insights into analysis for both Manuscripts 2 and 3. I again acknowledge Dr. Andrew Katz for his intellectual contribution to the framing of Manuscript 3 in the methods and analysis of the results.

Declaration of Generative AI in Dissertation

The author used ChatGPT in order to check grammar and spelling in the preparation of this work. The author also used ChatGPT to provide ideas for the names of the manuscripts within the dissertation, as well as the title of the dissertation. ChatGPT was used as a supplemental search for literature. An example prompt could be “Please provide me with references that pertain to engineering students’ time management and test anxiety.” The model would respond with citations to which the author would then search for, read, and decide its relevance to the manuscript. Furthermore, in certain subsections of the dissertation, ChatGPT was used to suggest some subheadings that could be contained in the subsection. For example, asking the model to suggest some subheadings in the literature of Manuscript 1. The author then decided whether any of the suggestions could be used in the manuscript. After using ChatGPT, the author reviewed and edited the content as needed and took full responsibility for the content of the publication.

Table of Contents

Acknowledgments v
Table of Contents viii
List of Tables xii
List of Figures xiv
Common Acronyms in Dissertation xv
Chapter 1 – Introduction 1
1.1 Background and Context 1
1.1.1 Self-regulated learning 2
1.1.2 Exam Wrappers 3
1.1.3 Natural Language Processing 6
1.2 Statement of the Problem 8
1.3 Purpose of this Dissertation 10
1.4 Research Questions 12
1.5 Significance of the Research 13
1.6 Summary 14
References 16
Chapter 2: Manuscript 1 23
2.1 Structured Abstract 23
2.2 Introduction 25
2.3 Background 26
2.3.1 Qualitative analysis 26
2.3.2 Natural Language Processing 26
2.3.3 Large Language Models 27
2.3.4 NLP and Qualitative Analysis 28
2.3.5 Significance of this Study 30
2.4 Methods 30
2.4.1 Data Sources 31
2.4.2 Manual Analysis 33
2.4.3 NLPCA 33
2.4.4 NLPGPT 37
2.6 Discussion 49
2.6.1 Study Implications 51

2.7 Conclusion 53
References 54
Chapter 3: Manuscript 2 60
3.1 Abstract 60
3.1 Introduction 61
3.1.1 Self-Regulated Learning 61
3.1.2 Exam Wrappers 62
3.1.3 Modern Natural Language Processing 63
3.1.4 Natural Language Processing Using Transformers 64
3.1.5 Zero-shot Learning Classification 65
3.2 Methods 66
3.2.1 Data Collection 66
3.2.2 Data Analysis 68
3.2.2.1 Questions analyzed 68
3.2.3 Codebook Background 70
3.2.3.1 Method Evaluation 70
3.2.4 Ethical Considerations 71
3.3 Results 71
3.3.1 Evaluation Results 71
3.3.1.1 Other Skills 73
3.3.1.2 Learning Process 73
3.3.1.3 Starting Action 74
3.4 Discussion 75
3.4.1 Accuracy of Different Codebooks 76
3.4.2 Implications of the NLP Workflow in Studying SRL 77
3.4.3 Future Research 78
3.4.4 Limitations 79
3.5 Conclusion 80
References 82
Chapter 4: Manuscript 3 86
4.1 Structured Abstract 86
4.2 Introduction 88

4.3.1 Theoretical Framework – Self-Regulated Learning 89
4.3.1.1 Forethought phase - Task analysis 91
4.3.1.2 Forethought phase - Self-motivation beliefs 93
4.3.1.3 Performance phase – Self-control and self-observation 94
4.3.1.4 Self-Reflection Phase - Self-judgment 96
4.3.1.5 Self-Reflection phase - Self-reaction 97
4.3.2 Interventions for SRL 98
4.3.2.1 The Role of Instructors in Promoting SRL 99
4.3.2.2 Integrating SRL Strategies Into the Course Content 99
4.3.2.3 Reflective Writing and Exam Wrappers 100
4.3.3 Natural Language Processing 103
4.4 Methods 104
4.4.1 Context of Study and Participants 104
4.4.2 Data Collection 105
4.4.3 Data Analysis 107
4.4.4 Codebook Background 109
4.4.5 Zero-shot Classification 110
4.4.6 SRL Constructs Discussed in Exam Wrappers 112
4.4.7 Student Performance Profiles 114
4.4.8 Distributions of Learning Strategies and SRL Constructs Across Profiles 115
4.5 Results 115
4.5.1 Student Performance Profiles 116
4.5.2 Learning Strategy and SRL Construct Distribution 119
4.5.2.1 Learning Strategies Across Different Student Profiles - Non-Physics Skills 120
4.5.2.2 Learning Strategies Across Exams – Non-Physics Skills 123
4.5.2.3 Self-Regulated Learning Constructs – Non-Physics Skills 126
4.5.2.4 Learning Strategies Across Different Student Profiles - Preparation Process 131
4.5.2.5 Learning Strategies Across Exams – Preparation Process 134
4.5.2.6 Self-Regulated Learning Constructs – Preparation Process 137
4.6 Discussion 141
4.6.1 Time Management 141
4.6.2 Attention and Lack of Focus 144

4.6.3 Exam Preparation Tasks 147
4.6.2 Limitations 149
4.6.2.1 Limitations of Zero-shot Learning Natural Language Processing 150
4.6.2.2 Self-report Bias in Exam Wrappers 150
4.6.2.3 Generalizability 151
4.6.2.4 Interpretational Limitations 152
4.6.3 Study Implications 152
4.6.3.1 Implications for Instructors and Students 153
4.6.4 Future Work 155
4.7 Conclusion 156
References 157
Chapter 5: Discussion and Conclusions 167
5.1 Introduction 167
5.2 Implications 168
5.2.1 Implications for Practice 169
5.2.2 Implications for Research 172
5.3 Future Work 173
5.4 Concluding Remarks 176
References 177
Appendices 179
Appendix A Full Exam Wrapper Questions 179
Appendix B Full Codebook for Manual Qualitative Analysis, NLPCA, and NLPGPT 180
Appendix C Full Distribution of topics 213

List of Tables

Table 1.1: Examples of SRL in each phase of Zimmerman's model.…………………………….4
Table 1.1: Research Plan and Outcomes for Dissertation……………………………………..13
Table 2.1: Exam wrapper questions and number of responses...……………………………...32
Table 2.2: Match, near-match, and unique codes for manual analysis and NLPCA for the Preparation Process question …..……………………………………………………………40
Table 2.3: Examples of match codes between manual analysis and NLPCA for the Preparation Process question …………………………………………………...…………………40
Table 2.4: Example of near-match codes found in NLPCA and their comparison to a similar code in manual analysis for the Preparation Process question……………………………….....…41
Table 2.5: Examples of unique codes found in NLPCA and manual analysis for the Preparation Process question………………………………………………………………………………42
Table 2.6: Match, near-match, and unique codes for manual qualitative analysis and NLPCA for the Exam Reflection question…………….…………………………………….………..….42
Table 2.7: Examples of match codes between manual analysis and NLPCA for the Exam Reflection question……………………..……………………………….………………..43
Table 2.8: Example of near-match codes found in NLPCA and their comparison to a similar code in manual analysis for the Exam Reflection question...……………………………………….43
Table 2.9: Examples of unique codes found in NLPCA and manual analysis for the Exam Reflection question…………………………………………….……………………….....44
Table 2.10: Match, near-match, and unique codes for manual analysis and NLPGPT for the Preparation Process question……………………………………………………………….45
Table 2.11: Examples of match codes between manual analysis and NLPGPT for the Preparation Process question………………………………………………………………………………45
Table 2.12: Examples of unique codes found in manual analysis for the Preparation Process question…………………………………………………………………………………..46
Table 2.13: Match, near-match, and unique codes for manual analysis and NLPGPT for the Exam Reflection question………………………………………………………………..47
Table 2.14: Examples of match codes between manual analysis and NLPGPT for the Exam Reflection question…………………………………………………………………………47
Table 2.15: Examples of unique codes between manual analysis and NLPGPT for the Exam Reflection question…………………………………………………………………………48
Table 3.1. Exam wrapper questions and different SRL constructs associated with those questions ……………………………………………………………………………………………67
Table 3.2. Questions analyzed for this study……………………………..……………………69
Table 3.3: Evaluation results table for zero-shot classifier of "Other skills"………..…………73
Table 3.4: Examples of correct, incorrect, and semantically related sentences for "Other skills" ……………………………………………………………………………………………73
Table 3.5. Evaluation results table for zero-shot classifier of "Learning process"………………73
Table 3.6. Examples of correct, incorrect, and semantically related sentences for "Learning process"……………………………………………………………………………………..74
Table 3.7. Evaluation results table for zero-shot classifier of "Start action"……………………74
Table 3.8. Examples of correct, incorrect, and semantically related sentences for "Start action" ……………………………………………………………………………………………75
Table 4.1: Summary of Zimmerman's cyclical phases model with definitions of key subprocesses (Zimmerman & Moylan, 2009)……………………………………………………97

Table 4.2: Exam wrapper questions for end-of-module exam wrappers and related SRL constructs……………………………………………………………………………………..106
Table 4.3: Questions analyzed for this study……………………………….……………………...109
Table 4.4: Codebook and SRL constructs for "Non-Physics Skills"…………………………...112
Table 4.5: Codebook and SRL constructs for "Preparation Process"…………………………..113
Table 4.6: Student Profile and Exam Results……………..……………………………………..118
Table 4.7: Average Number of Strategies Used by Each Profile Per Exam for "Non-Physics Skills"…………………………………………………………………………………………...119
Table 4.8: Average Number of Strategies Used by Each Profile Per Exam for "Preparation Process"………………………………………………………………………………………119

List of Figures

Figure 1.1: An Overview of the Problems Being Addressed and the Research Focus……..…...11
Figure 2.1: Natural Language Processing Cluster-Assisted Workflow……………………….….34
Figure 2.2: Overview of results and comparison of manual analysis to NLPCA and NLPGPT...39
Figure 4.1: Phases and processes of self-regulation (Zimmerman & Moylan, 2009)………….…..91
Figure 4.2: UMAP and HDBSCAN results indicating the four performance profiles based on exam performance across three exams in the semester……….……………………………...117
Figure 4.3: Cluster Analysis of Student Performance Based on Exam Grades…….…………..117
Figure 4.4: Distribution of topics that students of different performance profiles discuss in response to the "Non-Physics Skills" exam wrapper question …….……………….…….…121
Figure 4.5: Percentage of students of different performance profiles discuss in response to the "Non-Physics Skills" exam wrapper question broken down by exam …………………………………………………………………………………………..123
Figure 4.6: Distribution of SRL constructs from Zimmerman's model that students discuss in response to the "Non-Physics Skills" exam wrapper question………………….……………128
Figure 4.7: Distribution of SRL constructs from Zimmerman's model that students of different exam performance profiles discuss in response to the "Non-Physics Skills" exam wrapper question…………………………………………………...…………………………..129
Figure 4.8: Distribution of SRL constructs from Zimmerman's model that students of different exam performance profiles discuss in response to the "Non-Physics Skills" exam wrapper question broken down by exam…………………………………….….…….………….132
Figure 4.9: Distribution topics that students of different performance profiles discuss in response to the "Preparation Process" exam wrapper question……….…………………..………..…..134
Figure 4.10: Distribution of topics that students of different performance profiles discuss in response to the "Preparation Process" exam wrapper question broken down by exam …………………………………………………………………………………………..137
Figure 4.11: Distribution of SRL constructs from Zimmerman's model that students discuss in response to the "Non-Physics Skills" exam wrapper question……………………..…………139
Figure 4.12: Distribution of SRL constructs from Zimmerman's model that students of different exam performance profiles discuss in response to the "Preparation Process" exam wrapper question……………………………………………………………………………………140
Figure 4.13: Distribution of SRL constructs from Zimmerman's model that students of different exam performance profiles discuss in response to the "Preparation Process" exam wrapper question broken down by exam……………………………….……..…………………..141
Figure 5.1: Dissertation research focus with an overview of problems being addressed and research outcomes……………………………………………………………………….168

Common Acronyms in Dissertation

AI – Artificial Intelligence

BERT – Bidirectional Encoder Representations from Transformers

GPT – Generative Pretrained Transformers

HDBSCAN – Hierarchical Density-Based Spatial Clustering

LDA – Latent Dirichlet Allocation

LLMs – Large Language Models

NLP – Natural Language Processing

NLPCA – Natural Language Processing Cluster-Assisted

NLPGPT – Natural Language Processing using Generative Pretrained Transformers

PCA – Principal Component Analysis

SRL – Self-Regulated Learning

UMAP – Uniform Manifold Approximation and Projection

Chapter 1 – Introduction

1.1 Background and Context

Distinguishing between effective and ineffective study strategies is an important ability for students moving from secondary to tertiary education (Blackmore et al., 2021). Often this transition is accompanied by academic difficulty and an increased possibility of failure. At universities, students might face large classrooms, impersonal lectures, countless tests, and deadlines, with little access to teachers and tutors for personal guidance and feedback (Vosniadou, 2020). Findings by Shell et al. (2013) suggested that lack of regulation, which describes confused students, who have difficulty studying effectively and also need support from others, was associated with lower grades and knowledge retention. Additionally, better self-regulation has a positive association with academic achievement (Bergin et al., 2005; Nota et al., 2004; Paechter et al., 2010; Tilfarlioglu & Delbesoglugil, 2014). It is therefore important for undergraduate engineering students to establish a repertoire of effective learning strategies early in their academic career and to regulate their learning since better learning strategies lead to improved academic success.

One opportunity for improving learning strategies is when students receive a graded exam. Graded exams can be a valuable - often unrecognized - opportunity for students to discover how their study strategies and learning activities affected their grades by allowing them to reflect on their performance and how it relates to how they prepared for the exam (Lovett, 2013). Examples of poorly regulated study strategies include not being able to understand how to study or what they need to do when stuck or confused about their work while they are studying (Shell et al., 2013). Students can also use their graded exams to reflect on their learning activities in preparation for exams, which would also be an instance of self-regulation (Castellanos &

Enszer, 2013). While actively involving students in their learning process by having them reflect on their learning experiences can enhance academic success (Tuckman & Kennedy, 2011), this can be challenging in large foundational engineering classrooms (Mervis, 2013).

Large class sizes are often associated with a lower frequency of interaction between instructors and students (Grohs et al., 2018; Soledad et al., 2017). This results in instructors having less awareness of students' needs and strengths (Panadero et al., 2016). To address the problem of understanding students' learning strategies in large foundational engineering classrooms, I posit that a better understanding of students' self-regulated learning (SRL) while preparing for exams could pave the way to assisting students to improve their academic performance in foundational engineering courses. Additionally, exploring SRL using novel natural language processing (NLP) techniques could assist instructors and researchers in better understanding SRL and save time in analysis. This understanding could then be used to tailor tasks for students' learning (Panadero et al., 2016).

In the upcoming subsections of this chapter I shall be reviewing SRL, the use of a structured self-reflection tool called an exam wrapper to improve student's SRL strategies in preparation for exams, and the use of NLP as a tool for analyzing student exam wrappers to better understand SRL strategies while preparing for exams.

1.1.1 Self-regulated learning

Vosniadou (2020) proposed that one of the various ways the transition from secondary school to higher education can be facilitated is by improving students' SRL. A self-regulated learner explores tasks with confidence, diligence, and resourcefulness. Self-regulated learners are aware when they know a fact or possess a skill and when they do not (Zimmerman & Schunk, 1989). Although SRL has been studied by many scholars (Boekaerts, 1999; Butler & Winne, 1995; Pintrich, 2000; Schunk, 2008; Zimmerman, 1986) and there have been various definitions, I shall use a definition by one of the more prominent authors in the field, Barry J. Zimmerman, who defined SRL as: *"self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals."* (Zimmerman, 2000, p. 14). In addition to the many definitions, there exist many models for SRL (Panadero, 2017), but the most widely cited and used in the field of engineering education (Harding, 2018; Prather et al., 2020; Sáez-Delgado et al., 2020; Silverajah et al., 2022) is Zimmerman's Cyclical Phases Model (Zimmerman & Moylan, 2009). The reason for using Zimmerman's definition of SRL, as well as his model is that it is cyclical in nature, with three distinct phases. While other models such as Pintrich (2000) and Winne & Hadwin (1998) can be seen as cyclical in nature, they do not provide a clear distinction between the different phases and see SRL as a more "open" process (Efklides, 2011). Furthermore, the exam wrapper questions that I investigated in this dissertation were informed by Zimmerman's model.

1.1.2 Exam Wrappers

One of the ways that SRL can be improved is by using self-reflections (Sitzmann & Ely, 2010; van den Boom et al., 2007; Zumbrunn et al., 2011). Reflection and SRL have deeply intertwined histories in the discourse of teaching and learning since self-reflection forms one of the three subprocesses of SRL. The concept of self-reflection refers to either qualitative or quantitative in-depth self-monitoring of the learning process, learning outcomes, and the causes of one's errors or successes (van Loon, 2018). Through the process of self-reflection, students can acquire insights into the learning standards and competencies they are expected to develop, compare these to their progress, and learn how to act to close the gap between the two (Sadler, 1989; Sitzmann & Ely, 2010).

One of the reflection activities that incorporate SRL is an exam wrapper. Exam wrappers are structured reflection activities that prompt students to practice SRL strategies after they get their graded exams back (Lovett, 2013). Self-regulated learning strategies can be regarded as the implementation of activities aimed at achieving learning goals (Sebesta & Bray Speth, 2017). Each phase of Zimmerman's model can be associated with certain SRL strategies; Table 1.1, adapted from Sebesta & Bray Speth (2017), provides examples of strategies in each phase.

Table 1.1: Examples of SRL in each phase of Zimmerman's model (Sebesta & Bray Speth, 2017)

Phase in Zimmerman's Model	Strategy examples
Forethought	• Make a timeline to parse out study tasks and materials • Aim to keep up with assigned work and reading
Performance	• Studying early or in advance for the exam • Structuring the study environment to learn more effectively
Self-reflection	• Check the progress of his/her work, or generally monitor understanding of the material • Address or clarify confusion or gaps in knowledge by reviewing graded work

Exam wrappers can be different depending on the context. While there are different questions that students can be asked about their exam preparation and performance, there are generally three kinds of questions that they ask students: (a) how they prepared for the exam, (b) what kinds of errors they made on the exam, and (c) what they might do differently to prepare for the next exam (Lovett, 2013). These three questions are related to each phase of Zimmerman's model depending on how the student answers. For example, asking students how they prepared for the exam could relate to strategies in the Performance Phase if the student speaks about tasks and actions they took to prepare for the exam, but it could also be related to the Forethought Phase if students refer to setting goals and discuss the plans they made to achieve those goals. The question about the kinds of errors made during the exam could be related to self-reflection since it requires the student to look at their exam performance and self-evaluate where their errors were. Finally, in the questions of what they could do differently, they could set goals and plan to do things differently which would be associated with the forethought phase in Zimmerman's Model of SRL.

Through these three types of questions, exam wrappers aim to improve students' preparation for subsequent exams by asking them to reflect on their learning strategies, compare those strategies to the learning outcomes, and adjust their learning strategies for the next exam (Craig et al., 2016). Furthermore, instructors can review the exam wrappers to see which SRL strategies have led to success in the exam and adapt their teaching to assist students with better learning strategies for the next exam and also support students' SRL development (Lovett, 2013). In this way, exams can be used not only as a summative assessment, which is supposed to be used once the learning has already taken place to audit a student's performance (Dixson & Worrell, 2016) but also as an intervention to improve students' SRL strategies (Panadero et al., 2016).

In small classrooms, the instructor will administer the exam wrappers, collect them, and then provide students with some feedback on their responses. Sometimes exam wrappers are graded as part of the course grade to encourage participation (Davis, 2021). This method becomes challenging for instructors when class sizes are 100 students or more since the grading of exam wrappers becomes overwhelming for a single instructor (Carpenter et al., 2020). The challenge for the instructor is not only the time that it would take to grade exam wrappers for a large classroom but also to make meaningful connections to what SRL strategies students have used and how effective those strategies have been.

1.1.3 Natural Language Processing

A report by the National Center for Education Statistics (2022) indicated that total undergraduate enrollment is projected to increase from 15.9 million to 17.1 million students between 2020 and 2030 in the U.S. With this projected increase in undergraduate enrolment, and financial troubles within universities, such as budget cuts by states, large class sizes are an attractive option, however, Kokkelenberg et al. (2008) found that larger classes negatively affect students' grades. One possible reason for this is that larger classes can lead to an increase in student text-based data from homework, assignments, or open-ended responses on teaching evaluations which presents a challenge for educators who need to grade these assignments and gain insights into which areas students are doing well in and which areas need improvement. In the case of exam wrappers, Carpenter et al. (2020) addressed the challenge of providing students with large classroom feedback by using multiple-choice questions with pre-defined feedback which was released to the students based on their choices. While the authors found that there was a relationship between the students' use of the wrappers and their performance in the course, they noted that their metacognitive skills, which are a necessary part of SRL since they relate to knowledge and regulation of one's cognition (Dinsmore et al., 2008; Flavell, 1979) did not necessarily improve. Furthermore, the multiple-choice-style exam wrapper does not provoke as deep of a reflection as when using open-ended questions (Panadero et al., 2016).

The emergence of NLP has opened the door for new ways to think about processing large corpora of textual data. Natural language processing is a collection of approaches for analyzing natural language. There are various applications for NLP such as translating text from one language to another, text summarization, parts of speech tagging (POS) - which determines the part of speech of each part in a sentence, and co-reference resolution - which refers to a sentence or larger set of text that determines all words which refer to the same object (Khurana et al., 2022). Researchers and practitioners in engineering education have already looked at applications of some of these NLP tools for automatic short answer grading (ASAG), automatic essay scoring (AES; Haller et al., 2022), and automatic question generation (AQG; Tsai et al., 2021). Additional applications in engineering education research other than for assessment purposes have employed NLP techniques to summarize large corpora of data in text summarization (Katz et al., 2021), detect students' sentiments about a course (Ganesh et al., 2022), analyze engineering students' use of disciplinary discourse in their resumes (Berdanier et al., 2018), assess students' metacognitive development in the classroom (Bhaduri, 2018), summarizing students' reflections on confusing concepts using a mobile application (Menekse, 2020), and qualitative research (Katz et al., 2023).

Some of the work on NLP in engineering education that I have cited has used techniques that rely on monogram-based approaches that do not deal well with semantic meaning in text, therefore yielding limited insights into student textual data. For example, older NLP techniques could identify a common word such as "engineering" in a document, but this does not give any insights as to what specifically about engineering the document is referring to. Recent NLP techniques, using a transformer-based approach (Vaswani et al., 2017), have shown more sensitivity to detecting the semantic meaning of text when compared with monogram-based approaches (Becker et al., 2021; Ganesh et al., 2022; Wulff et al., 2022). Furthermore, the advent of large language models (LLMs) and generative AI, such as OpenAI's GPT-3.5 and GPT-4, have extended the capabilities and flexibility of performing NLP tasks more generally (Katz et al., 2023). My work aims to build on the use of these transformer-based models and LLMs to identify the different strategies and SRL constructs that students discuss in their exam wrappers and whether there is a difference in strategies and SRL constructs discussed by students of different exam performance profiles. This study provides an understanding of how we could use transformer-based NLP in research and practice, while also providing an indicator of effective learning strategies and hindrances to SRL that students describe in their exam wrappers.

1.2 Statement of the Problem

Despite the importance of SRL in student learning and academic achievement, I have identified that more research is needed in studying SRL strategies as they are discussed in student responses to exam wrappers. If we want to better support student SRL strategies in engineering because of how it has been linked to academic performance (Lawanto et al., 2014; Menekse, 2020), we can do so by understanding if and how students are engaging in SRL while preparing for exams (Grohs et al., 2018). If we better understand what SRL strategies students are using while they are in this process, it could lead to the development of specific pedagogical interventions that could develop students' SRL. These SRL skills could help them to better prepare for exams and also for better academic performance in other assignments (Azevedo & Cromley, 2004; Broadbent & Poon, 2015; Sebesta & Bray Speth, 2017).

Chew et al. (2016) identified learning strategies that students use in exam wrappers such as: being more deliberate in presenting a solution, reviewing past homework/solutions, starting preparation early enough, and studying with peers. While this information is useful, we do not know the distribution of these strategies. Knowing the distribution of strategies and SRL constructs using exam wrappers can assist instructors in identifying the most common strategies and SRL constructs that students are using while preparing for exams. Furthermore, the effectiveness of these strategies can be identified by looking at which strategies students of different exam performance profiles use.

Our ability to understand which SRL strategies engineering students are using through analyzing exam wrappers is further hampered by the current methodological approaches for studying SRL when looking at large corpora of textual data. Traditionally, qualitative reflections are used on small sample sizes because of the time taken to analyze textual data (Crowston et al., 2012) or larger sample sizes that use shorter textual responses. Since students' exam wrappers are challenging to analyze in large classrooms, tools and techniques for analyzing large corpora of textual data may help. Being able to analyze these SRL strategies on a larger scale allows us to see the distribution of these strategies to know what the most common strategies are that students use, and which strategies are less common. Furthermore, linking these strategies to students' grades in exams assisted me in identifying which are the most common strategies and SRL constructs that students from differing exam performance profiles use. As stated previously, having insights into the distribution of what strategies and SRL constructs students are engaging in is more insightful than just knowing what those strategies are. For example, knowing that a specific ineffective strategy is very common among students preparing for exams instructors can discourage students from using that strategy and instructors can use these insights to help students build more effective learning strategies. To further drive this point home if we observe that an effective strategy is not very common, then pedagogical changes can be made to address that issue by explicitly teaching that strong strategy within context. These insights could be valuable given the importance of specific teaching and learning strategies for specific contexts and students (Bager-Elsborg, 2017; Garcia & Pintrich, 1996).

1.3 Purpose of this Dissertation

The overarching purpose of this dissertation involves exploring the use of a novel NLP workflow for analyzing student strategies and SRL constructs in response to exam wrappers. I achieved this through three manuscripts. Manuscript 1 develops a method of studying strategies and SRL in exam wrapper responses. I have done this by comparing three methods of analyzing qualitative data: (i) manual qualitative coding, (ii) a state-of-the-art NLP workflow that uses transformers plus manual labeling, and (iii) a GPT-3.5-assisted approach. To analyze the potential for this NLP workflow to be used in analyzing self-reflections, the three methods were compared in terms of agreement. In the case of Manuscript 1, agreement refers to the degree of similarity between codes across different methods. Manuscript 2 uses the codebook developed in Manuscript 1 and applies a novel NLP technique called zero-shot learning (ZSL; Yin et al., 2019) where this ZSL approach will classify student responses into the different codes that were developed in Manuscript 1. I then evaluate the accuracy of this method in classifying student learning strategies in exam wrappers. Finally, Manuscript 3 explores and identifies strategies and SRL constructs that students describe in their exam wrapper responses. I use the results from Manuscript 2 to identify what learning strategies students are using while preparing for exams. I also link these learning strategies to SRL constructs using Zimmerman's model as a theoretical framework. Additionally, I use students' exam performances across three exams to examine the learning strategies and SRL constructs that students of different exam performance profiles use to see if any trends emerge in different profiles.

While significant research has suggested that there is an association between SRL and academic achievement, I have identified that a gap exists in using a novel NLP workflow to find the distribution of learning strategies and SRL constructs that students use on a larger scale than previously studied. Furthermore, by incorporating exam grades into these distributions, it becomes possible to assess the variations in strategies employed by students across different performance profiles. This could allow instructors and curriculum designers to adjust how they teach and design their courses based on the most common and effective SRL strategies. Additionally, the implementation of this NLP workflow could save instructors large amounts of time that they would usually take to manually read through individual responses to exam wrappers.

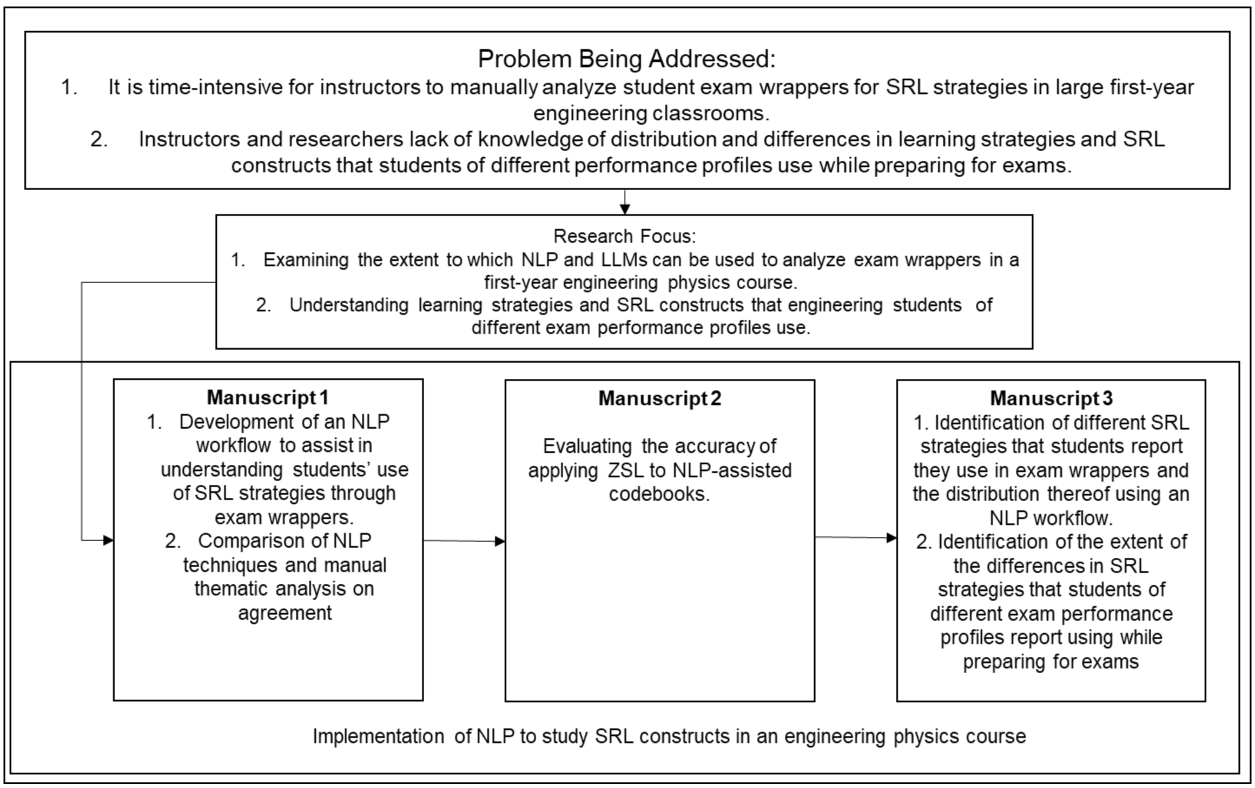


Figure 3.1: An overview of the Problems Being Addressed and the Research Focus

Figure 1.1 shows the research focus of this dissertation and what part each manuscript plays in addressing the problem. The problem being addressed is the time it takes instructors to (i) analyze exam wrappers in large first-year engineering classes and (ii) identify effective SRL strategies that students use while preparing for exams. This study, therefore, explores the use of a novel NLP workflow for analyzing student strategies and SRL constructs in responses to exam wrappers. The theoretical framework used to guide this dissertation is Zimmerman's three-phase cyclical model of SRL (Zimmerman & Moylan, 2009).

1.4 Research Questions

To accomplish the overarching purpose of exploring the use of a novel NLP workflow for analyzing student SRL strategies in responses to exam wrappers, I developed one overarching research question along with three sub-research questions which formed the research questions of each manuscript. Table 1.2 displays each of the three manuscripts and their research questions, how and what data was collected and analyzed, and the outcomes of each manuscript.

Table 1.1: Research Plan and Outcomes for Dissertation

Overarching Research Question: **How can we use a transformer-based NLP workflow to understand students' SRL strategies in exam wrappers?**				
Manuscript	**Research Question**	**Data Collection**	**Analysis**	**Outcomes**
Manuscript 1: Advancing Qualitative Analysis: An Exploration of the Potentials of Generative AI and NLP in Qualitative Coding	How does an NLP-assisted approach compare to traditional qualitative analysis of first-year engineering physics exam wrappers?	Student responses to exam wrappers on the end-of-module exams.	Manual qualitative analysis, a transformer-based NLP workflow, and GPT-3.5. Comparing codes that each method produces for similarity.	1. Development of an NLP-assisted codebook that can be used to assess students' use of SRL in end-of-module exam wrappers. 2. The level of agreement between two different transformer-based NLP codebooks and a codebook developed through manual qualitative analysis.
Manuscript 2: Utilizing Natural Language Processing to Examine Self-Reflections in Self-Regulated Learning	How does an NLP-assisted approach compare to a traditional qualitative analysis of first-year engineering physics exam wrappers?		Accuracy checks of implementing Transformer-based NLP workflow using zero-shot learning by reading through a subset of the data and ensuring that the researcher agrees with the codes assigned to the student response-text	An evaluation of the NLP workflow in whether the method can be used to classify students' exam preparation strategies and SRL constructs.
Manuscript 3: Understanding Performance Profiles through Self-Regulated Learning Constructs in Engineering Physics: A Large Language Model Approach	To what extent are there differences in the strategies and SRL constructs that students of different performance profiles report that they use in their exam wrapper responses?	Student responses to exam wrappers on the end-of-module exams and exam grades.	Transformer-based NLP workflow using zero-shot learning.	1. Identification of different SRL strategies that students report they use in exam wrappers and the distribution thereof. 2. Identification of differences in strategies and SRL constructs between students of different performance profiles in exam wrapper responses.

1.5 Significance of the Research

This research contributes toward helping engineering education researchers and instructors recognize additional ways to analyze qualitative data on a larger scale by using NLP. Not only can this research assist researchers and practitioners with the utility of NLP, but also the use of advanced techniques such as the transformer-based approach which can yield better insights because of its ability to process more complex text scales. In Manuscript 1 I demonstrate multiple NLP-assisted approaches to codebook development that can save time and enhance insights in qualitative analysis. Manuscript 2 provides insights into the utility of using ZSL, by manually checking how it classified a subset of the exam wrapper responses and if it was appropriate to apply ZSL to the full dataset.

Using the outcomes from Manuscripts 1 and 2, Manuscript 3 provides instructors and researchers with insights on strategies and SRL in engineering education by identifying strategies and SRL constructs discussed by students in their exam wrappers as well as the distribution of these strategies. Insights into which are the most commonly discussed strategies and SRL constructs can inform instructors on ways to teach differently to employ the more effective SRL-enhancing strategies for students in first-year engineering courses. Common pitfalls in students' ability to perform well in exams were also identified and can be dealt with in terms of encouraging students to use better strategies while preparing for exams. Manuscript 3 demonstrates the potential of using NLP to analyze large corpora of data instead of selecting a subset of the data to analyze by hand and also provides a potential avenue to save instructors time in manually analyzing students' qualitative reflections. Manuscript 3 also determines whether there are any associations between students' SRL and how they perform on exams. This information can inform instructors about whether different performing students need different types of pedagogy or interventions to improve exam success.

1.6 Summary

The purpose of this dissertation is to better understand the use of SRL strategies that students report on in their end-of-module exam wrappers using a transformer-based NLP workflow. I investigated the utility of using state-of-the-art NLP techniques by comparing them to manual qualitative analysis. I then evaluated a technique for qualitative analysis to check if it would yield accurate results for classifying students' strategies and SRL constructs. Using this NLP workflow, I addressed a research gap in SRL in engineering education, by identifying the extent of the differences in learning strategies and SRL constructs that students use of different performance profiles report in end-of-module exam wrappers and the distribution of those strategies across a large sample of first-year engineering students

This document is divided into five chapters. Chapters 2, 3, and 4 comprise Manuscripts 1, 2, and 3 respectively. The manuscripts provide greater details in the background and context of the studies, a review of the relevant literature to the study, the methods, results, discussion, and conclusion of each manuscript. Finally, Chapter 5 will discuss how the individual manuscripts answer their respective research questions, as well as how they contribute to the overarching research question. Chapter 5 also includes a discussion of implications, areas of future work, and concluding remarks.

References

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology*, *96*(3), 523.

Bager-Elsborg, A. (2017). Discipline context shapes meaningful teaching: A case study of academic law. *Journal of Further and Higher Education*, *43*, 1–13. https://doi.org/10.1080/0309877X.2017.1377162

Becker, J. P., Kahanda, I., & Kazi, N. H. (2021, July 26). *WIP: Detection of Student Misconceptions of Electrical Circuit Concepts in a Short Answer Question Using NLP*. 2021 ASEE Virtual Annual Conference Content Access. https://peer.asee.org/wip-detection-of-student-misconceptions-of-electrical-circuit-concepts-in-a-short-answer-question-using-nlp

Berdanier, C. G. P., Baker, E., Wang, W., & McComb, C. (2018). Opportunities for Natural Language Processing in Qualitative Engineering Education Research: Two Examples. *2018 IEEE Frontiers in Education Conference (FIE)*, 1–6. https://doi.org/10.1109/FIE.2018.8658747

Bergin, S., Reilly, R., & Traynor, D. (2005). Examining the role of self-regulated learning on introductory programming performance. *Proceedings of the First International Workshop on Computing Education Research*, 81–86. https://doi.org/10.1145/1089786.1089794

Bhaduri, S. (2018). *NLP in Engineering Education-Demonstrating the use of Natural Language Processing Techniques for Use in Engineering Education Classrooms and Research* [PhD Thesis]. Virginia Tech.

Blackmore, C., Vitali, J., Ainscough, L., Langfield, T., & Colthorpe, K. (2021). A Review of Self-Regulated Learning and Self-Efficacy: The Key to Tertiary Transition in Science, Technology, Engineering and Mathematics (STEM). *International Journal of Higher Education*, *10*(3), 169–177.

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, *31*(6), 445–457. https://doi.org/10.1016/S0883-0355(99)00014-2

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, *27*, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, *65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Carpenter, T. S., Beall, L. C., & Hodges, L. C. (2020). *Using the LMS for Exam Wrapper Feedback to Prompt Metacognitive Awareness in Large Courses*. https://doi.org/10.14434/jotlt.v9i1.29156

Castellanos, M., & Enszer, J. A. (2013). *Promoting Metacognition through Reflection Exercises in a Thermodynamics Course*. 23.999.1-23.999.13. https://peer.asee.org/promoting-metacognition-through-reflection-exercises-in-a-thermodynamics-course

Chew, K. J., Chen, H. L., Rieken, B., Turpin, A., & Sheppard, S. (2016). Improving students' learning in statics skills: Using homework and exam wrappers to strengthen self-regulated learning. *123rd ASEE Annual Conference and Exposition, June 26, 2016 - June 29, 2016*, *2016-June*.

Craig, M., Horton, D., Zingaro, D., & Heap, D. (2016). Introducing and evaluating exam wrappers in CS2. *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, 285–290.

Crowston, K., Allen, E. E., & Heckman, R. (2012). Using natural language processing technology for qualitative data analysis. *International Journal of Social Research Methodology*, *15*(6), 523–543.

Davis, K. C. (2021, July 26). *Examining the Efficacy of Exam Wrappers in a Computer Science Course*. 2021 ASEE Virtual Annual Conference Content Access. https://peer.asee.org/examining-the-efficacy-of-exam-wrappers-in-a-computer-science-course

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the Conceptual Lens on Metacognition, Self-regulation, and Self-regulated Learning. *Educational Psychology Review*, *20*(4), 391–409. https://doi.org/10.1007/s10648-008-9083-6

Dixson, D. D., & Worrell, F. C. (2016). Formative and Summative Assessment in the Classroom. *Theory Into Practice*, *55*(2), 153–159. https://doi.org/10.1080/00405841.2016.1148989

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational Psychologist*, *46*(1), 6–25.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Ganesh, A., Scribner, H., Singh, J., Goodman, K., Hertzberg, J., & Kann, K. (2022). Response Construct Tagging: NLP-Aided Assessment for Engineering Education. *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*, 250–261. https://aclanthology.org/2022.bea-1.29

Garcia, T., & Pintrich, P. R. (1996). Assessing Students' Motivation and Learning Strategies in the Classroom Context: The Motivated Strategies for Learning Questionnaire. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge* (pp. 319–339). Springer Netherlands. https://doi.org/10.1007/978-94-011-0657-3_12

Grohs, J. R., Knight, D. B., Young, G. D., & Soledad, M. M. (2018). Exploring Academic Performance Paths and Student Learning Strategies in a Large Foundational Engineering Course. *International Journal of Education in Mathematics, Science, and Technology*, 241–253. https://doi.org/10.18404/ijemst.428175

Haller, S., Aldea, A., Seifert, C., & Strisciuglio, N. (2022). *Survey on Automated Short Answer Grading with Deep Learning: From Word Embeddings to Transformers*.

Harding, S.-M. (2018). *Self-regulated learning in the classroom*.

Katz, A., Norris, M., Alsharif, A. M., Klopfer, M. D., Knight, D. B., & Grohs, J. R. (2021, July 26). *Using Natural Language Processing to Facilitate Student Feedback Analysis*. 2021 ASEE Virtual Annual Conference Content Access. https://peer.asee.org/using-natural-language-processing-to-facilitate-student-feedback-analysis

Katz, A., Shakir, U., & Chambers, B. (2023). *The Utility of Large Language Models and Generative AI for Education Research* (arXiv:2305.18125). arXiv. https://doi.org/10.48550/arXiv.2305.18125

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022). Natural language processing: State of the art, current trends, and challenges. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-022-13428-4

Kokkelenberg, E. C., Dillon, M., & Christy, S. M. (2008). The effects of class size on student grades at a public university. *Economics of Education Review*, *27*(2), 221–233. https://doi.org/10.1016/j.econedurev.2006.09.011

Lawanto, O., Santoso, H., Lawanto, K., & Goodridge, W. (2014). Self-Regulated Learning Skills and Online Activities Between Higher and Lower Performers on a Web Intensive Undergraduate Engineering Course. *The Journal of Educators Online*, *11*(3). https://doi.org/10.9743/JEO.2014.3.2

Lovett, M. C. (2013). Make exams worth more than the grade: Using exam wrappers to promote metacognition. In: Kaplan, M., Silver, N., LaVague-Manty, D., Meizlish, D. (eds) Using reflection and metacognition to improve student learning: Across the disciplines, across the academy. In *Using reflection and metacognition to improve student learning: Across the disciplines, across the academy,* (pp. 18–52).

Menekse, M. (2020). The Reflection-Informed Learning and Instruction to Improve Students' Academic Success in Undergraduate Classrooms. *Journal of Experimental Education*, *88*(2), 183–199. https://doi.org/10.1080/00220973.2019.1620159

Mervis, J. (2013). *Transformation is possible if a university really cares*. American Association for the Advancement of Science.

National Center for Education Statistics. (2022). *Undergraduate Enrollment*. U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/cha.

Nota, L., Soresi, S., & Zimmerman, B. J. (2004). Self-regulation and academic achievement and resilience: A longitudinal study. *International Journal of Educational Research*, *41*(3), 198–215. https://doi.org/10.1016/j.ijer.2005.07.001

Paechter, M., Maier, B., & Macher, D. (2010). Students' expectations of, and experiences in e-learning: Their relation to learning achievements and course satisfaction. *Computers & Education*, *54*(1), 222–229. https://doi.org/10.1016/j.compedu.2009.08.005

Panadero, E. (2017). A Review of Self-regulated Learning: Six Models and Four Directions for Research. *Frontiers in Psychology*, *8*, 422. https://doi.org/10.3389/fpsyg.2017.00422

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: When measurement and intervention come hand in hand. *Scandinavian*

Journal of Educational Research, *60*(6), 723–735. https://doi.org/10.1080/00313831.2015.1066436

Pintrich, P. R. (2000). The Role of Goal Orientation in Self-Regulated Learning. In *Handbook of Self-Regulation* (pp. 451–502). Elsevier. https://doi.org/10.1016/B978-012109890-2/50043-3

Prather, J., Becker, B. A., Craig, M., Denny, P., Loksa, D., & Margulieux, L. (2020). What do we think we think we are doing? Metacognition and self-regulation in programming. *Proceedings of the 2020 ACM Conference on International Computing Education Research*, 2–13.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/BF00117714

Sáez-Delgado, F., Mella-Norambuena, J., Loyer, S., Zambrano, & Zañartu-Canihuante, N. (2020). *Self-regulated learning in engineering students: A systematic review*. *41*, 7–22.

Schunk, D. H. (2008). Metacognition, Self-Regulation, and Self-Regulated Learning: Research Recommendations. *Educational Psychology Review*, *20*(4), 463–467. https://doi.org/10.1007/s10648-008-9086-3

Sebesta, A. J., & Bray Speth, E. (2017). How Should I Study for the Exam? Self-Regulated Learning Strategies and Achievement in Introductory Biology. *CBE Life Sciences Education*, *16*(2), ar30. https://doi.org/10.1187/cbe.16-09-0269

Shell, D. F., Hazley, M. P., Soh, L.-K., Ingraham, E., & Ramsay, S. (2013). Associations of students' creativity, motivation, and self-regulation with learning and achievement in college computer science courses. *2013 IEEE Frontiers in Education Conference (FIE)*, 1637–1643. https://doi.org/10.1109/FIE.2013.6685116

Silverajah, V. S. G., Wong, S. L., Govindaraj, A., Khambari, M. N. Md., Rahmat, R. W. B. O. K., & Deni, A. R. M. (2022). A Systematic Review of Self-Regulated Learning in Flipped Classrooms: Key Findings, Measurement Methods, and Potential Directions. *IEEE Access*, *10*, 20270–20294. https://doi.org/10.1109/ACCESS.2022.3143857

Sitzmann, T., & Ely, K. (2010). Sometimes you need a reminder: The effects of prompting self-regulation on regulatory processes, learning, and attrition. *Journal of Applied Psychology*, *95*(1), 132–144. https://doi.org/10.1037/a0018080

Soledad, M., Grohs, J., Bhaduri, S., Doggett, J., Williams, J., & Culver, S. (2017). Leveraging institutional data to understand student perceptions of teaching in large engineering classes. *Proc. Front. Educ. Conf. FIE*, *2017-October*, 1–8. Scopus. https://doi.org/10.1109/FIE.2017.8190608

Tilfarlioglu, F. Y., & Delbesoglugil, A. B. Ö. (2014). Questioning academic success through self-regulation, self-esteem and attitude in foreign language learning (a case study). *Theory and Practice in Language Studies*, *4*(11), 2223.

Tsai, D. C. L., Huang, A. Y. Q., Lu, O. H. T., & Yang, S. J. H. (2021). Automatic Question Generation for Repeated Testing to Improve Student Learning Outcome. *2021 International Conference on Advanced Learning Technologies (ICALT)*, 339–341. https://doi.org/10.1109/ICALT52272.2021.00108

Tuckman, B. W., & Kennedy, G. J. (2011). Teaching learning strategies to increase success of first-term college students. *Journal of Experimental Education*, *79*, 478–504. https://doi.org/10.1080/00220973.2010.512318

van den Boom, G., Paas, F., & van Merriënboer, J. J. G. (2007). Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learning and Instruction*, *17*(5), 532–548. https://doi.org/10.1016/j.learninstruc.2007.09.003

van Loon, M. (2018). *Self-Assessment and Self-Reflection to Measure and Improve Self-Regulated Learning in the Workplace*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Vosniadou, S. (2020). Bridging Secondary and Higher Education. The Importance of Self-regulated Learning. *European Review*, *28*(S1), S94–S103. https://doi.org/10.1017/S1062798720000939

Winne, P. H., & Hadwin, A. F. (1998). Studying as Self-Regulated Learning. In *Metacognition in Educational Theory and Practice*. Routledge.

Wulff, P., Mientus, L., Nowak, A., & Borowski, A. (2022). Utilizing a Pretrained Language Model (BERT) to Classify Preservice Physics Teachers' Written Reflections. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-022-00290-6

Yin, W., Hay, J., & Roth, D. (2019). *Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach* (arXiv:1909.00161). arXiv. https://doi.org/10.48550/arXiv.1909.00161

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology*, *11*(4), 307–313. https://doi.org/10.1016/0361-476X(86)90027-5

Zimmerman, B. J. (2000). Chapter 2 - Attaining Self-Regulation: A Social Cognitive Perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J., & Moylan, A. R. (2009). Self-Regulation: Where Metacognition and Motivation Intersect. In *Handbook of Metacognition in Education*. Routledge.

Zimmerman, B. J., & Schunk, D. H. (Eds.). (1989). *Self-Regulated Learning and Academic Achievement*. Springer New York. https://doi.org/10.1007/978-1-4612-3618-4

Zumbrunn, S., Tadlock, J., & Roberts, E. (2011). Encourage Self Regulated Learning in the Classroom. *MERC Publications*. https://scholarscompass.vcu.edu/merc_pubs/18

Chapter 2: Manuscript 1

Advancing Qualitative Analysis: An Exploration of the Potentials of Generative AI and NLP in Qualitative Coding

This manuscript includes intellectual contributions from Jennifer Case and Andrew Katz

2.1 Structured Abstract

Background

Traditional manual coding in qualitative data analysis can be labor-intensive and time-consuming, especially with large data sets. This research investigates the potential use of natural language processing (NLP) techniques and large language models (LLMs), such as GPT-3.5, to enhance efficiency and depth of insights during the qualitative data coding process.

Method

We compared traditional manual qualitative analysis with two NLP-assisted approaches, NLP Cluster Assisted (NLPCA) and NLP with GPT-3.5 (NLPGPT), using a dataset of 3,800 student responses on "exam wrappers" from an engineering physics course. Exam wrappers are structured reflection activities that prompt students to practice self-reflection after they get their graded exams back (Lovett, 2013). Agreement between the methods was evaluated based on the similarity of the generated codes.

Results

Both NLPCA and NLPGPT effectively identified similar codes in the student responses, demonstrating a promising alternative to traditional qualitative coding. Notably, the GPT-3.5 model exhibited strength in producing highly granular codes, which could offer deeper and more nuanced insights.

Discussion

The results of the study underscore the significant benefits of integrating NLP and LLMs into qualitative research. While the study identified challenges such as biases in language models, and resource constraints, the findings suggest these hurdles can be addressed with further research and refinement of the methodology. The application of NLP and LLMs across various research contexts needs validation, setting a promising direction for future studies. This research marks an important stepping stone in enhancing traditional qualitative research with AI technology, paving the way for more scalable, robust, and efficient research methodologies.

Keywords

Natural language processing, qualitative analysis, ChatGPT, Large Language Models

2.2 Introduction

Aspers & Corte (2019) characterize qualitative research as an iterative process that contributes to the scientific community's understanding by revealing significant distinctions through deep engagement with the phenomenon under study. A prevalent approach within qualitative research is the technique of qualitative coding. According to Saldaña (2014), a code in qualitative research often comprises a short phrase or word encapsulating an idea or a significant portion of language-based or visual data. This method of coding can be applied to a wide range of data types, including interview transcripts, participant observations, journal entries, literature, and illustrations.

Despite its wide application, the qualitative coding process is usually time-consuming and costly, and its findings are complex to replicate (Abram et al., 2020; Guetterman et al., 2018). Further, the volume of textual data being generated continues to grow, and novel techniques to analyze larger and more diverse types of data are steadily emerging (Abram et al., 2020). This study aims to further previous studies in exploring the comparison of long-standing techniques in qualitative analysis, with modern approaches to qualitative analysis. Specifically, we compare traditional qualitative analysis with modern NLP techniques and LLMs (Abram et al., 2020; Leeson et al., 2019; Xiao et al., 2023).

The advent of recent NLP technologies and LLMs such as ChatGPT presents a potential solution to the challenges of time and cost associated with analyzing extensive textual corpora. In this paper, we compare the results of traditional qualitative analysis with NLP-assisted techniques, particularly those using OpenAI's GPT-3.5 and open-source NLP tools. Our data set for conducting methodological evaluations comprises student reflections following exams in an engineering physics course.

2.3 Background

2.3.1 Qualitative analysis

The bedrock of qualitative research lies in the intricate process of coding, where researchers identify and label key ideas from qualitative data (Saldaña, 2014). These labels, known as codes or nodes, help researchers extract overarching themes and significant insights (Braun & Clarke, 2006). However, such in-depth analysis is labor-intensive and time-consuming, often demanding a team of analysts (Leeson et al., 2019). Analyzing a complete corpus of interviews can span over weeks. Moreover, disagreement among analysts must be reconciled, posing questions on the replicability of such studies (Armstrong et al., 1997). There is a concern that the results might be subject to disciplinary predilections or biases, which could potentially skew findings (Lincoln & Guba, 1985; Mackieson et al., 2019).

2.3.2 Natural Language Processing

Natural language processing and LLMs are being explored by the research community to counter some of these challenges. This set of methods leverages algorithmic and statistical techniques to decipher semantic meanings from textual data (Khurana et al., 2022). Unlike its early applications, which relied heavily on predetermined rule books (Crowston et al., 2012; Goodfellow et al., 2016) modern NLP embraces machine learning and deeper neural networks with more parameters to improve performance and adaptability (Mikolov et al., 2013). For example, the evolution of NLP took a significant leap forward with the introduction of the attention mechanism (Bahdanau et al., 2016). This innovative approach allowed models to focus on different parts of the input text sequence when producing output, mimicking human cognition during reading or listening. Attention mechanisms resolved a limitation of earlier sequence-to-sequence models which compressed all necessary information of a text into a fixed-length vector, struggling to handle longer sequences. This new development allowed models to weigh the importance of different words or phrases in a sentence, essentially enabling them to “pay attention” to context-relevant inputs.

Building on this, Vaswani et al. (2017) introduced the transformer model, a novel architecture that significantly expanded the possibilities for NLP. The transformer model was distinctive due to its self-attention mechanism, which gave the model the ability to weigh and relate different words in a sentence, regardless of their position, hence capturing the dependencies among words more effectively. This was a considerable advancement over recurrent neural networks (RNNs) and long short-term memory networks (LSTMs), which were limited by sequential computations (Hochreiter & Schmidhuber, 1997).

2.3.3 Large Language Models

OpenAI’s introduction of the Generative Pre-training Transformer (GPT; Radford et al., 2018) models and Google’s BERT (Devlin et al., 2019) marked a significant evolution in NLP, using vast amounts of text data and transformer capabilities to generate human-like text and predict subsequent words in sequences (Radford et al., 2018). These NLP models were given the name “Large Language Models” due to the vast amounts of data they have been trained on, the number of parameters the models have, and their ability to perform a wide range of NLP tasks, including qualitative research (Katz et al., 2023; Xiao et al., 2023). This technology was further refined in GPT-2, demonstrating the power of large-scale self-supervised learning through the generation of coherent, contextually relevant sentences (Radford et al., 2019). A significant leap forward came with ChatGPT, a model designed for generating conversational responses, fine-tuned on a dataset of internet conversations and reinforcement learning from human feedback. It indicated a promising potential for various applications, including virtual assistant technology and qualitative research analysis (Radford et al., 2021). The language model underlying ChatGPT, GPT-3.5 is what was used in part of this study.

2.3.4 NLP and Qualitative Analysis

Natural language processing presents several advantages over traditional methods. It is capable of rapidly processing a substantial volume of textual data and identifying intricate patterns that may elude human analysts. Past studies, such as those conducted by Crowston et al. (2012), have explored NLP's potential to automate qualitative data analysis using a dictionary-based approach. Other research has endeavored to summarize student responses (Luo et al., 2016), analyze disciplinary discourse in student resumes (Berdanier et al., 2018), and assess metacognitive development (Bhaduri, 2018; Cunningham et al., 2017).

The study by Leeson et al. (2019) bears a resemblance to the current study as it compared the traditional qualitative analysis with two other NLP approaches: Topic Modeling through Latent Dirichlet Allocation (LDA; Blei et al., 2003), and Word2Vec (Mikolov et al., 2013). This study demonstrated encouraging results for qualitative analysis utilizing NLP. Similarly, Abram et al. (2020) employed LDA with interview data for qualitative analysis, concluding that NLP holds considerable promise for qualitative data analysis. They suggested that researchers with an interest in NLP and basic programming skills could feasibly conduct similar studies. While these studies have shown promise in qualitative data analysis, they lack the nuance and accuracy that is achievable by more modern NLP technologies. We therefore sought to further the exploration of NLP in using transformer-based models which have been shown to achieve state-of-the-art results.

One of the main reasons why transformer-based models are state-of-the-art is their ability to process long-range dependencies, a common problem for earlier NLP methods. Long-range dependencies are a phenomenon where the understanding or interpretation of a word or a phrase in a sentence is influenced by another word or phrase that is located far away from it in the sentence (Lakretz et al., 2020). Transformer-based models have been shown to outperform older models because of their ability to overcome this challenge that older models could not address. For example, (Ganesh et al., 2022) successfully utilized RoBERTa, a pre-trained transformer model, to analyze student responses to their engineering experiences and their impact on their identities. The study used response construct tagging (RCT), an innovative classification task, which outperformed traditional Bag-of-Words models.

The RoBERTa-based approach provided nuanced categorization of students' responses, allowing educators a more refined assessment of curriculum effectiveness in shaping student perceptions and identities in engineering (M. Liu et al., 2019; Wolf et al., 2020). Becker et al. (2021) also found that the transformer model BERT performed well when identifying student misconceptions in a circuits course, outperforming rule-based approaches in precision. Additionally, Katz et al. (2021) emphasized the efficiency of a transformer-based approach, demonstrating its speed and accuracy in analyzing open-ended student feedback, and demonstrating the transformer-based approach's advantage over manual methods at scale. Wulff et al. (2023) explored NLP techniques using BERT to enhance writing analytics in science education and found utility in using transformer-based LLMs in assessing teachers' written reflections. A more recent study by Katz et al. (2023) demonstrated the utility of LLMs and generative AI in qualitatively analyzing unstructured text data of students' career interest essays using a combination of open-source transformer-based models from HuggingFace and GPT-3.5.

2.3.5 Significance of this Study

The proliferation of transformer-based LLMs and their increasing sophistication offers an exciting new avenue for conducting qualitative research. Comparing manual qualitative analysis of first-year engineering physics exam wrapper responses with the innovative NLP techniques of an open-source transformer-based language model and GPT-3.5 assisted approach, this study seeks to break new ground. Exam wrappers are structured reflection activities that prompt students to practice self-reflection after they get their graded exams back (Lovett, 2013). The potential for NLP to not only replicate but enhance and expedite the qualitative data analysis process is significant. By leveraging cutting-edge technology, we aim to streamline the analysis process, reduce labor-intensive coding, and uncover nuanced insights that may escape manual methods. While mindful of the potential limitations and challenges, this study may illuminate how the inherent capabilities of these advanced NLP techniques could overcome traditional bottlenecks in qualitative research, enhancing the reliability, efficiency, and depth of insights. We expect that this exploration will contribute to the ongoing dialogue on integrating AI technologies in qualitative research, forging a path toward a more robust, scalable, and insightful research landscape.

2.4 Methods

We explored the use of NLP and LLMs to facilitate qualitative analysis and compared these novel approaches with a traditional qualitative method that used the first part of a grounded theory approach called open coding (Case & Light, 2011). The process of open coding is the breaking down of the data into discrete parts and then analyzing the phenomena through a close examination of the data (Seale, 2004). The aim is to compare similarities and differences in the data that could lead to discoveries. We explored and compared three approaches - (1) manual qualitative analysis, (2) NLP cluster-assisted (NLPCA), and (3) NLP using generative pre-trained transformers (NLPGPT) - to analyze student exam wrappers in an engineering physics course. We sought to answer the research question, “How does an NLP-assisted approach compare to a traditional qualitative analysis of first-year engineering physics exam wrappers?” We sought to determine the agreement between the manual qualitative analysis and the two NLP-assisted approaches. For this study, agreement is defined as codes produced between methods being the same or similar enough as determined by the researcher. For example, if manual qualitative analysis returned the code “Better Conceptual Understanding” and either or both of the NLP approaches returned codes including “Increased Awareness of Conceptual Understanding”, “Understanding Concepts”, and/or “Improving Material Comprehension”, we judged these codes to agree.

2.4.1 Data Sources

The data we analyzed were 3,800 responses which comprised three exam wrapper responses from each student in Spring 2021, Fall 2021, and Spring 2022. These exam wrappers can be regarded as self-reflections that aid the implementation of activities aimed at achieving learning goals (Sebesta & Bray Speth, 2017). The exam wrapper responses come from Physics for Engineers II which is the second physics course that all incoming first-year students take in the College of Engineering at a large R1 university in the Southeastern region of the United States. The course is a calculus-based study of basic physics concepts that includes rotational dynamics, statics, oscillations, waves, fluids, heat and temperature, and the first and second laws of thermodynamics. The course is assessed mainly through exams, which count for 54% of the final grade. The course is delivered in-person and online through lectures three times per week, and labs that happen twice per week.

The exam wrappers included 9 questions that asked students to reflect on their performance on the exam, reflect on future strategies, and develop a strategic plan for preparing for the next exam. The student responses were written in paragraph form and stored in the university learning management system with a unique identifier for students. For this study, we analyzed two exam wrapper questions. The reason for this was to ensure that we could replicate this workflow for more than one question. The two questions analyzed were “Exam Reflection” and “Preparation Process” which can be found in Table 2.1. The full exam wrapper with all the questions can be found in Appendix Table A1. We also include the number of responses to the respective questions and the number of sentences we analyzed per question. The reason for the large difference in the number of responses and sentences between the two questions was that there would be no “Exam Reflection” question in the first exam wrapper because students would only have written one exam. The data was stored on Google Drive as comma-separated value files (CSVs) and were only accessible to members of the research team.

Table 2.1: Exam wrapper questions and number of responses

Name	Question	No. of Responses	No. of Sentences
Exam Reflection	What did you do differently between this exam and the previous exam? Did the changes that you made make an impact? Did you reach your goal from the last Exam Wrapper?	2,805	4,442
Preparation Process	Describe your process for preparing to take the module exam. Can you identify any areas of improvement that could strengthen your preparation activities?	3,667	6,578

2.4.2 Manual Analysis

We performed a manual qualitative analysis on the questions in Table 2.1 as the first method of analysis using the first step in the grounded theory methodology called open coding (Case & Light, 2011; Corbin & Strauss, 1994). This type of manual qualitative analysis was also used by Leeson et al. (2019) when comparing traditional qualitative analysis to NLP methods. Open coding was done by inductively developing codes though line-by-line open coding of responses. This allowed us to develop codes from the data rather than applying pre-existing concepts. A subset of the total collected set of responses - 270 responses - was manually coded for each of the questions presented in the data collection section. The number of responses was chosen as a starting point since most qualitative studies analyzing exam wrappers manually were in the range of 80-300 (Chew et al., 2016; Colthorpe et al., 2018; Craig et al., 2016; Davis, 2021). We decided to analyze the data to theoretical saturation so if more than 270 responses were needed we would analyze more (Guest et al., 2006). For sampling, we used stratified random sampling since we had data from three exam wrappers spanning three semesters (Cochran, 1977). We sampled 30 random responses to the two exam wrapper questions for each of the nine exam wrappers we had, totaling 270 exam wrapper responses. For the analysis, one of the researchers read the responses and then, using qualitative analysis, identified codes and assigned these to codes that formed a codebook - one for each question. New codes were created for responses that fall into categories that are not already covered in pre-existing codes. Responses with codes for each question shall be assigned multiple codes.

2.4.3 NLPCA

The NLPCA method we used is based on previous works using similar software (Katz et al., 2021). This approach is sometimes called a computer-assisted approach or a human-in-the-loop approach. The reason for this is that the process of analyzing qualitative data using this approach required human judgment so the process was not fully automated; instead, it aimed to reduce the number of human hours taken to analyze large corpora of qualitative data by first clustering responses into hypothetically similarly groups to which a researcher appends a label. In this section, I shall present an overview of how the NLPCA approach works. Figure 2.1 shows the pre-processing, embedding, dimension reduction, and clustering steps.

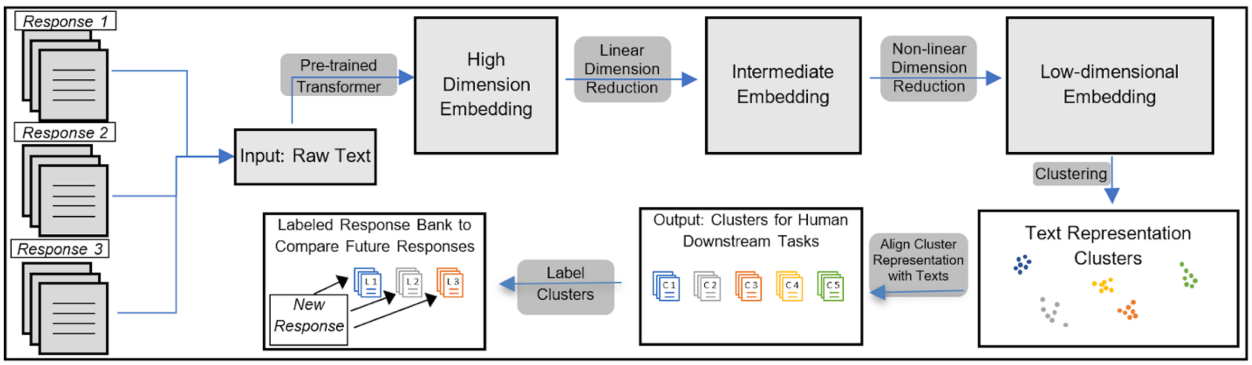


Figure 4.1: Natural Language Processing Cluster-Assisted Workflow

The overarching goal of the NLPCA approach was to take semantically similar responses from students' exam wrapper responses and group them into codes that will be easier to analyze for the researcher. For example, if there were 10000 responses the NLP workflow would cluster semantically similar responses together, which allowed the researcher to quickly skim what the topic is about. This workflow resulted in a faster way for the individual researcher to qualitatively label thousands of responses. For example, consider a scenario where each cluster has 30-40 responses with the same meaning. The human-in-the-loop was then able to skim the clusters to ensure that each sentence was of similar meaning. Once confirming that the sentence in the cluster is talking about a similar topic, the researcher would then find other clusters to group the initial cluster with.

What made NLPCA more attractive for use in qualitative research was that it didn't rely on older monogram-based techniques which grouped responses based on the individual words in the sentence. Instead, it analyzed students' responses at the sentence level (Haller et al., 2022). The NLPCA approach derived meaning by considering the sentence as a whole, a process known as text embedding (Mikolov et al., 2013). This approach was crucial because sentences can use identical words yet convey different meanings. For instance, older NLP techniques might have incorrectly grouped the sentences "Studying engineering is hard" and "I study engineering hard". Even though three out of the four words in these sentences are identical, the meanings are distinct. Such mis-grouping was a common flaw of traditional monogram-based approaches. However, NLPCA differentiated between these sentences, taking into account not just the words but also their order and the overall semantic meaning.

As mentioned above, to utilize the NLP workflow for identifying and categorizing sentences based on their meaning and not just on similar words, a process of text embedding was employed. Text embedding takes a string of text and represents it in a high-dimensional space as a vector (Bujokas, 2020). The specific numbers within this high-dimensional vector were not significant on their own, but their relationships to other vectors or embedded texts were key to grouping strings of text. Consequently, semantically similar sentences ended up being embedded similarly within this high-dimensional space. The NLP workflow then located text vectors with the smallest distances (Euclidean) or smallest angles (cosine similarity) between them.

The NLPCA approach began by first pre-processing the student exam wrapper responses. We first separated all of the responses into sentences, this was done to try and mitigate responses with multiple ideas being clustered together under the assumption that one sentence would most likely contain one idea. Next, the NLPCA embedded each of the student responses into a high-dimensional space using pre-trained, transformer-based models. The fact that the models were pre-trained indicates that they had been trained on a large corpus of text (e.g., all of Wikipedia) to generate the embeddings. This approach simplified the process significantly, as it eliminated the need for us to train a model, which could be a very time-consuming task.

After the students' exam wrapper responses were embedded into a high-dimensional space, it was necessary to cluster similar sentences. However, clustering in a high-dimensional space proved challenging due to the "curse of dimensionality," a phenomenon where high-dimensional vectors are far apart and thus difficult to cluster (Verleysen & François, 2005). To overcome this issue, the number of dimensions of the vectors had to be reduced while still preserving as much information as possible in the compression process.

This was achieved through a series of dimension reduction steps employing Principal Component Analysis (PCA) and Uniform Manifold Approximation Projection (UMAP; McInnes et al., 2020). Both techniques were utilized because PCA tends to lose too much information once the dimensions are reduced to around 80. Therefore, after PCA had brought the dimensions of the sentence vectors down to approximately 80, UMAP was then employed to further reduce the dimensions down to five. Once the vectors' dimensionality was reduced to around five, the system was then able to generate clusters of sentences with similar meanings. This allowed for a more in-depth and nuanced analysis of the student responses.

Once the responses were clustered, the process of manually coding the sentence clusters into topics commenced. It is important to note that the processes described earlier only produced groupings of sentences with similar meanings but did not assign those groupings with meaningful labels - this task fell to the researcher. As the human-in-the-loop, the researcher read the groupings and attached a label to each grouping, a task that was carried out in Google Sheets.

This workflow proves more efficient than manual qualitative analysis, given that the sentences were pre-grouped by NLPCA. For instance, if 70 students talked about a goal-setting strategy with phrases such as “study for 30 minutes every day after class”, “daily studying in the evenings”, or “once a day I shall study”, NLPCA would categorize them into the same cluster despite the varying phrasing describing a similar strategy. It was only necessary to read a few of the responses to ascertain that the topic could be labeled as “Studying every day”. In contrast, the traditional qualitative analysis would require each of those 70 responses to be read at different times during the coding process.

This feature of NLPCA, the capacity to gather like responses together, represented one of its main advantages over the traditional approach. While we coded to saturation using manual qualitative analysis, there was the potential to miss out on topics students discussed because we only analyzed a subset of the data. The NLPCA approach embedded and clustered all 4,442 sentences for “Exam Reflection and all 6,578 sentences for “Preparation Process.”

2.4.4 NLPGPT

For the NLPGPT approach, we used the clusters that were produced by NLPCA, but instead of having the researcher manually label the clusters produced, we prompted GPT-3.5 with the following prompts:

System role: “You are an expert text summarizer.”

User: “You will be given a group of comments that students wrote about their study strategies. Write a two to five-word phrase characterizing what strategy this cluster of comments is about. Your response should start with ‘Phrase:’. \nCLUSTER OF COMMENTS:”

This prompt, along with the cluster of student responses of the same meaning, was sent to the GPT-3.5 application programming interface (API) which then returned a response with the topic of that cluster. This process was repeated for each cluster for both questions using the Python programming language. We merged semantically similar clusters, as the NLPGPT generated multiple clusters that conveyed the same ideas but used different wording. The final step of our analysis was to compare the two NLP approaches with the manual qualitative analysis. Agreement was gauged by the similarity between the codes identified by the three methods and those derived from manual coding. It's important to note that the exact matching of words within the codes wasn't necessary, but the conveyed ideas needed to be the same to be deemed similar. For instance, if the NLP workflow labeled a response as "Practice test questions" and the manual labeling classified that same response as "Do past test problems", they would be counted as similar since they expressed the same idea. A precedent for this was established by Katz et al. (2021), who discovered that the NLP workflow was able to identify similar codes in student responses to end-of-semester survey questions, even though the wording used to describe the codes varied.

2.5 Results

In this section, we present the results of comparing the three qualitative methods: manual qualitative analysis, NLPCA, and NLPGPT. The overview of the comparison is shown in Figure 2.2. Our comparative approach used manual analysis as the baseline for comparing the two NLP methods. We compared the two approaches in both the Preparation Process and Exam Reflection questions. We used this comparison to evaluate how the NLP methods can be used in qualitative analysis. Our evaluation criteria involved examining the total codes generated by each method, assessing the matching of codes between methods, identifying near-match codes, and identifying unique codes between the manual approach and the NLP approaches. Match codes represented the codes agreed upon by both methods in each pair. Near-match codes, on the other hand, are not identical but share similarities across methods. Unique codes are codes that are found only in the method being referenced. In the following sections, we provide specific examples of match, near-match, and unique codes to the two comparison pairs of NLPGPT and manual analysis, and NLPCA and manual analysis. The comparison aims to emphasize the various insights each method can offer, highlight their areas of convergence, and illuminate the unique contributions of each approach.

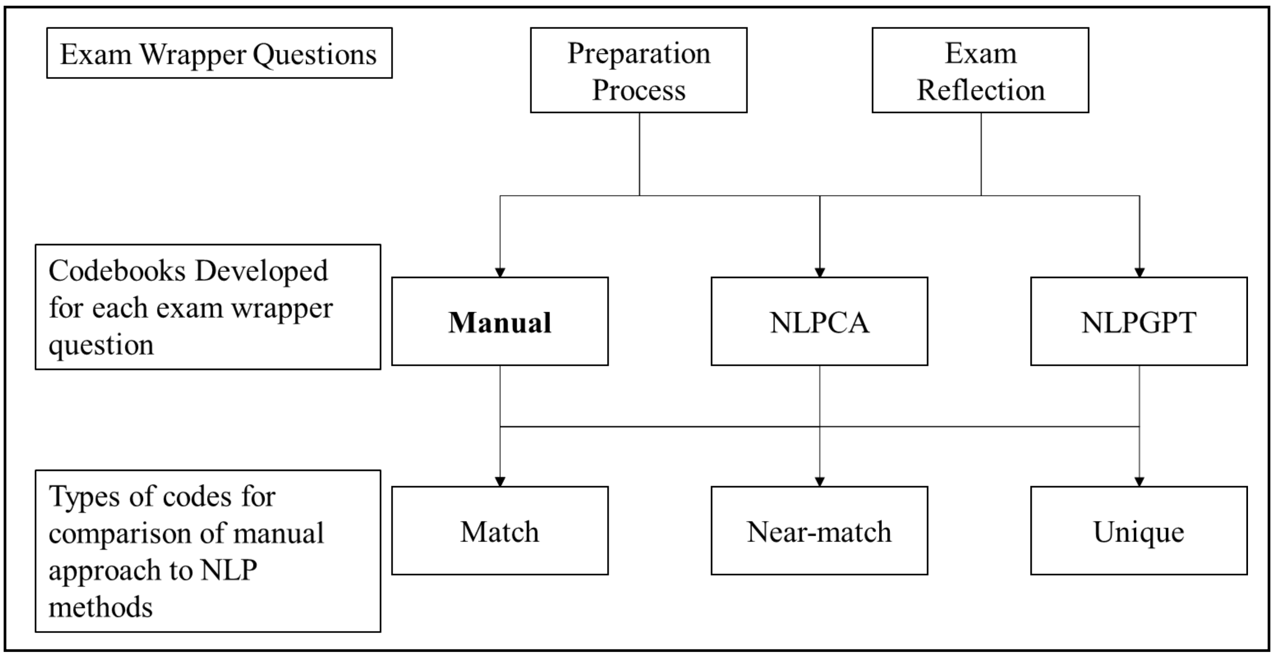


Figure 2.2: Overview of results and comparison of manual analysis to NLPCA and NLPGPT

2.5.1 Manual Analysis Versus NLPCA

Preparation Process

Table 2.2 represents a summary of the match, near-match codes, and unique codes for the comparison between manual qualitative analysis and NLPCA. The manual approach had 27 codes and NLPCA produced 28. Both methods shared 19 match codes, NLPCA had seven near-match codes indicating that seven of the NLPCA nearly matched the codes in the manual approach. The manual approach had eight codes that are unique to this method while NLPCA has two unique codes.

Table 2.2: Match, near-match, and unique codes for manual analysis and NLPCA for the Preparation Process question

	Manual Analysis	NLPCA
Total Codes	27	28
Match Codes	19	19
Near Match Codes	-	7
Unique Codes	8	2

We included some examples of match, near-match, and unique codes in tables 2.3, 2.4, and 2.5 for the manual and NLPCA approach for the Preparation Process question. The examples we showed in Table 2.3 of match codes were both related to practicing old exams or past tests and studying more. Turning to the near-match codes we found that seven code codes in NLPCA have a related meaning to codes in the manual approach. The example we presented in Table 2.4 showed the manual approach labeling the code as Review Notes, whereas NLPCA had two codes related to reviewing notes. One talks about reviewing notes and watching videos and the other is related to reviewing notes and practicing questions. The reason for this naming of the codes in the NLPCA method was because, in the clusters that were analyzed, some of the clusters would have students speaking about both topics in the clusters.

Table 2.3: Examples of match codes between manual analysis and NLPCA for the Preparation Process question

Manual Codes	Manual Examples	NLPCA Codes	NLPCA Examples
Practiced Old Exams/Past Tests	In preparing for this exam I just made my equation sheet. In reviewing for the next exam I could also do practice exams.	Practice Past/Old Exams	Preparing for the exam I went over many practice exams. My process for preparing for this Exam included a lot of practice exams. My process for preparing for the

			exam involved looking over practice problems and reviewing previous exam examples.
Study More and Do More Practice	I study alot the night before the exams. I could study more over the week to succeed more often.	Study More	A way I could strengthen this would be studying the exam questions more intently and spending more time on the practice exams. I can definitely strengthen my preparation by reviewing notes and taking a practice exam each day the week of the exam.

Table 2.4: Example of near-match codes found in NLPCA and their comparison to a similar code in manual analysis for the Preparation Process question

Manual Code	Manual Example	NLPCA Codes	NLPCA Example
Review Notes	I review notes and learning pages to make my equation sheet, and once I make my equation sheet, I do 2-3 practice exams.	Review Notes/Watch Videos	Then I watch videos over the previous exams. and then I watch all of the videos that are associated with previous exams. After that, I usually review old exams by trying some of the problems and watching videos for the ones that have them.
		Review Notes/Practice Questions	I usually take at least two practice exams the night before. I finished all the prep and practice the week before the exam, so that I could start reviewing and going through notes.

Concerning the unique codes, the manual approach had more, with some examples for both methods shown in Table 2.5. The NLPCA method had two unique codes that were not found in the manual approach which discussed doing a few practice exams and putting example problems on the formula sheet. The manual approach however had eight unique codes that discussed watching conceptual videos, studying alone, clicker questions, other commitments, problem understanding, asking for help, nothing to improve on, and getting enough sleep.

Table 2.5: Examples of unique codes found in NLPCA and manual analysis for the Preparation Process question

Manual Codes	Manual Example	NLPCA Codes	NLPCA Examples
Watch Conceptual Videos	For exams I make my equation sheet and do a few practice exams and watch the videos for the problems I don't understand. Studying with a friend could help strengthen that preparation.	Put Examples Problems On Equation Sheet	Then whichever topics that I feel that I need help memorizing , I put examples of problems we have done on my equation sheet. Then I take a practice test and after I have it completed I add practice problems to my equation sheet.
Study Alone	I tried studying in a group this time but found it to be very distracting specifically because it was a group of people that I didn't know well. In the future, I will know to study alone or with closer friends that I feel more comfortable with.	Few practice exams	I only did a few practice exams this time, which led to my absolute destruction on exam 3.

Exam Reflection

Table 2.6 shows the comparison between the manual qualitative analysis and the NLPCA method for the Exam Reflection question. The manual approach had 36 codes in total while NLPCA had 38. Both methods had 25 match codes that matched one another, and the NLPCA method had seven near-match codes which we will discuss below. The NLPCA approach had six unique codes which cannot be found in the manual analysis results while the manual approach had eleven unique codes.

Table 2.6: Match, near-match, and unique codes for manual qualitative analysis and NLPCA for the Exam Reflection question

	Manual Analysis	NLPCA
Total Codes	36	38
Overlapping Codes	25	25
Near Match Codes	-	7
Unique Codes	11	6

We included some examples of match, near-match, and unique codes in tables 2.7, 2.8, and 2.9 for the manual and NLPCA approach for the Exam Reflection question. The examples we show in Table 2.7 of match codes were both related to better conceptual understanding and increasing their grade. The example of near-match codes we present in Table 2.8 compares the manual code of “Study More” to the NLPCA codes of “Did More Practice”, “Spend More Time on Past Exams”, and “Studied More Consistently”. In this case, we concluded that these codes were near-match codes to the manual code “Study More” so we counted one of the codes as a match code and the other two were counted as near-match codes.

Table 2.7: Examples of match codes between manual analysis and NLPCA for the Exam Reflection question

Manual Codes	Manual Examples	NLPCA Codes	NLPCA Examples
Better Conceptual Understanding	I solved all the previous exams to get an idea of what's going to be asked. Additionally, I also went through the concept questions and videos which helped me understand concepts properly. I achieved 80% of my goal. And it helped me a lot.	Increased Awareness Of Conceptual Understanding	I believe it helped show me topics I wasn't confident in. This helped me understand what concepts I was confident in and the ones I was less confident. I was able to slow down and think about my answers and the processes I was going through.
Increased Grade	I watched the example videos on the learning pages much more carefully. This helped a little bit because my exam grade improved a slightly from exam 2. I still did not reach my goal from the last wrapper.	Increased Grade	I increased my grade by 6 points. My grade went up one point. So my grade went up by about 2 points.

Table 2.8: Example of near-match codes found in NLPCA and their comparison to a similar code in manual analysis for the Exam Reflection question

Manual Code	Manual Example	NLPCA Codes	NLPCA Examples

Study More	What I did differently between Exam 1 and Exam 2 was study more multiple choice questions/prep questions. This definitely made a difference because I only missed one multiple choice. My goal was to do 5 points better and I did 15 points better than Exam 1.	Did More Practice	I did more practice exams. I also did more of the practice exams. I also did more practice exams this time.
		Spend More Time On Past Exams	I did two practice exams and timed them. I did an extra practice test.
		Studied More Consistently	I started studying earlier than I did last time. I studied much more and I went about studying differently. All I did differently was studying sooner and more often.

There were also many unique codes to each method – Table 2.9 shows a few examples. The manual approach had codes related to doing past exams without looking at the solutions, asking more questions, feeling less stressed, putting examples on formula sheet, watching videos, timing practice exams, improving time management, clicker questions, problem-solving, paying more attention, and having strong math skills. The NLPCA had unique codes such as not studying enough, exam harder than expected, missed points on exam, not enough time in test, not completing the exam wrapper, and being in a bad study environment.

Table 2.9: Examples of unique codes found in NLPCA and manual analysis for the Exam Reflection question

Manual Codes	Manual Examples	NLPCA Codes	NLPCA Examples
Did Past Exams Without Looking At Solutions	I looked through more tests and actually took one without the solutions instead of just looking over it and saying yeah I can do that. The changes made a huge impact because I made mainly small errors on this test instead of missing an entire problem. I	Didn't Study As Much	I did not spend enough time running through the material I needed to in order to really understand the concepts.

	did reach my goal from the last exam wrapper.		
Asked More Questions	I was more on top of my work throughout the learning pages. I stayed after lecture to ask questions about practice problems. I did more practice exams, but did not reach my goal of the amount of practice exams I wanted to complete.	Exam Harder Than Expected	This change probably helped but the exam was just harder than I expected and I got confused on a few of the problems.

2.5.2 Manual Qualitative Analysis Versus NLPGPT

Preparation Process

In this section, we compared the manual qualitative analysis to the NLPGPT approach. A summary of the results is shown in Table 2.10 where NLPGPT produced 22 codes that matched the manual approach. It is worth noting that this matching was after the researcher grouped similar codes that the NLPGPT produced together. This was because the GPT-3.5 API would sometimes label clusters that express the same idea with slightly different wording. For example, in Table 2.11 we can see that all the NLPGPT codes were about practicing past exams. Finally, the manual approach had five unique codes that were not found in the NLPGPT approach. Near-match codes were only found in the NLPCA method so there will be no reference to near-match codes in this section.

Table 2.10: Match, near-match, and unique codes for manual analysis and NLPGPT for the Preparation Process question

	Manual Analysis	NLPGPT
Total Codes	27	22
Match Codes	22	22
Unique Codes	5	0

Table 2.11: Examples of match codes between manual analysis and NLPGPT for the Preparation Process question

Manual Codes	Manual Examples	NLPGPT Codes	NLPGPT Examples
Practiced Old Exams/Past Tests	In preparing for this exam I just made my equation sheet. In reviewing for the next exam I could also do practice exams.	Practice Exam Emphasis. Practice exams as preparation. Using practice exams for review. Practice exams. Practice exams as preparation. Practice Exam Emphasis. Practice old exams. Reviewing old exams and learning pages. Practice exams for preparation. Using Past Exams for Practice Practice with old exam problems.	Before this exam, I took a couple of practice exams and graded myself on the exams. Before this exam, I took a couple of practice exams and graded myself on the exams. Before the exam I take multiple practice exams without looking at the answers.
Study More And Do More Practice	I study alot the night before the exams. i could study more over the week to succeed more often.	Problem-solving through practice and video tutorials. Practice Problems. Practice and Review. Improving through Review and Practice. Need for more studying. Practice question review. Practice Problem Review Lack of Practice Problems. Practice problem repetition. Practice problem review. Practice exams and problem review.	To improve, I could've studied longer. I think if I was on time and had more study time, I could have done better. I could have studied more. I definitely could have studied more. I definitely could have studied more.

Table 2.12 shows the unique codes produced by the manual qualitative analysis compared to NLPGPT. There were five codes not found in NLPGPT which were studying alone, clicker questions, other commitments, nothing to improve on, and doing lab questions.

Table 2.12: Examples of unique codes found in manual analysis for the Preparation Process question

Manual Codes	Examples
Study Alone	I tried studying in a group this time but found it to be very distracting specifically because it was a group of people that I didn't know well. In the future, I will know to study alone or with closer friends that I feel more comfortable with.
Clicker Questions	I start a week before. I practice past exams and study clicker questions. I also rework prep and practice questions. An area where I can improve

	would be to take more practice exams and test myself on conceptual questions.
Other Commitments	I tried to study a couple days before, but like I said I had 3 other exams around this time and was very flustered. More time.
Nothing To Improve On	My process for taking the module exam is to make sure I am prepared for it. I do this by going over at least two of the previous exams in the past semesters. I feel no need for improvement as this works for me.
Do Lab Questions	I usually take 3 practice exams, and I cover over half of the concept questions on the other practice exams. In addition, I take a look at the lab concept questions as well as they have proved extremely helpful.

Exam Reflection

Between NLPGPT and the manual approach, there were 29 overlapping codes. The manual approach had 36 total codes, with seven of them being unique while NLPGPT had 31 total codes with two unique codes.

Table 2.13: Match, near-match, and unique codes for manual analysis and NLPGPT for the Exam Reflection question

	Manual Analysis	NLPGPT
Total Codes	36	31
Overlapping Codes	29	29
Unique Codes	7	2

In classifying codes in NLPGPT we put any codes that had the same meaning into one category and then matched them with the manual codes. In Table 2.14 one can see that codes such as “Consistent and focused studying”, “Increased study time”, and “Improved study consistency” were put into the same category as “Study More” found in the manual approach. Additionally, we show an example of the different ways that the NLPGPT labeled conceptual understanding-related codes.

Table 2.14: Examples of match codes between manual analysis and NLPGPT for the Exam Reflection question

Manual Codes	Manual Examples	NLPGPT Codes	NLPGPT Examples

Study more	What I did differently between Exam 1 and Exam 2 is study more multiple choice questions/prep questions. This definitely made a difference because I only missed one multiple choice. My goal was to do 5 points better and I did 15 points better than Exam 1.	Consistent and focused studying. Increased study effort. Exam improvement strategies. Practice-based study strategy. Consistent studying over time. Increased Study Time. Consistent study habits. Intensive Pre-Exam Preparation. Improved study consistency. Exam preparation. Consistent study approach.	I practiced more problems and went over old exams, and I also set aside a consistent time to study. I spent more time studying, took more notes on my note sheet, and worked on practice exam questions.
Better conceptual understanding	I solved all the previous exams to get an idea of what's going to be asked. Additionally, I also went through the concept questions and videos which helped me understand concepts properly. I achieved 80% of my goal. And it helped me a lot.	Utilizing Resources and Understanding Concepts Improved Conceptual Understanding. Understanding concepts through preparation. Conceptual Understanding and Application. Improving Material Comprehension. Lack of Understanding and Effort.	I worked on understanding the concepts instead of focusing on how to do the problem. I also went to the ef study room to get help when I was confused on problems instead of just looking at the discussion board.

Table 2.15 shows the table of unique codes found in the manual approach and the NLPGPT approach. Two codes were in the NLPGPT method which were mistake prevention and lack of preparation, whereas there were seven codes found in the manual approach that were not in the found in the NLPGPT approach. These codes included doing past exams without looking at solutions, asking more questions, being less stressed, watching videos, looking at clicker questions, improved problem solving, and strengthening math skills.

Table 2.15: Examples of unique codes between manual analysis and NLPGPT for the Exam Reflection question

Manual Codes	Manual Examples	NLPGPT Codes	NLPGPT Examples
Did Past Exams Without	I looked through more tests and actually took one without the solutions instead of just looking over it and saying yeah I can	Mistake Prevention.	It was because I made simple mistakes.

Looking At Solutions	do that. The changes made a huge impact because I made mainly small errors on this test instead of missing an entire problem. I did reach my goal from the last exam wrapper.		but I ended up making stupid mistakes.
Asked More Questions	I was more on top of my work throughout the learning pages. I stayed after lecture to ask questions about practice problems. I did more practice exams, but did not reach my goal of the amount of practice exams I wanted to complete.	Lack Of Preparation.	I did not prepare as well as I did for exam 2. I didn't have nearly as much time to study due to other assignments, so I came into the exam less prepared than Exam 1.
Less Stressed	I managed time very well on this exam. I was able to complete much more of the exam, and my grade was 20 points higher. I reached my goal from last exam wrapper, and it helped me so much stress-wise.		

2.6 Discussion

In this section, we compare the results of the three methods of qualitative analysis in manual qualitative analysis, NLPCA, and NLPGPT. The purpose is to compare a long-established practice in qualitative analysis to NLP methods. The results reveal that manual analysis and the two NLP methods show a high degree of matching codes, with some unique codes identified in each method. There is a significant overlap in codes between the manual approach and both NLPCA and NLPGPT indicating that both NLP methods can be used in qualitative analysis. Furthermore, there were unique codes identified by each method, suggesting that NLP techniques could complement the manual approach by producing some insights that may not be seen with the manual approach. The codes generated by NLPGPT were also accurate in labeling the clusters when it was checked by the researcher. Looking at the Preparation Process question, manual analysis, and NLPCA had 19 codes that matched, while NLPGPT had 22 out of 27 manual codes matching. This shows that NLPGPT performed better than NLPCA when compared to the manual qualitative analysis in this question in terms of reproducing similar codes to the manual approach. The reason for this could be the way that the researcher grouped the first round of codes in the NLPGPT approach. In the NLPGPT approach, the researcher grouped codes using the manual approach as the comparison group. On the other hand, the NLPCA codes were compared completely independently which could be why there were more unique codes and fewer matching ones. Both NLPCA and manual analysis produced unique codes, indicating that NLPCA could be used to gather more insights in qualitative analysis that might not be identified if only using the manual approach.

In the Exam Reflection question, the NLP methods again showed matching codes to a high degree when compared to manual analysis. For this question, NLPCA had 25 matching codes and NLPGPT had 29 matching codes when compared to the manual approach. In both questions, NLPGPT had over 80% of its codes matching with the manual approach and NLPCA close to 70%. However, there were 11 unique codes identified in the manual approach not identified by NLPGPT. The level of matching between NLPGPT and the manual approach shows a positive avenue to use this NLP method for qualitative analysis since there were above 80% matching codes. This is deemed acceptable if one uses the concept of inter-rater reliability (IRR), which is the level of agreement between the codes that researchers produce (McAlister et al., 2017).

In both questions, the manual approach was able to identify more unique codes than NLPCA and NLPGPT. A possible reason for this is the way that the responses were clustered. In the manual approach, the researcher would read through each response and code the responses at an individual level. In the NLPCA approach, the researcher would cluster the data, and due to the column of data, these clusters could contain a large number of responses therefore when the researcher labels these clusters, we could have lost some of the nuance that would have been seen when coding the data on an individual response level. Therefore, some nuanced information could have been missed when considering clusters as a whole. For example, the unique manual code “Did past exams without looking at the answers” is a specific strategy that may not have been mentioned by enough students to form its cluster, therefore it may have been clustered with responses related to past exams so that response would have been coded as “Did Past Exams” in the NLPCA or NLPGPT method.

One of the major benefits of the NLP approaches we used is that the language models that were used in this study were pre-trained. Pre-trained models save time for users since the training of an NLP model can be resource and time-intensive. Furthermore, the pre-trained models could be used in various contexts since many trained models will only perform well on specific NLP tasks (Crowston et al., 2012). Specifically, NLPGPT uses GPT-3.5, a general-purpose LLM pretrained that showed high match codes with the manual approach. Another benefit to NLPGPT is that using the GPT-3.5 API would require less programming knowledge than previous studies (Leeson et al., 2019). Previous studies using NLP have also required a lot of backend programming (Leeson et al., 2019) or they were resource intensive through making a dictionary that the NLP model could follow (Crowston et al., 2012). Finally, the high level of agreement between the manual approach and the NLP methods that we have seen in this study is in line with (Katz et al., 2023) who used a similar workflow to code student career interest essays. This confirms that the NLP workflows used in this study could be further explored in different contexts with acceptable accuracy when compared to traditional manual approaches.

2.6.1 Study Implications

This study offers valuable insights into the potential of leveraging LLMs for qualitative codebook generation and shows versatility in how LLMs can be used. Large language models such as GPT-3.5 have become more prevalent in their use of various tasks involving natural language, including qualitative data analysis (Sok & Heng, 2023). As demonstrated by the comparison of manual qualitative analysis, NLPCA, and NLPGPT each method was able to produce similar codes to a significant degree. The NLP methods that we compare can significantly automate qualitative analysis while still producing meaningful insights. This could increase the efficiency of analysis. However, due to the higher volume of unique codes produced by the manual approach, human input is still crucial in the qualitative analysis process. At present, NLP can complement, not replace manual analysis since NLP methods are particularly useful for handling large volumes of data for qualitative analysis, where manual coding alone may not be feasible (Katz et al., 2023).

The advantage of LLMs such as GPT-3.5 is that they are more complex and flexible than NLP methods that previously used a rule-based or word-based approach. Due to the advances in these LLMs, and the potential gain in efficiency, researchers could spend more time studying more complex phenomena in the data that is only currently possible with human cognition which could lead to new insights could deepen the understanding researchers have when studying complex phenomena as was previously shown in other areas (Hovy & Lavid, 2010).

Lastly, our study illuminates potential areas for future investigation. The refinement and application of NLP tools in qualitative research warrants further exploration. Additionally, their potential use in other domains, beyond the context of this study, could yield valuable insights (Liu et al., 2019). One of the potential avenues for further research would be to investigate various prompts and see how they affect how the qualitative codes are generated The way LLMs such as GPT-3.5 are prompted affects the model's output, and these outputs could change the codes or summaries produced by the model (Liu et al., 2023). Another avenue for further research would be combining different manual and automated techniques in quaThis study has explored this multi-methods approach by manually labeling clusters in the NLPCA approach and merging codes produced by GPT-3.5 into overarching codes in the NLPGPT approach. The NLPGPT approach. An example of this future exploration could run the first round of codes into a language model to further merge codes into broader themes (Katz et al., 2023).

The current study used MPNet which is a transformer-based model, for one of the methods we tested, as well as GPT-3.5 to assist in generating a qualitative codebook. For further investigation, we could test models such as Anthropic's Claude, Google's Bard, open-source models, or the latest version of the GPT models, GPT-4. Another avenue for potential for further investigation would be to test the generalizability of the NLP methods. This study produced favorable results for the NLP-based methods we tested when comparing them to manual qualitative analysis, but the NLP approaches were only shown to be effective in this specific context. Therefore, it would be beneficial to replicate this study in different research contexts, with varying types of qualitative data, to assess the applicability and robustness of using LLMs and NLP techniques in various research scenarios. Finally, we believe this method could result in much time saved and enhanced insights in qualitative analysis. An economic analysis of using LLMs and NLP could assist researchers in deciding whether these techniques are worth exploring from the perspective of time and money.

2.7 Conclusion

This study was conducted to compare the effectiveness of a well-known qualitative method of manual qualitative analysis, an NLPCA approach, and the NLPGPT approach for coding qualitative data. Our results demonstrated promising potential for using state LLMs and NLP to automate some parts of qualitative coding, providing an avenue to have a complementary approach to qualitative analysis using LLMs and a traditional manual approach. These findings expand our understanding of how NLP and LLMs can be effectively applied in the context of qualitative data analysis of first-year engineering student exam wrapper responses. This study also demonstrates potential avenues to explore the NLP workflows' ability to conduct qualitative analysis in other contexts, using diverse types of qualitative data sources

While this study demonstrates positive implications for using LLMs in qualitative analysis, we must be aware of the limitations of these methods. Firstly, we have studied a specific context and type of data – exam wrappers. The robustness of the NLP workflow needs to be tested on diverse types of qualitative data such as interview data, observational notes, and student essays to further confirm LLMs' applicability. A human also needs to be involved in checking the codebook generated to ensure the LLMs' accuracy. With the limitations of NLP and LLMs considered, this study represents a valuable step towards a more integrative approach in qualitative research, where human expertise and LLMs can be combined to unlock a deeper and more nuanced understanding of qualitative data while also cutting down on the time taken to conduct qualitative analysis.

References

Abram, M. D., Mancini, K. T., & Parker, R. D. (2020). Methods to Integrate Natural Language Processing Into Qualitative Research. *International Journal of Qualitative Methods*, *19*, 1609406920984608. https://doi.org/10.1177/1609406920984608

Armstrong, D., Gosling, A., Weinman, J., & Marteau, T. (1997). The Place of Inter-Rater Reliability in Qualitative Research: An Empirical Study. *Sociology*, *31*(3), 597–606. https://doi.org/10.1177/0038038597031003015

Aspers, P., & Corte, U. (2019). What is Qualitative in Qualitative Research. *Qualitative Sociology*, *42*(2), 139–160. https://doi.org/10.1007/s11133-019-9413-7

Bahdanau, D., Cho, K., & Bengio, Y. (2016). *Neural Machine Translation by Jointly Learning to Align and Translate* (arXiv:1409.0473). arXiv. https://doi.org/10.48550/arXiv.1409.0473

Becker, J. P., Kahanda, I., & Kazi, N. H. (2021, July 26). *WIP: Detection of Student Misconceptions of Electrical Circuit Concepts in a Short Answer Question Using NLP*. 2021 ASEE Virtual Annual Conference Content Access. https://peer.asee.org/wip-detection-of-student-misconceptions-of-electrical-circuit-concepts-in-a-short-answer-question-using-nlp

Berdanier, C. G. P., Baker, E., Wang, W., & McComb, C. (2018). Opportunities for Natural Language Processing in Qualitative Engineering Education Research: Two Examples. *2018 IEEE Frontiers in Education Conference (FIE)*, 1–6. https://doi.org/10.1109/FIE.2018.8658747

Bhaduri, S. (2018). *NLP in Engineering Education-Demonstrating the use of Natural Language Processing Techniques for Use in Engineering Education Classrooms and Research* [PhD Thesis]. Virginia Tech.

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

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bujokas, E. (2020, May 30). *Creating Word Embeddings: Coding the Word2Vec Algorithm in Python using Deep Learning*. Medium. https://towardsdatascience.com/creating-word-embeddings-coding-the-word2vec-algorithm-in-python-using-deep-learning-b337d0ba17a8

Case, J. M., & Light, G. (2011). Emerging Research Methodologies in Engineering Education Research. *Journal of Engineering Education*, *100*(1), 186–210.

Crowston, K., Allen, E. E., & Heckman, R. (2012). Using natural language processing technology for qualitative data analysis. *International Journal of Social Research Methodology*, *15*(6), 523–543.

Cunningham, P., Matusovich, H. M., Hunter, D.-A. N., Blackowski, S. A., & Bhaduri, S. (2017, June 24). *Board # 28: Beginning to Understand Student Indicators of Metacognition*. 2017 ASEE Annual Conference & Exposition. https://peer.asee.org/board-28-beginning-to-understand-student-indicators-of-metacognition

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. https://doi.org/10.48550/arXiv.1810.04805

Ganesh, A., Scribner, H., Singh, J., Goodman, K., Hertzberg, J., & Kann, K. (2022). Response Construct Tagging: NLP-Aided Assessment for Engineering Education. *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*, 250–261. https://aclanthology.org/2022.bea-1.29

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Guetterman, T. C., Chang, T., DeJonckheere, M., Basu, T., Scruggs, E., & Vydiswaran, V. G. V. (2018). Augmenting Qualitative Text Analysis with Natural Language Processing: Methodological Study. *Journal of Medical Internet Research*, *20*(6), e9702. https://doi.org/10.2196/jmir.9702

Haller, S., Aldea, A., Seifert, C., & Strisciuglio, N. (2022). *Survey on Automated Short Answer Grading with Deep Learning: From Word Embeddings to Transformers*.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hovy, E., & Lavid, J. (n.d.). *Towards a 'Science' of Corpus Annotation: A New Methodological Challenge for Corpus Linguistics*.

Katz, A., Norris, M., Alsharif, A. M., Klopfer, M. D., Knight, D. B., & Grohs, J. R. (2021, July 26). *Using Natural Language Processing to Facilitate Student Feedback Analysis*. 2021 ASEE Virtual Annual Conference Content Access. https://peer.asee.org/using-natural-language-processing-to-facilitate-student-feedback-analysis

Katz, A., Shakir, U., & Chambers, B. (2023). *The Utility of Large Language Models and Generative AI for Education Research* (arXiv:2305.18125). arXiv. https://doi.org/10.48550/arXiv.2305.18125

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-022-13428-4

Lakretz, Y., Dehaene, S., & King, J.-R. (2020). What Limits Our Capacity to Process Nested Long-Range Dependencies in Sentence Comprehension? *Entropy*, *22*(4), Article 4. https://doi.org/10.3390/e22040446

Leeson, W., Resnick, A., Alexander, D., & Rovers, J. (2019). Natural Language Processing (NLP) in Qualitative Public Health Research: A Proof of Concept Study. *International Journal of Qualitative Methods*, *18*, 1609406919887021. https://doi.org/10.1177/1609406919887021

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. sage.

Liu, M., Shum, S. B., Mantzourani, E., & Lucas, C. (2019). Evaluating machine learning approaches to classify pharmacy students' reflective statements. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11625 LNAI*, 220–230. Scopus. https://doi.org/10.1007/978-3-030-23204-7_19

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, *55*(9), 195:1-195:35. https://doi.org/10.1145/3560815

Lovett, M. C. (2013). Make exams worth more than the grade: Using exam wrappers to promote metacognition. In: Kaplan, M., Silver, N., LaVague-Manty, D., Meizlish, D. (eds) Using reflection and metacognition to improve student learning: Across the disciplines, across the academy,. In *Using reflection and metacognition to improve student learning: Across the disciplines, across the academy*, (pp. 18–52).

Luo, W., Liu, F., Liu, Z., & Litman, D. (2016). Automatic Summarization of Student Course Feedback. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 80–85. https://doi.org/10.18653/v1/N16-1010

Mackieson, P., Shlonsky, A., & Connolly, M. (2019). Increasing rigor and reducing bias in qualitative research: A document analysis of parliamentary debates using applied thematic analysis. *Qualitative Social Work*, *18*(6), 965–980.

McAlister, A. M., Lee, D. M., Ehlert, K. M., Kajfez, R. L., Faber, C. J., & Kennedy, M. S. (2017, June 24). *Qualitative Coding: An Approach to Assess Inter-Rater Reliability*. 2017 ASEE Annual Conference & Exposition. https://peer.asee.org/qualitative-coding-an-approach-to-assess-inter-rater-reliability

McInnes, L., Healy, J., & Melville, J. (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* (arXiv:1802.03426). arXiv. http://arxiv.org/abs/1802.03426

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space* (arXiv:1301.3781). arXiv. https://doi.org/10.48550/arXiv.1301.3781

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). *Learning Transferable Visual Models From Natural Language Supervision* (arXiv:2103.00020). arXiv. https://doi.org/10.48550/arXiv.2103.00020

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (n.d.). *Improving Language Understanding by Generative Pre-Training*.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (n.d.). *Language Models are Unsupervised Multitask Learners*.

Saldaña, J. (2014). Coding and Analysis Strategies. In P. Leavy (Ed.), *The Oxford Handbook of Qualitative Research* (pp. 580–598). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199811755.013.001

Seale, C. (2004). *Social Research Methods: A Reader*. Psychology Press.

Sebesta, A. J., & Bray Speth, E. (2017). How Should I Study for the Exam? Self-Regulated Learning Strategies and Achievement in Introductory Biology. *CBE Life Sciences Education*, *16*(2), ar30. https://doi.org/10.1187/cbe.16-09-0269

Sok, S., & Heng, K. (2023). *ChatGPT for Education and Research: A Review of Benefits and Risks* (SSRN Scholarly Paper 4378735). https://doi.org/10.2139/ssrn.4378735

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Verleysen, M., & François, D. (2005). The Curse of Dimensionality in Data Mining and Time Series Prediction. In J. Cabestany, A. Prieto, & F. Sandoval (Eds.), *Computational*

Intelligence and Bioinspired Systems (pp. 758–770). Springer. https://doi.org/10.1007/11494669_93

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., … Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

Wulff, P., Westphal, A., Mientus, L., Nowak, A., & Borowski, A. (2023). Enhancing writing analytics in science education research with machine learning and natural language processing—Formative assessment of science and non-science preservice teachers' written reflections. *Frontiers in Education*, *7*. https://www.frontiersin.org/articles/10.3389/feduc.2022.1061461

Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P.-Y. (2023). Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 75–78). Association for Computing Machinery. https://doi.org/10.1145/3581754.3584136

Chapter 3: Manuscript 2

Utilizing Natural Language Processing to Examine Self-Reflections in Self-Regulated Learning

This manuscript includes intellectual contributions from Rachel McCord Ellestad and Andrew Katz

3.1 Abstract

Self-regulated learning (SRL) and reflection are crucial skills for first-year engineering students given their association with academic success. Exam wrappers are a simple tool that can help students develop their SRL skills by encouraging them to reflect on their exam performance and identify areas where they need to improve. Students can learn to regulate their learning by setting goals, developing study strategies, and monitoring their progress. Exam wrappers can also be useful to instructors and researchers since they offer insights into the learning process that students undergo and can allow both instructors and curriculum designers to make adjustments that can increase student success. Unfortunately, analyzing written exam wrapper responses can be time-intensive in large classrooms or research settings due to the large volumes of text that need to be analyzed. Advances in natural language processing (NLP) and modern neural network architectures present opportunities for higher education teaching and research communities to address this challenge. In this study, we investigated the use of a transformer-based NLP workflow on written exam wrapper responses. We found that the NLP workflow demonstrated high accuracy on a dataset of exam wrapper responses. These results suggest the investigated technique of using pre-trained zero-shot classification models can be used in other instructional and research settings to help analyze student writing.

Keywords: natural language processing; self-regulated learning; exam wrappers

3.1 Introduction

3.1.1 Self-Regulated Learning

Vosniadou (2020) suggested that enhancing students' self-regulated learning (SRL) can help facilitate the transition from secondary school to higher education. Self-regulated learners approach tasks with confidence, diligence, and resourcefulness. They are aware of their knowledge and skill levels, characteristically recognizing when they know a fact or possess a skill and when they do not (Zimmerman & Schunk, 1989). Although numerous scholars have studied SRL (Boekaerts, 1999; Butler & Winne, 1995; Pintrich, 2000; Schunk, 2008; Zimmerman, 1986), and various definitions exist, in this study we adopt the definition provided by a prominent author in the field, Barry J. Zimmerman. He described SRL as "self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals" (Zimmerman, 2000, p. 14). Zimmerman's SRL model comprises three subprocesses: the Forethought Phase, the Performance Phase, and the Self-Reflection phase. The Forethought Phase involves planning, goal-setting, and integrating motivational beliefs. The Performance Phase occurs during task execution and encompasses time management, help-seeking, and task strategies. Lastly, the Self-Reflection phase includes self-evaluation, self-satisfaction, and adaptive reactions.

Numerous studies have investigated the implications of SRL for learning across various fields (Boekaerts, 1999; Schunk & Zimmerman, 2011; Wallin & Adawi, 2018; Zimmerman, 1986) including engineering (Ellestad, 2016; Gynnild et al., 2008; Lawanto et al., 2014; Menekse, 2020; Nelson et al., 2015; Wedelin et al., 2015). All cited research indicates an association between SRL and academic performance. Additionally, most studies call for further exploration to better comprehend the distinct aspects of SRL and how they specifically relate to students' performance in courses.

3.1.2 Exam Wrappers

Receiving graded exams presents an often-overlooked opportunity for students to improve their learning strategies. These exams allow students to reflect on their performance and analyze how their study strategies and learning activities contributed to their results (Lovett, 2013). For example, insufficient regulation of study strategies might involve not understanding how to study or how to address confusion while learning (Shell et al., 2013). Students can use graded exams to reflect on their learning activities in preparation for exams, which is another aspect of regulation (Castellanos & Enszer, 2013). Actively involving students in their learning process through reflection can enhance academic success (Tuckman & Kennedy, 2011). However, this can be challenging in large foundational engineering classrooms (Mervis, 2013), where limited interaction between instructors and students often occurs. This lack of interaction results in instructors having reduced awareness of students' needs and strengths (Panadero et al., 2016).

To address the challenge of understanding student learning strategies in large foundational engineering classrooms, we propose that gaining insight into students' SRL strategies during exam preparation could help them improve their academic performance in foundational engineering courses. Vosniadou (2020) suggested that enhancing students' SRL is one way to facilitate the transition from secondary school to higher education. Self-regulated learners approach tasks with confidence, diligence, and resourcefulness, and are aware of their knowledge and skill levels (Zimmerman and Schunk, 1989).

Exam wrappers present a solution for providing feedback to students in large classrooms. Carpenter et al. (2020) addressed this challenge by using multiple-choice questions with predefined feedback, which was released to students based on their selections. Although the authors found a relationship between students' use of exam wrappers and their course performance, they noted that students' metacognitive skills - a crucial aspect of SRL related to knowledge and regulation of cognition (Dinsmore et al., 2008) - did not necessarily improve. Additionally, multiple-choice exam wrappers do not promote as deep a reflection as open-ended questions do (Panadero et al., 2016).

3.1.3 Modern Natural Language Processing

The development of NLP and LLMs, such as OpenAI's GPT-4 and Google's Bard, has created new opportunities for processing extensive textual data. NLP encompasses various techniques for analyzing natural language, with applications ranging from text translation and summarization to part-of-speech tagging (POS) and co-reference resolution (Khurana et al., 2022). In engineering education, researchers and practitioners have explored NLP tools for tasks like automatic short answer grading (ASAG), automatic essay scoring (AES; Haller et al., 2022), and automatic question generation (AQG). Beyond assessment purposes, NLP has been employed in engineering education research to summarize large text corpora (Katz et al., 2021), detect student sentiments about courses (Ganesh et al., 2022), analyze engineering students' use of disciplinary discourse in resumes (Berdanier et al., 2018), assess students' metacognitive development in classrooms (Bhaduri, 2018), evaluate student conceptual understanding (Arbogast & Montfort, 2016), and summarize student reflections on confusing concepts using mobile applications (Butt et al., 2022; Menekse, 2020).

3.1.4 Natural Language Processing Using Transformers

Some earlier work on NLP in engineering education has relied on monogram-based approaches, which may not adequately capture semantic meaning in sentences, thus limiting insights into student textual data. For instance, older NLP techniques might identify a common word like "engineering" in a document but not provide insights into the specific aspects of engineering being discussed. Recent NLP techniques, using transformer-based approaches (Vaswani et al., 2017), have demonstrated greater sensitivity in detecting semantic meaning in sentences compared to monogram-based methods (Becker et al., 2021; Ganesh et al., 2022). Transformer-based approaches can discern meaning in longer sentences (Haller et al., 2022), making them more advanced than their monogram-based counterparts.

Transformers are neural network architectures employing attention mechanisms (Vaswani et al., 2017), which are deep-learning techniques capable of modeling long-range dependencies in sentences, regardless of word distance. Long-range dependencies occur when words in a sentence are not sequentially related. For example, in the sentence "The **dog** that the cat chased **ran** away," a center-embedded clause ('the cat chased') creates a long-range dependency between the main subject and verb ('dog' and 'ran,' respectively – in bold; Lakretz et al., 2020). Another advantage of transformer-based approaches is their use of text embeddings, which extend the concept of word embedding to entire text segments. The objective is to identify when students' comments share a common topic and distinguish those from comments expressing different ideas without relying solely on lexicographical similarity.

Recent studies have employed transformer-based models to analyze reflections in various contexts. Wulff et al. (2022) used a pre-trained transformer model called BERT on preservice teachers' written reflections and found it outperformed other deep learning architectures and word-based algorithms for reflective writing classification. Wang et al. (2019) applied a transformer-based model to student peer evaluations, achieving a 61.5% accuracy score, which they considered satisfactory given the small dataset (480 instances) used. Magooda et al. (2022) improved student reflection quality with a transformer-based NLP model called Distil-BERT by analyzing student reflections and providing automated, timely feedback on reflection quality. Nehyba & Štefánik (2022) attempted to build automated transformer-based models to classify student-teacher reflections, achieving 76.56-79.37% accuracy. While these studies utilized state-of-the-art transformer-based NLP techniques, they relied on existing labeled data or manual classification of the full dataset to train their models and provide accuracy metrics, a time-intensive and application-specific process. To facilitate the widespread adoption of transformer-based models in educational settings, a more general framework is needed. In this study, we explore a method requiring only a codebook and unlabeled text as inputs to label text based on the codebook codes.

Our work aims to build on the use of transformer-based models to identify various SRL strategies students discuss in their exam wrappers. These insights could be valuable for instructors interested in understanding the SRL strategies students employ while preparing for exams. Currently, instructors must read each response, a tedious and time-consuming task, especially in larger classes. The pre-trained, zero-shot classification (ZSL) model we investigate in this study offers a low-resource option for instructors and researchers to evaluate SRL strategies in large classes or study samples, respectively.

3.1.5 Zero-shot Learning Classification

To label input sentences from exam wrappers, we employed ZSL to classify the complete dataset of student responses to a specific exam wrapper question (Yin et al., 2019). Zero-shot learning is a state-of-the-art NLP method that embeds both the text to be labeled and the codebook (i.e., labels) into a high-dimensional space. Yin et al. (2019) refer to the text we want to label as the “premise” and the label text strings as the “hypothesis.” The task is then to determine if the hypothesis is true or false, given the premise - whether the text being labeled matches a particular label in the codebook within the high-dimensional space. Both high-dimensional vectors are passed through the model, which outputs vectors of logits, later converted to probabilities. The resulting probability distribution indicates the model’s confidence that a sentence corresponds to a given label. The sentence-label pairs with the highest probabilities are then selected as the model’s prediction for that sentence. A key advantage of the ZSL classifier is its ability to assign multiple labels to a single sentence if more than one idea is present, an improvement over previous NLP approaches.

3.2 Methods

3.2.1 Data Collection

Data for this study were collected in a physics for engineering students’ course at a large R1 university in the Southeastern region of the United States. The course is required for all incoming first-year students in the College of Engineering at the university. The course uses a flipped-classroom approach and is delivered in person three times per week, and labs happen twice per week. Summative assessment mainly occurs through exams, which count for 54% of the final grade. The study sample included any student who submitted an exam wrapper during the Spring 2021, Fall 2021, and Spring 2022 semesters.

Exam wrappers were given to students to complete after the first three exams in the course. Each wrapper included 9 questions that asked students to reflect on their performance on the exam, reflect on future strategies, and develop a strategic plan for preparing for the next exam. Table 3.1 summarizes the exam wrapper questions. Table 3.1 also shows which SRL constructs from Zimmerman's model most align with that specific question. The student responses were written in paragraph form and stored in the university learning management system with a unique identifier for students.

Table 3.1. Exam wrapper questions and different SRL constructs associated with those questions

Wrapper Section	**Wrapper Question**	**SRL Construct**
Reflection	Reflection - What did you do differently between Exam 1 and Exam 2? Did the changes that you made make an impact? Did you reach your goal from the last Exam Wrapper?	Self-evaluation Adaptive reactions
Exam Dissection	There are skills other than physics knowledge necessary to complete this exam. Can you identify any skills or fundamental knowledge (non-physics) that are weak that impeded your ability to show what you know about physics concepts? What evidence do you have to backup your answer?	Task strategies Time management
Exam Preparation Reflection	Describe your process for learning/engagement during the regular week for this module. Can you identify any areas of improvement that could strengthen your learning during the regular week moving forward?	Task strategies Time management Environmental structuring Help-seeking Interest incentives Goal setting Strategic planning
	Describe your process for preparing to take the module exam. Can you identify any areas of improvement that could strengthen your preparation activities?	Task strategies Time management Environmental structuring Help-seeking Interest incentives Goal setting Strategic planning Self-evaluation
	How confident were you when the exam was passed out that you were ready to show what you knew about this module? What is one thing YOU could do over the next three weeks to support building confidence? What is one thing your instructor could help with to support building your confidence?	Self-efficacy
Strategic Plan	Define a measurable goal you would like to achieve during our next class module. This goal should be measurable and attainable in the next three-week period.	Goal setting Strategic planning Time management

	Identify one action you want to START doing that may better support your learning in this next module. Can you describe a specific action plan to support you in starting this action?	Strategic planning Adaptive reactions
	Identify one action you want to STOP doing that is detrimental to your learning in this next module. Can you describe a specific action plan to support you in stopping this action?	Defensive reactions Strategic planning

3.2.2 Data Analysis

In this study, we aimed to pinpoint the learning strategies that students mentioned in their exam wrappers. To achieve this, we employed a three-step workflow:

(1) Develop a codebook by applying NLP and text clustering techniques.

(2) Assess the phrasing of codebook labels by examining a subset of the model's output. Modify the codebook iteratively based on the accuracy of code predictions.

(3) Utilize the improved codebook to categorize the entire dataset and evaluate the final labeling accuracy.

3.2.2.1 Questions analyzed

We evaluated the model's classification performance based on three exam wrapper questions, shown in Table 3.2. These questions were selected as they encompass most of the SRL constructs in Zimmerman's model. The chosen questions include:

(1) "Other skills" – This question primarily addresses the Reflection Phase of Zimmerman's model, as it prompts students to consider skills beyond physics knowledge needed to complete the exam

(2) "Learning process" – This question relates to the Performance Phase, as it inquires about the students' actual learning process during the module. Additionally, it pertains to the

Forethought Phase by encouraging students to identify ways to enhance their learning throughout the module.

(3) "Start action" – This question is associated with the Forethought Phase, as it urges students to pinpoint actions they can initiate to bolster their learning for the subsequent module. It also requests a detailed plan to improve their learning in the next module.

It is important to note It is important to note that while we categorize each question under a specific phase of Zimmerman's model, there are multiple sub-questions within each question name. For instance, "Other skills" not only asks students to identify non-physics skills that may have hindered their exam performance but also requests evidence to support their response. These additional sub-questions often overlap with other phases of Zimmerman's model, serving as a bridge between phases. The purpose of this design is to guide students through the full SRL cycle for their varied responses and help them grasp the comprehensive objectives of the exam wrapper.

Table 3.2. Questions analyzed for this study

Question name	**Question**	**SRL Constructs**
Other skills	There are skills other than physics knowledge necessary to complete this exam. Can you identify any skills or fundamental knowledge (non-physics) that are weak that impeded your ability to show what you know about physics concepts? What evidence do you have to back up your answer?	Task strategies Time management
Learning process	Describe your process for learning/engagement during the regular week for this module. Can you identify any areas of improvement that could strengthen your learning during the regular week moving forward?	Task strategies Time management Environmental structuring Help-seeking Interest incentives Goal setting Strategic planning
Start action	Identify one action you want to START doing that may better support your learning in this next module. Can you	Strategic planning Adaptive reactions

	describe a specific action plan to support you in starting this action?	

3.2.3 Codebook Background

To generate a codebook, one member of the research team read a subset of the data. This reading produced codes to describe each study strategy that students mentioned. Further details on codebook generation can be found at Gamieldien et al. (2023). For this paper and the method investigated, it is sufficient to assume one has a codebook without needing to know whether it was generated through traditional analysis, informed by existing literature, suggested by an automated NLP process, or arrived at by any other means.

3.2.3.1 Method Evaluation

We evaluated the ZSL classifier accuracy on 10% of the responses. The process of evaluation was done by reading and comparing the sentence with its model-assigned label(s). If the label and the sentence matched in meaning, a score of "1" was given. If the label and the sentence did not match, i.e., the sentence did not match the code, it was assigned a "-1". We also created an intermediate label where the sentence and the label were semantically related, but not an exact match. Those instances were assigned a "0". The prediction accuracy of the ZSL was measured by the proportion of the sentences assigned a label of "1" for each question and the total number of labels assigned.

Once the accuracy of the codebook was calculated, we evaluated which codes in the codebook were producing the most inaccurate labels. We did this by checking if there was an association between a certain code and the prediction accuracy of the ZSL classifier. In the case that a label was producing a large number of inaccuracies - a high number of "-1" labels - we would reword that label or remove the label. We would then run the ZSL classifier with the new iterated codebook and evaluate whether the prediction accuracy would increase.

3.2.4 Ethical Considerations

Ethical approval was obtained from the university’s Institutional Review Board (IRB-22-07020-XP). The data we received was de-identified by one of the researchers before it was shared with the research team to protect the identity of the participants.

3.3 Results

In this section, we will first present the final codebooks that we developed, which were created from the clustering of the NLP workflow results after iterating on certain codes. We then evaluated the topics in the codebook of the ZSL classifier by looking at the accuracy of 10% of the full dataset. We did this to evaluate whether using the ZSL classifier produced results that would allow us to extract meaningful insights from the data accurately.

3.3.1 Evaluation Results

The evaluation of the codebooks is presented in Tables 3.3, 3.5, and 3.7 below. The most accurately labeled dataset was “Start action” with 91%, then “Other skills” with 87%, and then “Learning process” with 85%. What we also found was that even though “Learning process” had a 6% lower number of exact matches than “Start action” it only had around a 1% difference in incorrectly labeled responses. This is because most of the inaccuracy of the “Learning Process” came from the “0” class which is neither an exact match nor an incorrect classification. Even though we used the same 10% of the total data when evaluating the ZSL NLP classifier you will notice that the totals for each question differ. The highest count was “Start action” and the lowest count was “Other skills”. The reason for this difference in total counts is that some sentences that are parsed by the ZSL NLP classifier would get assigned multiple labels. We have found that the more codes we have in our codebook for each question, the more codes get assigned to a single sentence. For example, the sentence “I will start reviewing the material weekly that way I won’t forget the older material and I can master it better." will get assigned the labels "Increase mastery on topics", "Study more", and "Better study habits" which increased the total count of the labels assigned.

Examples of correctly, incorrectly, and semantically related sentences are shown in Tables 3.4, 3.6, and 3.8 which correspond to the exam wrapper questions on "Other skills", "Learning process, and "Start action" respectively. The scores for the corresponding sentences are in the "Score" column and the "Label" column is the code that the ZSL classifier identified for that particular sentence. In Tables 3.4, 3.6, and 3.8 the sentences corresponding to a score of "1" were exact matches to how the researcher would have labeled the sentences. For example in Table 3.4, the sentence "I sometimes struggle with remembering conversion factors if I do not have them written on my formula sheet" was assigned the label "Better Equation Sheet" in response to the "Other skills" question. Some sentences were labeled incorrectly and were assigned a score of "-1". An example of this is in Table 3.6 where the sentence "I take notes early and often over the module" was given the label "Pay More Attention" which is an incorrect description of that sentence. Finally, some sentences were not exact matches but had some semantic relation to the label they were assigned. For example, In Table 3.8, the sentence "I also want to start working on my equation sheet as I am learning the new equations" got assigned the label "More Time on Learning Pages" by the ZSL classifier. While this is not completely correct since the student is mainly talking about their equation sheet and the timing of that, which we would probably have assigned "Equation Sheet" or "Time Management" as a label, we do see a relationship since the learning pages could be where the equations for the equation sheet could be found, and since the label mentions time, it is associated with time management.

3.3.1.1 Other Skills

Table 3.3: Evaluation results table for zero-shot classifier of “Other skills”

Score	Count	Proportion
1	664	87%
0	62	8%
-1	38	5%
Total	764	

Table 3.4: Examples of correct, incorrect, and semantically related sentences for “Other skills”

Sentence	Label	Score
This means that I should brush up on using my calculator under stressful conditions to help minimize the number of mistakes I make on the exam.	Attention to Detail	1
I sometimes struggle with remembering conversion factors if I do not have them written on my formula sheet.	Better Equation Sheet	1
I lose around 5 points or more due to a simple mistake in calculation.	Calculator Skills	0
I get very nervous when I take exams which really impedes my performance on exams.	Lack of Focus	0
My math skills failed me on number 15 of my version of the exam, and I should have recognized that a squared x would mean there is more than one answer.	No Weak Fundamental Knowledge lacking	-1
I need to be able to manage my time better and look over the test and find the questions that I can get the most points on instead of going in sequential order.	Study Skills	-1

3.3.1.2 Learning Process

Table 3.5. Evaluation results table for zero-shot classifier of “Learning process”

Score	Count	Proportion
1	741	85%

0	100	11%
-1	34	5%
Total	875	

Table 3.6. Examples of correct, incorrect, and semantically related sentences for "Learning process"

Sentence	Label	Score
I bring questions I have about prep or practice questions to class.	Asking Questions	1
Attending every class and lab alongside my studying and prep goals from the last exam wrapper will most definitely strengthen my learning.	Attend All Classes/Labs	1
I do the learning pages and practice problems last minute.	Do More Practice Problems	0
I take good notes on the learning page and write down examples which I used to help me with the practice questions.	Get Help With Questions	0
This is so if I get stuck I can put it away and come back to it with a fresh mind.	Improve Time Management	-1
I take notes early and often over the module.	Pay More Attention	-1

3.3.1.3 Starting Action

Table 3.7. Evaluation results table for zero-shot classifier of "Start action"

Score	Count	Proportion
1	896	91%
0	55	6%
-1	33	3%
Total	984	

Table 3.8. Examples of correct, incorrect, and semantically related sentences for “Start action”

Sentence	Label	Score
I would like to start asking for help more often when I do not understand a homework problem in order to actually learn the material instead of just trying to get through the work and get the assignment done.	Ask for Help	1
I’m going to start the practice before it is discussed in class so that way I can ask questions in class, and then I would like to finish the rest of the practice that day.	Ask Questions	1
Even skimming through the practice problems could help me realise or figure out how to do the problem or what equation I should use.	Improve on Conceptual Understanding	0
I also want to start working on my equation sheet as I am learning the new equations.	More Time on Learning Pages	0
This might sound stupid, but I think I need to take a few calming breaths before I even start my exam.	Prepare for the Exam Earlier	-1
Whether I need to wake up earlier to go before class starts or stay later after my classes end, this will allow me to get more help and a better understanding of the material.	Study More	-1

3.4 Discussion

In this section, we discuss the results obtained from the study. In terms of answering the research question: “How can a state-of-the-art NLP workflow be used to analyze exam wrappers for self-regulated learning?” we found that the codebook was created in a more streamlined way than the traditional qualitative analysis as outlined by Braun & Clarke (2006). Using NLP to cluster responses allowed us to glean the main topics that students were speaking about, which gave us more access to information across the entire dataset, rather than randomly coding to saturation. We found that SRL constructs overlapped with one another, and this was expected given the cyclical nature of Zimmerman’s model of SRL. Concerning the accuracy of labeling, the NLP workflow performs at an acceptable level since we achieved accuracies between 85 -

91%. Generally, a level of 80 percent agreement indicates an acceptable level of reliability (McAlister et al., 2017).

3.4.1 Accuracy of Different Codebooks

The codebooks received a relatively high accuracy when evaluating all three questions on 10% of the dataset for each question. There are a few reasons why accuracy can differ using the zero-shot method. As an example, Pushp & Srivastava (2017) achieved an accuracy of between 60-70%, while attempting to classify Tweets into different categories. This classification task is more open-ended since the codes in their dataset varied with topics including business, health, sports technology, and entertainment. Our classifier focused explicitly on exam preparation and the questions were developed using the same theoretical framework that we used to analyze the data in Zimmerman's Cyclical Phases Model.

In addition, the way that the labels are created is important. Liu et al. (2023) refer to the classification task accuracy depending on the prompts that are used in the classification task. To generate the codes, the researchers needed to not only be knowledgeable about the context of the data but also have an idea of how the zero-shot classifier works (Gao et al., 2021). This means that implementing an NLP workflow on textual data will require domain knowledge as well as knowledge of the inner workings of the NLP classifier, thus we were able to achieve high accuracies through a few rounds of iteration informed by literature on SRL and domain knowledge in the engineering education space. Additionally, what we have observed in the results is that more general labels were shown to be accounted for more, for example, the label "Study More" was the most common in the "Start" question. This label is a general strategy that many students would reference, but studying more would also be an overarching theme that encompasses some other topics that students spoke about such as spending more time on practice questions, spending more time on learning pages, watching videos on topics, etc.

The accuracies achieved in the codebooks imply that the codebook and the ZSL NLP classifier that we have developed have a practical use for future exam wrappers given that the questions in the exam wrappers are similar to the ones we did our analysis on. This will save the time it takes for instructors to identify the most common SRL strategies that students are using while preparing for exams, as well as some of the concepts and ideas that students may struggle with.

3.4.2 Implications of the NLP Workflow in Studying SRL

We have demonstrated that the analysis of a large corpus of exam wrapper data can be done using this NLP workflow to an acceptable degree of accuracy. We have also demonstrated that the expected findings were in line with what has been found in the literature when student exam wrappers were analyzed qualitatively. While qualitative studies have identified similar themes, the distribution of these themes has not been discussed in previous work due to the time-intensive nature of the qualitative analysis. All previous qualitative studies on SRL and self-reflections in any form have a sample size of less than 200, whereas this study has around 3,800 participants. For example, Chew et al. (2016) found that students were able to reflect on their process of learning much like what we have found, but the sample of 69 students does not reflect a good representation of an entire student body in larger courses such as first-year engineering courses. Furthermore, the NLP approach offers more insights since Chew et al. (2016) mostly found instances in the reflection phase while our study found that students had improvements to make in all phases of the exam preparation process. This study, in comparison, confirms the SRL strategies found in other qualitative studies that students report using while preparing for exams, and provides more generalizability to the whole student population by showing the distribution of these strategies. These insights can point instructors in the right direction as to which areas of improvement to prioritize and which SRL strategies students must engage in while preparing for exams. For example, instructors could proactively use the exam wrappers as an intervention since the NLP Workflow will save them a lot of grading time and could provide instructors with insights immediately after the exam wrappers are turned in. This presents an opportunity for the instructor to adjust the course based on the insights gained from the exam wrapper.

The use of the NLP workflow could also lower the barrier to entry for researchers who want to use exam wrappers to gain insights about what students view as important when preparing for exams and which gaps students identify in their strategies when preparing for exams. Previously, the reason for selecting smaller samples for qualitative research was due to the time taken to analyze qualitative data, but in the case of NLP, the amount of data only affects the researcher when evaluating the accuracy of the codebook.

3.4.3 Future Research

The successful implementation of the NLP workflow on exam wrappers within a theoretical framework could open doors for numerous qualitative analyses in various fields. Specifically, in the context of SRL and exam wrappers, there is an opportunity to examine students' exam grades in conjunction with their exam wrappers to determine any correlations between SRL strategies and exam scores. Additionally, we could analyze other exam wrapper questions to uncover new findings and insights. The results from this study can be compared with subsequent research to further aid instructors in their teaching methods and provide researchers with valuable information about student SRL in first-year physics for engineering courses.

Another promising avenue is to develop a user-friendly application programming interface (API) that enables researchers and instructors to input their codebooks and responses for classification. The accuracy of this classification can be validated by manually analyzing a sample from the full dataset, saving educators and researchers time in assessing student reflections. Moreover, we could explore the accuracy of ChatGPT as a tool for analyzing large volumes of student data, providing a valuable resource for instructors with limited coding experience.

3.4.4 Limitations

Several limitations in this study warrant attention. Unlike traditional research scenarios, where a researcher creates a codebook with accompanying definitions for their team, the zero-shot classifier merely interprets codes at face value when classifying student responses. This necessitates prompt engineering, whereby labels are systematically generated based on an understanding of how the classifier interprets them (Liu et al., 2023). In future applications, we aim to investigate the use of various prompt engineering techniques to establish a systematic method for creating codebooks for classification. Second, the absence of a systematic approach to codebook creation might have led to skewed results, depending on whether the code was simple or nuanced. For instance, the top label “Study more” for “Start” questions might have more counts due to the various ways one can study more. In contrast, a specific label like “Make equation sheet during the module,” which is more detailed and informative for instructors advising students, maybe more challenging for the zero-shot classifier to detect, resulting in potential inaccuracies. Third, there is the potential for bias in the training data. The zero-shot classifier relies on pre-training based on web sources such as Wikipedia, which could contain bias that might inadvertently transfer to the model’s latent representations and subsequent classification tasks. Future research could examine exam wrappers or other reflective data to identify any biases originating from the zero-shot classifier. During the analysis, no noticeable bias was found in the data.

3.5 Conclusion

In answering the research question: "How can a state-of-the-art NLP workflow be used to analyze exam wrappers for self-regulated learning?" we were able to accurately analyze the three exam wrappers question responses of 3800 students over three semesters of an engineering physics course for SRL constructs. We achieved accuracies of 91%, 87%, and 85% when evaluating 10% zero-shot NLP classified responses. This process would have taken multiple researchers tens of hours to do while most of the analysis was mainly by one researcher in a few hours. Furthermore, once the initial setup phase of codebook generation is completed, analyzing new exam wrapper data for entire cohorts could be done in less than an hour.

The ZSL NLP workflow offers a useful way for faculty, administrators, and researchers to draw useful insights from large corpora of qualitative data. Since we were able to draw insights from all the students in each cohort, instead of a small sample as in traditional qualitative studies, we believe that these insights are more generalizable to larger populations.

The implications of this work hold considerable potential for improving the understanding and assessment of SRL in academic settings. By utilizing the ZSL NLP workflow, this study demonstrates that it is possible to efficiently and effectively analyze vast amounts of qualitative data from student exam wrappers, which can lead to the following outcomes. First, this could help lead to enhanced pedagogical strategies. By gaining a deeper understanding of students' SRL strategies, instructors can tailor their teaching methods to better support student learning. This could involve offering more targeted guidance, designing activities to foster specific SRL skills, or providing individualized feedback to students. Second, the work could support informed administrative decisions. Administrators can use the insights derived from large-scale qualitative data analysis to make evidence-based decisions regarding curriculum development, student support services, and resource allocation. This could lead to more effective and student-centered educational programs. Third, there is broader applicability in research. In particular, the NLP workflow presented in this study can be adapted and applied to other domains within education or extended to different fields that involve the analysis of qualitative data. This versatility can facilitate the exploration of new research questions and foster interdisciplinary collaboration. Fourth, along the same lines, there is further empowerment of researchers. The efficient analysis of large corpora of qualitative data allows researchers to conduct more comprehensive studies, drawing conclusions from entire cohorts instead of relying solely on small samples. This, in turn, enhances the generalizability and reliability of research findings, leading to more robust insights that can inform future studies.

Overall, the successful application of the NLP workflow in analyzing exam wrappers for SRL constructs has the potential to significantly impact educational research, teaching practices, and administrative decision-making. By enabling the efficient analysis of large-scale qualitative data, this approach can contribute to a better understanding of student learning processes, ultimately leading to improved educational outcomes. While modern LLMs such as ChatGPT will enable similar analysis, this approach still merits consideration for its comparatively low overhead and ability to run locally, thereby mitigating privacy concerns associated with sending student data to a third party.

References

Arbogast, C. A., & Montfort, D. (2016). Applying natural language processing techniques to an assessment of student conceptual understanding. *2016 ASEE Annual Conference & Exposition*.

Becker, J. P., Kahanda, I., & Kazi, N. H. (2021, July 26). *WIP: Detection of Student Misconceptions of Electrical Circuit Concepts in a Short Answer Question Using NLP*. 2021 ASEE Virtual Annual Conference Content Access. https://peer.asee.org/wip-detection-of-student-misconceptions-of-electrical-circuit-concepts-in-a-short-answer-question-using-nlp

Berdanier, C. G. P., Baker, E., Wang, W., & McComb, C. (2018). Opportunities for Natural Language Processing in Qualitative Engineering Education Research: Two Examples. *2018 IEEE Frontiers in Education Conference (FIE)*, 1–6. https://doi.org/10.1109/FIE.2018.8658747

Bhaduri, S. (2018). *NLP in Engineering Education-Demonstrating the use of Natural Language Processing Techniques for Use in Engineering Education Classrooms and Research* [PhD Thesis]. Virginia Tech.

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, *31*(6), 445–457. https://doi.org/10.1016/S0883-0355(99)00014-2

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, *65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Butt, A. A., Anwar, S., & Menekse, M. (2022). WIP: Role of digital nudging strategies on STEM students' application engagement. *2022 ASEE Annual Conference & Exposition*.

Carpenter, T. S., Beall, L. C., & Hodges, L. C. (2020). *Using the LMS for Exam Wrapper Feedback to Prompt Metacognitive Awareness in Large Courses*. https://doi.org/10.14434/jotlt.v9i1.29156

Castellanos, M., & Enszer, J. A. (2013). *Promoting Metacognition through Reflection Exercises in a Thermodynamics Course*. 23.999.1-23.999.13. https://peer.asee.org/promoting-metacognition-through-reflection-exercises-in-a-thermodynamics-course

Chew, K. J., Chen, H. L., Rieken, B., Turpin, A., & Sheppard, S. (2016). Improving students' learning in statics skills: Using homework and exam wrappers to strengthen self-regulated learning. *123rd ASEE Annual Conference and Exposition, June 26, 2016 - June 29, 2016*, *2016-June*.

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the Conceptual Lens on Metacognition, Self-regulation, and Self-regulated Learning. *Educational Psychology Review*, *20*(4), 391–409. https://doi.org/10.1007/s10648-008-9083-6

Ellestad, R. M. (2016). The Impact of Teaching Self-Regulated Learning Skills to First Year Engineering Students. *2016 ASEE Annual Conference & Exposition*.

Gamieldien, Y., Case, J. M., & Katz, A. (2023). *Advancing Qualitative Analysis: An Exploration of the Potential of Generative AI and NLP in Thematic Coding* (SSRN Scholarly Paper 4487768). https://doi.org/10.2139/ssrn.4487768

Ganesh, A., Scribner, H., Singh, J., Goodman, K., Hertzberg, J., & Kann, K. (2022). Response Construct Tagging: NLP-Aided Assessment for Engineering Education. *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*, 250–261. https://aclanthology.org/2022.bea-1.29

Gao, T., Fisch, A., & Chen, D. (2021). *Making Pre-trained Language Models Better Few-shot Learners* (arXiv:2012.15723). arXiv. http://arxiv.org/abs/2012.15723

Gynnild, V., Holstad, A., & Myrhaug, D. (2008). Identifying and promoting self-regulated learning in higher education: Roles and responsibilities of student tutors. *Mentoring & Tutoring: Partnership in Learning*, *16*(2), 147–161. https://doi.org/10.1080/13611260801916317

Haller, S., Aldea, A., Seifert, C., & Strisciuglio, N. (2022). *Survey on Automated Short Answer Grading with Deep Learning: From Word Embeddings to Transformers* (arXiv:2204.03503). arXiv. https://doi.org/10.48550/arXiv.2204.03503

Katz, A., Norris, M., Alsharif, A. M., Klopfer, M. D., Knight, D. B., & Grohs, J. R. (2021, July 26). *Using Natural Language Processing to Facilitate Student Feedback Analysis*. 2021 ASEE Virtual Annual Conference Content Access. https://peer.asee.org/using-natural-language-processing-to-facilitate-student-feedback-analysis

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-022-13428-4

Lakretz, Y., Dehaene, S., & King, J.-R. (2020). What Limits Our Capacity to Process Nested Long-Range Dependencies in Sentence Comprehension? *Entropy*, *22*(4), Article 4. https://doi.org/10.3390/e22040446

Lawanto, O., Santoso, H., Lawanto, K., & Goodridge, W. (2014). Self-Regulated Learning Skills and Online Activities Between Higher and Lower Performers on a Web Intensive Undergraduate Engineering Course. *The Journal of Educators Online*, *11*(3). https://doi.org/10.9743/JEO.2014.3.2

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, *55*(9), 195:1-195:35. https://doi.org/10.1145/3560815

Lovett, M. C. (2013). Make exams worth more than the grade: Using exam wrappers to promote metacognition. In: Kaplan, M., Silver, N., LaVague-Manty, D., Meizlish, D. (eds) Using reflection and metacognition to improve student learning: Across the disciplines, across the academy,. In *Using reflection and metacognition to improve student learning: Across the disciplines, across the academy*, (pp. 18–52).

Magooda, A., Litman, D., Ashraf, A., & Menekse, M. (2022). Improving the Quality of Students' Written Reflections Using Natural Language Processing: Model Design and Classroom Evaluation. In M. M. Rodrigo, N. Matsuda, A. I. Cristea, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (pp. 519–525). Springer International Publishing. https://doi.org/10.1007/978-3-031-11644-5_43

McAlister, A. M., Lee, D. M., Ehlert, K. M., Kajfez, R. L., Faber, C. J., & Kennedy, M. S. (2017, June 24). *Qualitative Coding: An Approach to Assess Inter-Rater Reliability*. 2017 ASEE Annual Conference & Exposition. https://peer.asee.org/qualitative-coding-an-approach-to-assess-inter-rater-reliability

Menekse, M. (2020). The Reflection-Informed Learning and Instruction to Improve Students' Academic Success in Undergraduate Classrooms. *Journal of Experimental Education*, *88*(2), 183–199. https://doi.org/10.1080/00220973.2019.1620159

Mervis, J. (2013). *Transformation is possible if a university really cares*. American Association for the Advancement of Science.

Nehyba, J., & Štefánik, M. (2022). Applications of deep language models for reflective writings. *Education and Information Technologies*, 1–39.

Nelson, K. G., Shell, D. F., Husman, J., Fishman, E. J., & Soh, L.-K. (2015). Motivational and Self-Regulated Learning Profiles of Students Taking a Foundational Engineering Course. *Journal of Engineering Education*, *104*(1), 74–100. https://doi.org/10.1002/jee.20066

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: When measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research*, *60*(6), 723–735. https://doi.org/10.1080/00313831.2015.1066436

Pintrich, P. R. (2000). The Role of Goal Orientation in Self-Regulated Learning. In *Handbook of Self-Regulation* (pp. 451–502). Elsevier. https://doi.org/10.1016/B978-012109890-2/50043-3

Pushp, P. K., & Srivastava, M. M. (2017). *Train Once, Test Anywhere: Zero-Shot Learning for Text Classification* (arXiv:1712.05972). arXiv. https://doi.org/10.48550/arXiv.1712.05972

Schunk, D. H. (2008). Metacognition, Self-Regulation, and Self-Regulated Learning: Research Recommendations. *Educational Psychology Review*, *20*(4), 463–467. https://doi.org/10.1007/s10648-008-9086-3

Schunk, D. H., & Zimmerman, B. (2011). *Handbook of Self-Regulation of Learning and Performance*. Taylor & Francis.

Shell, D. F., Hazley, M. P., Soh, L.-K., Ingraham, E., & Ramsay, S. (2013). Associations of students' creativity, motivation, and self-regulation with learning and achievement in college computer science courses. *2013 IEEE Frontiers in Education Conference (FIE)*, 1637–1643. https://doi.org/10.1109/FIE.2013.6685116

Tuckman, B. W., & Kennedy, G. J. (2011). Teaching learning strategies to increase success of first-term college students. *Journal of Experimental Education*, *79*, 478–504. https://doi.org/10.1080/00220973.2010.512318

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Vosniadou, S. (2020). Bridging Secondary and Higher Education. The Importance of Self-regulated Learning. *European Review*, *28*(S1), S94–S103. https://doi.org/10.1017/S1062798720000939

Wallin, P., & Adawi, T. (2018). The reflective diary as a method for the formative assessment of self-regulated learning. *European Journal of Engineering Education*, *43*(4), 507–521. https://doi.org/10.1080/03043797.2017.1290585

Wang, R., Wei, S., Ohland, M., & Ferguson, D. (2019, October 25). *Natural language processing system for self-reflection and peer-evaluation*.

Wedelin, D., Adawi, T., Jahan, T., & Andersson, S. (2015). Investigating and developing engineering students' mathematical modelling and problem-solving skills. *European Journal of Engineering Education*, *40*(5), 557–572. https://doi.org/10.1080/03043797.2014.987648

Wulff, P., Mientus, L., Nowak, A., & Borowski, A. (2022). Utilizing a Pretrained Language Model (BERT) to Classify Preservice Physics Teachers' Written Reflections. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-022-00290-6

Yin, W., Hay, J., & Roth, D. (2019). *Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach* (arXiv:1909.00161). arXiv. https://doi.org/10.48550/arXiv.1909.00161

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology*, *11*(4), 307–313. https://doi.org/10.1016/0361-476X(86)90027-5

Zimmerman, B. J. (2000). Chapter 2 - Attaining Self-Regulation: A Social Cognitive Perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J., & Schunk, D. H. (Eds.). (1989). *Self-Regulated Learning and Academic Achievement*. Springer New York. https://doi.org/10.1007/978-1-4612-3618-4

Chapter 4: Manuscript 3

Understanding Performance Profiles through Self-Regulated Learning Constructs in Engineering Physics: A Large Language Model Approach

This manuscript includes intellectual contributions from Rachel McCord Ellestad and Andrew Katz

4.1 Structured Abstract

Background

Self-regulated learning (SRL) strategies have proven instrumental in academic success. However, their implementation varies among students, and there is potential to study this variation at scale using large language models (LLMs) within the context of a first-year engineering physics course.

Purpose

This study aims to explore and understand the differences in strategies and SRL constructs reported by students across different performance profiles in an engineering physics course.

Design/Method

Based on exam performance, students were clustered into four distinct performance profiles—Low, Average, High, and Improvers. Subsequently, we used a text classification natural language processing (NLP) technique called zero-shot learning (ZSL) to categorize students' responses to two exam wrapper questions after each of their three exams. These questions asked about non-physics skills that students lacked that impeded their ability to perform well in the module exam and what preparation activities they did before the exam as well as improvement activities for the next exam. The ZSL model allowed us to identify diverse learning strategies and SRL constructs among the four performance groups.

Results

The analysis revealed a broad spectrum of study strategies and an inconsistent emphasis on SRL constructs across the profiles. Task-oriented strategies, such as practicing past exams, enhancing conceptual understanding, and creating equation sheets, were frequently highlighted, and non-physics skills such as lack of focus, test-taking skills, and time management were highly discussed, while SRL constructs like Environmental Structuring, Help-Seeking, and Self-Efficacy received less focus.

Conclusions

Despite limitations like the selectivity of self-reported data and potential loss of nuance in ZSL analysis, the study underscores the dynamic nature of learning strategies and varied application of SRL constructs among differing performance profiles. This work also extends our understanding of SRL in engineering education and offers valuable implications for pedagogical practices, curriculum design, and future research.

Keywords

self-regulated learning; natural language processing; exam wrappers; large language models

4.2 Introduction

Understanding SRL is integral to advancing education because of its association with academic achievement (Bergin et al., 2005; Nota et al., 2004; Paechter et al., 2010; Tilfarlioglu & Delbesoglugil, 2014). Furthermore, SRL provides a lens through which we can examine how students become masters of their learning processes (Zimmerman, 2002) and how these processes can differ for different students (Nelson et al., 2015). In the field of engineering education, the importance of SRL is amplified when compared to other fields due to complex problem-solving (Jonassen et al., 2006), continual learning (Adams & Felder, 2008), and team collaboration (Chowdhury & Murzi, 2019). Engineering students are required to tackle complex problems, which necessitates a strategic approach to learning and a high degree of self-regulation (Ellestad, 2016; Falkner et al., 2014; Kosnin, 2007; Lawanto et al., 2014; Wallin & Adawi, 2018).

Exam preparation is a critical component of the learning process in engineering education. It is here where students not only revisit the technical knowledge they have accumulated but also apply non-technical skills, such as time management, organization, and metacognitive strategies (Dembo & Seli, 2008). A helpful tool in promoting SRL is the use of exam wrappers - metacognitive exercises that encourage students to reflect on their study strategies and exam performance (Gezer-Templeton et al., 2017; Lovett, 2013). Exam wrappers are mutually beneficial to students and instructors in enhancing the teaching and learning experience (Chew et al., 2016). While valuable, the grading of these wrappers poses a challenge. Given the typically large class sizes in many institutions, assessing student responses becomes a logistical issue.

Enter the potential of LLMs and NLP. These innovative technologies, including ZSL, offer the ability to analyze large amounts of text data, similar to those generated by exam wrappers, in a scalable and efficient way (Devlin et al., 2019; Radford et al., 2021; Yin et al., 2019). The objective of this study is to leverage a ZSL text classification model to analyze the SRL strategies employed by engineering students of differing exam performance profiles. The guiding research question is: "To what extent do the SRL strategies reported by students of various performance profiles, as captured in their exam wrapper responses, differ while preparing for exams?" By answering this question, we aim to augment the existing body of knowledge on SRL in engineering education and derive practical implications for both pedagogy and curriculum design (Schunk & Zimmerman, 2011). The ultimate goal is to enhance the educational experience and boost academic performance within the field of engineering.

The structure of this paper is as follows: First, we review relevant literature focused on our theoretical framework of SRL, as well as on the topics of exam wrappers, NLP, and LLMs within the context of education. We will then describe the specifics of our study, including our data collection and analysis steps for investigating students' SRL strategies across various performance profiles. Subsequently, we will present our findings and discuss results by connecting them to existing literature. The paper concludes by identifying the limitations of our study, elucidating the implications of our research, and suggesting potential avenues for future exploration.

4.3.1 Theoretical Framework – Self-Regulated Learning

Since the field of SRL is large, and since engineering education is an emerging field (Borrego & Bernhard, 2011; Nager & Atkinson, 2016), we sometimes draw on literature from outside these fields to provide an overview of SRL (Panadero, 2017; Silverajah et al., 2022).

Recognizing that SRL has been defined in several ways in the literature, we define SRL based on Zimmerman (2000) since we wanted to align the definition and framework that we used in the study. Furthermore, the exam wrappers we describe later were designed by the course instructor through the lens of Zimmerman's model. Zimmerman (2000) proposed a cyclical model of self-regulation and related it to learning. According to Zimmerman, SRL is defined as: *"self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals."*(Zimmerman, 2000, p. 14).

The model, shown in Figure 4.1, cyclically consists of three phases: the Forethought Phase, the Performance Phase, and the Self-Reflection Phase. The Forethought Phase incorporates strategic planning and goal-setting while combining affective aspects which can be likened to self-regulation. The Performance Phase consists of metacognitive skills such as monitoring, as well as controlling. Finally, the Self-Reflection Phase incorporates the metacognitive elements of evaluation with emotional elements of self-satisfaction.

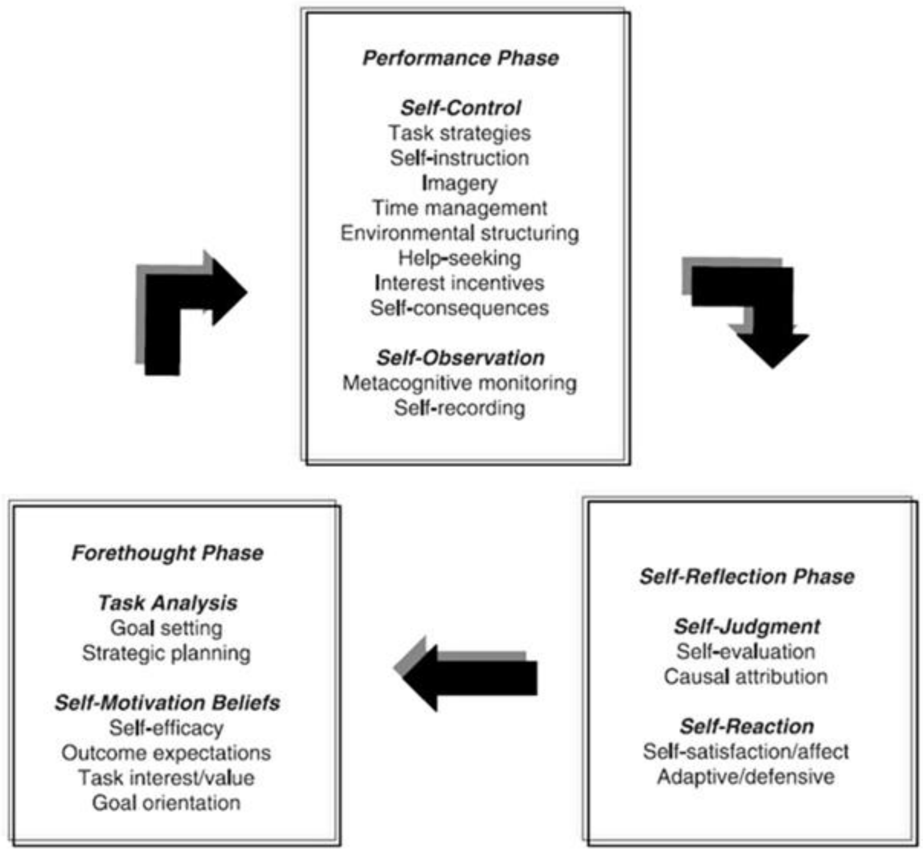


Figure 4.1: Phases and processes of self-regulation (Zimmerman & Moylan, 2009)

4.3.1.1 Forethought phase - Task analysis

Task analysis in the Forethought Phase can be broken up into two sub-processes: goal setting and strategic planning. *Goal setting* refers to deciding upon a specific outcome of learning or performance, such as "being able to achieve an A on the module exam". Schunk (1990) described goals as having properties that affect behavior, namely: specificity, proximity, and difficulty level. Goals that have specific performance standards are more likely to enhance learning and activate self-evaluations than general goals. For example, a statement such as "I want to do well in this course" is a general goal, whereas a more specific goal would be "I want to be able to represent complex systems at the end of this section of the course". In a study on goal proximity, Bandura & Schunk (1981) gave a group of children with low subtraction skills instructions to practice over seven sessions. There was a group with the proximal goal of completing one set of questions each session and a group was given a distal goal of completing all sets of questions by the end of the last session. They found that the group with the proximal goals had the highest subtraction skills and self-efficacy. Self-efficacy is a self-motivation belief that will be described in detail in the next section. The goal systems for highly self-regulated individuals are organized hierarchically, such that process goals operate as regulators for the more distal outcome goals (Zimmerman, 2000). These process goals are more than just checkpoints, they become invested with personal meaning because they convey evidence of progress. Process goals are less abstract than outcome goals and since measuring progress is easier it leads to greater levels of self-efficacy (Zimmerman & Campillo, 2003). Additionally, in a study in a large economics class, Amann & Rzepka (2023) found that students who were sent goal-setting prompts with bi-weekly quizzes outperformed the control who did not receive these prompts. The students who received these prompts were 18.8% more likely to pass the exam and earned 6.7% more points on the exam indicating that having process goals throughout a course can be more effective than having set no goals or setting general goals.

Strategic planning is a form of task analysis that requires students to have appropriate methods for the task and the setting. Appropriately selected strategies enhance performance by aiding in cognition, controlling effect, and directing motoric execution (Pressley, 1990). The planning and selection of strategies require cyclical adjustments because of the fluctuation of the individual. For example, making a study timetable can be a simple way to strategically plan to prepare for an exam, but a more advanced self-regulated learner would focus on specific tasks that will be carried out during those study sessions instead of generally setting aside time to study. Students should also link their strategic plans to the proximal and distant goals they have set, which will lead to them practicing effectively by themselves for longer periods by increasing motivation (Zimmerman & Moylan, 2009). Thus, self-regulated individuals must continuously adjust their goals and choices of strategies (Zimmerman, 2000). The planning and selection of strategies in exam preparation require ongoing adjustments due to cyclical feedback from earlier efforts because no self-regulatory strategy will work optimally in all contexts. It is therefore important for students to be highly self-regulatory to be successful (Zimmerman & Campillo, 2003).

4.3.1.2 Forethought phase - Self-motivation beliefs

Motivation is a key component of SRL according to prominent researchers in this area (Boekaerts, 1999; Pintrich, 2000; Schunk, 2008; Zimmerman, 1995). Definitions of self-regulation and close synonyms such as self-control and self-discipline are related to the control of one's present conduct and motivation (Schunk & Zimmerman, 2008). Highly motivated students are more attentive to their learning processes and outcomes than poorly motivated students (Bouffard-Bouchard et al., 1991). A student can be taught certain SRL skills, but if they are not attentive to their learning processes and the feedback they receive, they are unlikely to act on what they have learned. As shown in Figure 4.1, *self-motivation beliefs* can be categorized into multiple subprocesses that affect a student's motivation: self-efficacy, outcomes expectation, task value/interest, and goal orientation (Zimmerman & Moylan, 2009).

Bandura & Wessels (1994) defined perceived self-efficacy as people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives. *Outcome expectation* is a source of motivation that refers to the beliefs about the ultimate ends of performance (Bandura, 1991). To illustrate the interrelatedness of self-efficacy and outcome expectation the example of studying for an exam can be used. Having high self-efficacy would mean that you are confident about achieving a good grade on the exam, whereas the outcome expectation is perhaps the belief that you will be known as “smart” if you achieve a high grade on the exam. *Task interest/value* is an internal motivation construct that is linked to the individual’s enjoyment of the task in itself regardless of the outcome (Zimmerman & Moylan, 2009). The final self-motivation source is goal orientation, broken down into competency and performance goals. *Goal orientation* refers to whether the student wants to develop competence or optimize short-term performance. With a competency goal, students are focused on developing new skills, trying to understand their work, improving their level of competence, and achieving a sense of mastery. In contrast, a performance goal is concerned with being able in comparison to others and receiving public recognition for that ability (Zimmerman & Campillo, 2003). In the context of engineering education, Nelson et al. (2015) found that students who adopted negative motivational and SRL profiles indicated minimum effectiveness in learning the course content, maladaptive goal orientation to avoid learning course material, and a lack of SRL behaviors.

4.3.1.3 Performance phase – Self-control and self-observation

The second phase of SRL according to Zimmerman’s model is the Performance Phase. The Performance Phase can be broken down into two major classes: self-control and self-observation. Students’ use of self-control methods involves a variety of different task-specific and general strategies for learning (Zimmerman & Moylan, 2009). Task strategies refer to the developing of a systematic process for addressing specific components of a task. For example, if reworking homework problems or past exam questions when studying for an exam. Self-control strategies include self-instruction, imagery, time management, environmental structuring, help-seeking, interest incentives, and self-consequences (Zimmerman & Moylan, 2009).

Self-control processes such as self-instruction, imagery, attention focusing, and task strategies help learners and performers to focus on the physical task and optimize their solution effort (Zimmerman & Campillo, 2003). Self-instruction refers to the overt or covert descriptions of how to proceed as one executes a task. Examples of these are when a student questions themselves as they read class material (Zimmerman & Moylan, 2009). Imagery is a self-control strategy that involves forming pictures from textual data, examples of imagery are concept maps, flow diagrams, and animations (Zimmerman & Moylan, 2009). Time management involves prioritizing and planning learning strategies, as well as study strategies to achieve one’s goals. Prior research suggests that effective time management may improve students’ academic self-efficacy (Eissa, 2015). Environmental structuring is a self-control method for increasing the effectiveness of one’s immediate environment, and this alludes to the importance of the environment in which the learning takes place. This does not refer exclusively to the physical environment. An example of this would be to use a citation manager to keep your references in order instead of doing the process manually which will save time and resources in the writing process.

Help-seeking involves soliciting assistance when learning or performing, such as asking a peer to coach you through a calculation in an engineering classroom (Zimmerman & Moylan, 2009). From these two elements in the performance phase, we can see that there are social and motivational elements to becoming a self-regulated learner who is successful, as well as more of an emphasis on the context and environment in which the learner is. In an overview of his work, Zimmerman described his views on SRL as a socio-cognitive approach to SRL which further emphasizes that SRL operates on both the individual and social level, especially considering motivational factors.

4.3.1.4 Self-Reflection Phase - Self-judgment

The third and final phase of SRL in Zimmerman's model is the Self-Reflection phase. This phase has the categories of self-judgments and self-reactions. There are different standards that students may evaluate themselves on which could be based on prior levels of performance, social comparisons with the performance of others, and mastery of all components of a skill (Bandura, 1986). The self-reaction element has two categories that deal with behavior and affect: self-satisfaction and adaptive/defensive decisions. Self-satisfaction is simply defined as the cognitive and affective reactions to one's self-judgments (Zimmerman & Moylan, 2009). For example, students would be more likely to participate in activities that have led to satisfaction and positive affect, and they tend to avoid activities that produce dissatisfaction and negative affect. Adaptive reactions refer to students modifying strategies or continuing to use a strategy whereas defensive reactions take place when students feel helpless, procrastinate, avoid tasks, or feel apathy. Self-evaluation is the student's assessment of their performance based on the assessment criteria and modulated by their performance level goal (Panadero & Alonso-Tapia, 2014). Self-evaluation is the comparison of self-observed performances against some standard, such as goals set, prior performance, peer performance, or an absolute standard of performance.

Another form of self-judgment involves causal attribution which refers to beliefs about the cause of a student's errors or successes. For example, if a student believes that the limitation of their performance on a task is due to a fixed ability, this can be troublesome motivationally since it implies that any efforts to improve in future tasks will not be effective. Attributing the causes of poor performance to controllable processes will sustain motivation because it implies that other strategies could lead to success (Zimmerman, 2002). Task analysis also affects attribution judgments while studying for an exam. People who plan to use a specific strategy in the forethought phase and implement its use during the performance phase are more likely to attribute failures to that strategy than to low ability (Zimmerman & Campillo, 2003).

4.3.1.5 Self-Reflection phase - Self-reaction

Adaptive/defensive reactions, which were briefly mentioned earlier, take the form of adaptive or defensive responses. Defensive reactions refer to avoiding or protecting one's affect, for example, withdrawing from a course or avoiding challenging tasks that could enhance learning. In contrast, adaptive reactions refer to adjustments designed to increase the effectiveness of one's method of learning, such as discarding or modifying an ineffective learning strategy (Zimmerman, 2002). The aim of using this construct is to represent how students evaluate the effectiveness of their strategies and how they might modify strategies based on what they have already done to achieve better results on something such as exams. On the other hand, students may also talk about defensive decisions such as avoiding a task for fear of experiencing new failures (Panadero & Alonso-Tapia, 2014)**.** As stated above, Zimmerman's model will be used as the theoretical framework for this study. Table 4.1 summarizes the subprocesses in Zimmerman's model with definitions for the key subprocesses of SRL.

Table 4.1: Summary of Zimmerman's cyclical phases model with definitions of key subprocesses (Zimmerman & Moylan, 2009)

Phase	**Construct**	**Definition**
Forethought	Goal setting	Specific outcomes that the learner expects to attain.
	Strategic Planning	Choosing or constructing advantageous learning methods that are appropriate for the task and environment setting.
	Self-efficacy	Beliefs about one's capabilities to learn or perform at the desired level.

	Outcome expectation	Beliefs about the ultimate ends of one's performance.
	Task interest/value	Liking or disliking a task because of its inherent properties
	Goal orientation	Beliefs or feelings about the purpose of learning
Performance	Task strategies	A systematic process for addressing specific components of a task
	Time management	Strategies for accomplishing learning tasks on schedule.
	Environmental structuring	Increasing effectiveness of one's immediate environment.
	Help-seeking	Soliciting assistance when learning or performing a task.
	Interest incentives	Making mundane tasks more attractive to do.
Self-reflection	Self-evaluation	Comparing one's performance to a standard.
	Causal attribution	Beliefs about the causal implication of personal outcomes.
	Self-satisfaction/affect	Cognitive and affective reactions to self-judgments.
	Adaptive/defensive	Willingness to engage in further cycles of learning by continuing or modifying a strategy. In contrast, defensive decisions avoid further efforts to learn.

4.3.2 Interventions for SRL

Self-regulated learning is a skill that can be developed, and to do so we need to have interventions that foster SRL in students (Pintrich, 1995). Research suggests that self-regulatory processes can lead to increases in students' motivation and achievement (Schunk & Zimmerman, 1998). In this next section, we review several interventions that have been used to foster SRL in students. These types of interventions include the role of the instructor in promoting SRL, integrating SRL into the course through assignments and activities, reflective writing exercises, and exam wrappers. Since the main focus of this dissertation is exam wrappers, the literature reviewed on other interventions for SRL will not be as detailed and aims to give an overview of these interventions.

4.3.2.1 The Role of Instructors in Promoting SRL

In the study of SRL, instructors play a significant role. For instance, Harding (2018), who studied the influence of teachers on students' use of SRL strategies stressed the need for educators to foster a deliberate practice of SRL. Perry et al. (2008) explored a similar line of research by exploring how novice teachers could be mentored to foster SRL, revealing how post-observation discussions can be harnessed for this purpose. Other studies have examined dedicated courses aimed at enhancing students' SRL skills, with both Hofer & Yu (2003) and Ellestad (2016) reporting positive outcomes, such as increased self-efficacy, mastery orientation to learning, organization, and effort regulation. Despite these gains, Ellestad (2016) also observed increased test anxiety, suggesting the need for tailored anxiety management strategies. Lastly, Gynnild et al. (2008) examined the role of tutors in fostering SRL, by suggesting that trained tutors can be instrumental in enhancing students' SRL by offering pedagogical support. These studies collectively underscore the importance of an integrated approach involving instructors, mentors, and tutors in promoting SRL among students (Harding, 2018; Perry et al., 2008; Hofer & Yu, 2003; Ellestad, 2016; Gynnild et al., 2008).

4.3.2.2 Integrating SRL Strategies Into the Course Content

The integration of SRL strategies into the course content is associated with fostering better learning outcomes (Pintrich, 1995; Zumbrunn et al., 2011). This can be achieved through targeted interventions that prompt students to think metacognitively as they are engaging with course content. Bui et al. (2021) highlighted the efficacy of a technology-integrated project-based learning (PBL) intervention, RealLabs, which improved students' SRL skills, such as goal-setting, task strategies, time management, and self-evaluation. Pedrosa et al. (2017) explored a similar approach, the SimProgramming method, which promoted active participation and meaningful engagement, aiming to enhance SRL strategies like information search, work reviews, and time management among students transitioning from entry-level to advanced programming. Lastly, Habib et al. (2021) replicated Zimmerman's SRL model in a 400-level Civil Engineering course by integrating the model's phases into the tasks in the course. First, the students set goals and analyzed the main features of the task. The next step was to independently implement the task by themselves, and the final step was to submit a self-reflection. The researchers found that this implementation had a positive impact on the student's awareness and understanding of task requirements. These studies collectively suggest that course design and teaching methods that emphasize SRL strategies can effectively enhance student learning and performance (Bui et al., 2021; Pedrosa et al., 2017; Habib et al., 2021).

4.3.2.3 Reflective Writing and Exam Wrappers

Menekse (2020) examined the role of reflection-informed learning and instruction (RILI) using the CourseMIRROR mobile system in a statistics class. The CourseMIRROR application, which utilizes NLP to summarize reflective inputs, showed a positive correlation between the quantity and quality of students' reflections and their academic success, as measured by exam results. Another form of reflective writing is the idea of the reflective diary. Wallin & Adawi (2018) used reflective diaries in a master's level course. These diaries, utilized in a tissue engineering course at the Chalmers University of Technology, were used as a tool for fostering SRL. The reflective diary prompts proved promising for promoting students' SRL since the authors found that the diaries provided the students and instructors with insights into students' beliefs about learning, and strategies for monitoring and regulating learning.

Exam wrappers represent another form of reflective practice designed to promote SRL. These assessments, which ask students about their preparation for an exam, the errors they made, and what they plan to do differently next time, are used to encourage self-reflection and growth (Lovett, 2013). In their book, *Creating Self-Regulated Learners*, Nilson & Zimmerman (2013) also refer to a similar intervention called posttest analysis, a test autopsy, or a test postmortem. The post-test analysis requests students to make predictions about their grades on the exams, the amount of time spent on studying, and the strategies used. The authors also provide a template that students can use to identify test or exam items where they lost points, how many points they lost, and why they lost them. According to the authors, four common reasons for difficulties on exams are: carelessness (lack of concentration, rushing), unfamiliar material (what the student failed to study), misinterpreted questions (misreading or overcomplicating the questions), or not completing (due to poor time management or inadequate reading skills).

In a study by Chew et al. (2016), the use of exams and homework wrappers in a statics course found promising results in the implementation of homework and exam wrappers in a statics course. The findings from their study suggest that exam and homework wrappers can positively influence student confidence in statics concepts and also decrease the types of mistakes identified in their assignments. Liao et al. (2018) also found that the use of exam wrappers in a statics and dynamics course could improve exam performance through increased self-reflection. The authors reported that a third of the students who adjusted their study habits from the first exam to the second exam reported that they improved their exam performance.

However, the direct impact of exam wrappers on course performance is not always conclusive. A study conducted by El Bojairami et al. (2019) across five engineering classrooms found varied results when implementing exam wrappers in those courses. The authors found varying results in the association between exam grades and the quality of the exam wrappers, possibly due to differences in course types and the way exam wrapper exercises were delivered. Grandoit et al. (2020) also found that while exam wrappers could promote critical self-reflection, this did not necessarily lead to improved course performance. The authors state that while students may report critically on their learning activities for exams, it did not necessarily lead to improvement in course performance. However, the authors did find that exam wrappers did provide useful insights into students' reported SRL strategies that were lacking in their exam preparation activities. Despite this, exam wrappers still provided valuable insights into students' SRL strategies.

Large-scale studies have also been conducted to identify any associations between wrappers and exam performance exam wrappers. Carpenter et al. (2020) found a significant association between course grades and exam wrapper completion in a study involving 800 students. Hodges et al. (2020) also found a modest statistically significant relationship between exam wrapper use and course grades in a study involving over 1,100 students. Interestingly, the latter study also found that higher-performing students tended to make more use of optional exam wrappers than their lower-performing counterparts.

In summary, while the direct impact of exam wrappers on course performance is not conclusive, they do provide valuable insights into students' SRL strategies and foster awareness that can lead to greater confidence and boost academic achievement. There is also an emerging opportunity to automate exam wrapper written response summarization through NLP, similar to what was done by Menekse (2020). Building on the insights in this section, the next section will look into the usage of NLP in previous studies, thus demonstrating its potential in analyzing student exam wrappers to enhance our understanding of SRL and potentially unlocking new avenues of improving academic achievement and instruction.

4.3.3 Natural Language Processing

Research by Carpenter et al. (2020) and Hodges et al. (2020) utilized multiple-choice exam wrappers for student self-reflection in large classrooms, leaving an opportunity for open-ended reflection analysis NLP. Natural language processing offers the potential to identify common SRL constructs in student exam preparation, potentially influencing pedagogical strategies (Khurana et al., 2022).

Natural language processing has found applications in engineering education, including automatic essay scoring (AES), automatic short answer grading (ASAG; Haller et al., 2022), automatic question generation (AQG; Tsai et al., 2021), and qualitative data analysis (Anakok et al., 2022; Gamieldien, et al., 2023; Katz et al., 2023, 2021). It has been used in sentiment analysis to assess student feedback (Ganesh et al., 2022) and in metacognitive development evaluation (Bhaduri, 2018). Studies such as Cunningham et al. (2017) and Autrey et al. (2017) have used NLP techniques like term frequency-inverse document frequency (TF-IDF) to analyze student responses. However, these approaches focus on individual word usage and lack semantic context (Zhang et al., 2005). More sophisticated models like Latent Dirichlet Allocation (LDA) and Word2Vec (Mikolov et al., 2013) provide an improvement over individual word analysis (Blei et al., 2003). De Lin et al. (2021) showed the superiority of Word2Vec with K-means clustering in analyzing student self-reflections, outperforming LDA models. Yet, these models struggle with long-range dependencies in sentence structures, a problem addressed by neural network-based transformer models (Haller et al., 2022; Lakretz et al., 2020; Vaswani et al., 2017).

In various studies, transformer-based models have demonstrated their efficacy in analyzing reflective writing. For instance, Wulff et al. (2022) utilized a pre-trained BERT model to classify preservice teachers' written reflections, surpassing other deep learning and word-based algorithm performances. Wang et al. (2019) adopted a transformer-based model for analyzing student peer evaluations, securing a satisfactory accuracy score of 61.5%, notwithstanding the limited dataset (480 instances) used. In another application, Magooda et al. (2022) employed the Distil-BERT model, a derivative of transformer-based NLP, to provide automated feedback on student reflection quality, resulting in improved reflection quality. Similarly, Nehyba & Štefánik (2022) built automated transformer-based models that achieved an accuracy range of 76.56%-79.37% in classifying student-teacher reflections. Transformers-based models have also been applied in engineering education with promising results, demonstrating superior precision to manual qualitative methods and older NLP techniques (Ganesh et al., 2022; Becker et al., 2021; Gamieldien et al., 2023; Katz et al., 2021; Katz et al., 2023). The application of transformer-based NLP in analyzing exam wrappers is largely unexplored, presenting an opportunity to uncover SRL constructs and correlate them with student exam performance. Such insights could help educators enhance SRL strategies and performance outcomes in first-year engineering students.

4.4 Methods

4.4.1 Context of Study and Participants

Physics for Engineers II is a calculus-based introductory course, mandatory for first-year engineering students at a prominent R1 university in the Southeastern United States. The curriculum covers rotational dynamics, statics, oscillations, waves, fluids, heat, temperature, and basic thermodynamics. Students are introduced to various engineering disciplines, design issues, and technical communication, both written and oral, alongside a component of teamwork through projects. Assessment involves exams (54% of final grade), learning assignments, practice assignments, team projects (12% each), and labs (10%). The course blends in-person and online modes of delivery, comprising thrice-weekly lectures and twice-weekly labs. The study sample included students who submitted all exam wrappers during the Spring 2021, Fall 2021, and Spring 2022 semesters for a specific instructor. Students who did not write the exam or submit an exam wrapper over the three module exams were excluded from the study.

4.4.2 Data Collection

The data we used included existing responses to an exam wrapper given after the first three module exams of the Physics for Engineers II course. The exam wrapper data included students' responses to end-of-module exam wrappers. Each exam wrapper was given to students to complete after the first three exams in the course. The exam wrappers included 9 questions that asked students to reflect on their performance on the exam, reflect on future strategies, and develop a strategic plan for preparing for the next exam. The student responses were written in paragraph form and stored in the university learning management system with a unique identifier for students.

Table 4.2 summarizes the exam wrapper questions. All exam wrappers in this study ask the same questions except for the first exam wrapper of each semester, which did not contain the "Reflection" part since there were no previous exams to reflect on in the first exam wrapper. Table 4.2 also shows which SRL constructs from Zimmerman's model most align with that specific question. It should be noted, however, that students' responses may discuss other constructs of Zimmerman's model. For example, a student may respond to a question about their strategic planning, which is an SRL construct in the Forethought Phase, with a response that is more about a topic related to the Performance Phase. In other words, while a specific question could attempt to elicit a response related to a specific construct, students' responses may sometimes spill into other constructs in Zimmerman's model.

Table 4.2: Exam wrapper questions for end-of-module exam wrappers and related SRL constructs

Part	Question	SRL construct
Reflection	Reflection - What did you do differently between Exam 1 and Exam 2? Did the changes that you made make an impact? Did you reach your goal from the last Exam Wrapper?	Self-evaluation Adaptive reactions
Exam Dissection	There are skills other than physics knowledge necessary to complete this exam. Can you identify any skills or fundamental knowledge (non-physics) that are weak that impeded your ability to show what you know about physics concepts? What evidence do you have to backup your answer?	Task strategies Time management
Exam Preparation Reflection	Describe your process for learning/engagement during the regular week for this module. Can you identify any areas of improvement that could strengthen your learning during the regular week moving forward?	Task strategies Time management Environmental structuring Help-seeking Interest incentives Goal setting Strategic planning
	Describe your process for preparing to take the module exam. Can you identify any areas of improvement that could strengthen your preparation activities?	Task strategies Time management Environmental structuring Help-seeking Interest incentives Goal setting

		Strategic planning Self-evaluation
	How confident were you when the exam was passed out that you were ready to show what you knew about this module? What is one thing YOU could do over the next three weeks to support building confidence? What is one thing your instructor could help with to support building your confidence?	Self-efficacy
Strategic Plan	Define a measurable goal you would like to achieve during our next class module. This goal should be measurable and attainable in the next three-week period.	Goal setting Strategic planning Time management
	Identify one action you want to START doing that may better support your learning in this next module. Can you describe a specific action plan to support you in starting this action?	Strategic planning Adaptive reactions
	Identify one action you want to STOP doing that is detrimental to your learning in this next module. Can you describe a specific action plan to support you in stopping this action?	Defensive reactions Strategic planning
	How will you plan to celebrate if your goal is achieved?	Interest incentives Outcome expectations

4.4.3 Data Analysis

In this research, our objective was to pinpoint the various SRL constructs referred to by students of different performance profiles in their exam wrappers. To accomplish this, we utilized a three-step method:

1. Apply a codebook, devised by using the same method from a prior study, to categorize students' learning strategies into distinct SRL constructs (Gamieldien et al., 2023).
2. Examine students' grades across three module exams and then employ dimension reduction and cluster analysis techniques to discern patterns in student performance.

3. Compare how different learning strategies and the corresponding SRL constructs were distributed amongst these identified performance profiles.

We utilized the aforementioned three-step process on the two exam wrapper questions shown in Table 4.3. These questions were selected because they encompass domain-general skills as well as specific learning strategies that students use while preparing for exams. The chosen questions include:

1. "Non-Physics Skills" – This question primarily addresses the Reflection Phase of Zimmerman's model, as it prompts students to consider skills beyond physics knowledge needed to complete the exam.

2. "Preparation Process" – This question relates to the Performance Phase, as it inquires about the students' actual process for preparing for the module exam. Additionally, it pertains to the Forethought Phase by encouraging students to identify ways to enhance their preparation activities for the module exam.

It is important to note that while we categorize each question under a specific phase of Zimmerman's model, there are multiple sub-questions within each question name. For instance, "Non-Physics Skills" not only asks students to identify non-physics skills that may have hindered their exam performance but also requests evidence to support their response. These additional sub-questions often overlap with other phases of Zimmerman's model, serving as a bridge between phases. The purpose of this exam wrapper design is to guide students through the full SRL cycle for their varied responses and help them grasp the comprehensive objectives of the exam wrapper.

Table 4.3: Questions analyzed for this study

Question name	**Question**	**SRL Constructs**
Non-Physics Skills	There are skills other than physics knowledge necessary to complete this exam. Can you identify any skills or fundamental knowledge (non-physics) that are weak that impeded your ability to show what you know about physics concepts? What evidence do you have to back up your answer?	Task strategies Time management
Preparation Process	Describe your process for preparing to take the module exam. Can you identify any areas of improvement that could strengthen your preparation activities?	Task strategies Time management Environmental structuring Help-seeking Interest incentives Goal setting Strategic planning Self-evaluation

4.4.4 Codebook Background

To generate a codebook, one member of the research team read a subset of the student responses. This reading produced codes to describe each study strategy that students mentioned. Further details on codebook generation can be found in Gamieldien et al. (2023). For this paper and the method investigated, it is sufficient to assume one has a codebook without needing to know whether it was generated through traditional analysis, informed by existing literature, suggested by an automated NLP process, or arrived at by any other means. The codebooks generated for “Non-Physics Skills” and “Preparation Process” are shown in Tables 4.4 and 4.5 respectively.

4.4.5 Zero-shot Classification

Once we created the codebook, we used a ZSL classification model to identify the SRL strategies mentioned in the full dataset of student responses to a specific exam wrapper question (Yin et al., 2019). Zero-shot learning is a type of machine learning that allows a model to recognize and classify text that is out of distribution from its training set (Xian et al., 2017). The idea is that new input data can be categorized using a model that has been trained on potentially unrelated categories. In a traditional supervised machine learning paradigm, a classifier would be trained on labeled examples of input and target output. Such models typically do not transfer well to new data or new classifications beyond their original intended use case. The ZSL class of models addresses that non-transferability limitation by training a general-purpose model that theoretically has a broader understanding of language. It uses the idea of “attribute space”, a high-level description of the categories, to bridge the gap between seen and unseen categories.

The ZSL model we used is a state-of-the-art NLP model that works by embedding both the sentences that we want to label and the codes in the codebook into a high-dimensional space. Yin et al. (2019) refer to the sentence we want to label as the “premise” and the codebook sentences will be the “hypothesis”. The task is then to determine whether the hypothesis is true or false given the premise. In the current context, that translates to determining whether the sentence we are labeling matches a particular label in the codebook in the high-dimensional space. The ZSL classifier generated a probability score for whether each sentence belonged to a certain code and assigned that code to a sentence based on generating a probability above a cutoff threshold. For example, for the code “Lack Of Focus” the attribute space might include keywords or phrases such as “distraction”, “procrastination”, “multitasking,” etc. If a premise (student exam wrapper response) contains any of the words or phrases (or synonyms thereof) in the attribute space, the ZSL classifier might correctly label it as "Lack Of Focus". One of the advantages of the ZSL classifier is that it can assign multiple labels to one sentence if there is more than one idea in that sentence. Another advantage is the potential to use the model off-the-shelf, without additional fine-tuning or training. This is important for improving accessibility to using this general approach because most teachers and researchers will not have the means or opportunity to train their models from scratch. In a previous study (Gamieldien et al., 2023), we evaluated the "Non-Physics Skills" question along with two other exam wrapper questions and achieved accuracies of 87%, 85%, and 91% respectively as a result using the ZSL approach. The accuracies were determined by having one of the researchers read through 10% of the responses and rating them as a "1" for an accurate classification, "0" for something that was related but not an exact match, and a "-1" for a response that did not match with a topic that the ZSL method gave it. More details of this study can be found in (Gamieldien et al., 2023). The accuracies for each question were then determined by the following formula:

$$Evaluation\ Accuracy = \frac{Number\ of\ Accuracte\ Classifications}{Total\ Number\ of\ Labels\ Assigned}$$

In achieving the aforementioned accuracies, we thus concluded that the ZSL method was appropriate for the classification of the rest of the sentences and would be confident that at least 85% of the sentences fed to the model would be classified as exact matches. This conclusion is based on the idea of inter-rater reliability (Miles & Huberman, 1994) where some researchers deem an agreement of at least 80% between researchers on 95% of the codes is sufficient agreement to continue coding the rest of the data. In our case, we have between 85% - 91% agreement between the researcher and ZSL which indicates that the ZSL can be used for the analysis of the full dataset.

4.4.6 SRL Constructs Discussed in Exam Wrappers

After evaluating the accuracies of the codebooks, we then used Zimmerman's SRL model and the definitions of the different SRL constructs from Table 4.1 to code the responses for the "Non-Physics Skills" and "Preparation Process" questions. For example, if a student stated that they had "Strong Fundamental Skills" we assumed this was based on their own experiences and how they felt about their exam performance, we then labeled that topic as Self-efficacy. The results of our categorizing the topics in the two exam wrapper questions are presented in Tables 4.4 and 4.5.

Table 4.4: Codebook and SRL constructs for "Non-Physics Skills"

Topic	**SRL Construct**
ADHD	Causal attributions
Weak Fundamental Skills - Equations, Free Body Diagrams, Careless Errors	
Weak Fundamental Skills - Mathematics	
Mindset Before Exam - Confidence, Time Management, Distraction	Defensive reactions
Strong Fundamental Skills	Self-efficacy
Test Anxiety	
Lack Of Conceptual Understanding	Self-evaluation
Missed Points On Exam	
No Lack Of Fundamental Knowledge	
No Weak Fundamental Knowledge Lacking For Exam	
Strong Math Skills	
Weak Fundamental Skills - Multiple-Choice Strategies	
Organization Skills	Strategic planning
Lack Of Focus	Task interest

Attention To Detail	Task strategies
Better Equation Sheet	
Calculator Skills	
Conceptual Understanding	
Equations	
Physics Topics	
Problem-Solving Skills	
Required Non-Physics Skills	
Required Math Skills	
Study Skills	
Test-Taking Skills	
Unit Conversion	
Time Management	Time management
Time Management And Test Anxiety	

Table 4.5: Codebook and SRL constructs for "Preparation Process"

Topic	SRL Construct
Do Prep And Practice Questions	Task Strategies
Doing Problems That Are Hard	
Effective Study Strategies	
Make Equation Sheet	
Practice Exam and Check Answers	
Practice Past/Old Exams	
Put Examples Problems on Equation Sheet	
Review Notes and Practice Questions	
Review Notes and Watch Videos	
Reviewing Notes and Learning Pages	
Do More Practice Exams	Adaptive Reactions
Improve Conceptual Understanding	

Improve On Multiple Choice	
Improve Study Strategies	
Things To Improve On	Self-evaluation
Confidence	Self-efficacy
Go To EF Study Room	Environmental Structuring
Make Equation Sheet Earlier	Time Management
Study More or Earlier	
Time Management	
Time Practice Exams	

4.4.7 Student Performance Profiles

To comprehensively analyze student performance profiles, it was critical to have access to each student's scores from the three module exams conducted throughout the semester. Therefore, we excluded any students missing at least one exam score from our dataset to ensure our analysis was based on complete records. With the filtered dataset, we identified distinctive student performance profiles. To achieve this, we clustered students into different profiles using a non-linear dimension reduction method known as Uniform Manifold Approximation Projection (UMAP; McInnes et al., 2020) alongside a density-based clustering technique called Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN; Campello et al., 2013; McInnes & Healy, 2017) on their exam grades.

The purpose of using UMAP was to reduce the data of the three exams per student to a two-dimensional level so that we could visualize any patterns in the data. UMAP is an advanced nonlinear dimensionality reduction method designed for visualizing high-dimensional data in two or three dimensions (McInnes, & Healy, 2020). UMAP was chosen because it can maintain the structure of the data in lower-dimensional space. This was important for ensuring that students with similar grade profiles across the exams remained clustered together in the two-dimensional representation.

After reducing the data to two dimensions with UMAP, we grouped students into clusters using HDBSCAN. This method is a density-based hierarchical clustering algorithm. Density-based algorithms outperform methods such as k-means clustering in tasks involving noisy data, or data that do not have spherical-like shapes (He et al., 2022). We also did not need to specify the number of clusters, and the algorithm can handle noise. In this case, noise would be any students who do not fit into a specific cluster based on their exam grades. We used HDBSCAN to cluster the data into different student performance profiles based on the density of the points in our visualization. The parameter settings for UMAP and HDBSCAN can be found in Appendix Table C1.

4.4.8 Distributions of Learning Strategies and SRL Constructs Across Profiles

After we found the distinct student performance profiles, we investigated the distribution of learning strategies across these groups. This analysis aimed to identify any emerging topics of discussion of different strategies, skills, and SRL constructs across the different performance profiles, specifically focusing on the “Non-Physics Skills” and “Preparation Process” aspects of learning. We extended this investigation by looking at the same metrics across the three different exams to see if we could identify any temporal changes as the semester progressed. Finally, we used Table 4.1 as a point of reference to examine the SRL constructs associated with each strategy to perform a similar comparative analysis among student profiles and across exams.

4.5 Results

In this section we present (1) the student performance profiles, (2) the distribution of the various topics that students discussed in their responses to the exam wrapper questions “Non-Physics Skills” and “Preparation Process”, and (3) connections between these topics and the different SRL constructs outlined in Zimmerman’s model. We conclude the section by providing the distribution of SRL constructs for each exam wrapper question, categorizing them according to student profile and exam.

4.5.1 Student Performance Profiles

Removing students with missing grade data left 585 students with grades for all three module exams. The NLP workflow we used was still justified as we analyzed two exam wrapper questions for each of the three module exams, resulting in 3,510 exam wrapper responses analyzed. After using UMAP to reduce the number of dimensions for each student from three to two, clustering the students using HDBSCAN produced four clusters. The results from the dimension reduction and clustering are shown in Figure 4.2. We see that there were four clusters and each cluster is colored in a different color. Upon reviewing Figure 4.2, there could have been additional clusters. However, when we tested with a larger number of clusters, not all profiles exhibited a clear pattern like those identified with our current selection. Additionally, increasing the number of clusters could have led to overly detailed and specific insights, which might have made it difficult for the researchers to interpret and understand.

We utilized the cluster assignments from the UMAP and HDBSCAN analysis shown in Figure 4.2 to group students based on their performance profiles. For each of the identified clusters, we then plotted histograms showing the distribution of exam grades for the students within a particular cluster for each of the three exams which is shown in Figure 4.3. The student profile names were determined based on the average exam grade for each exam across clusters and the visual appearance of the grade distribution across exams.

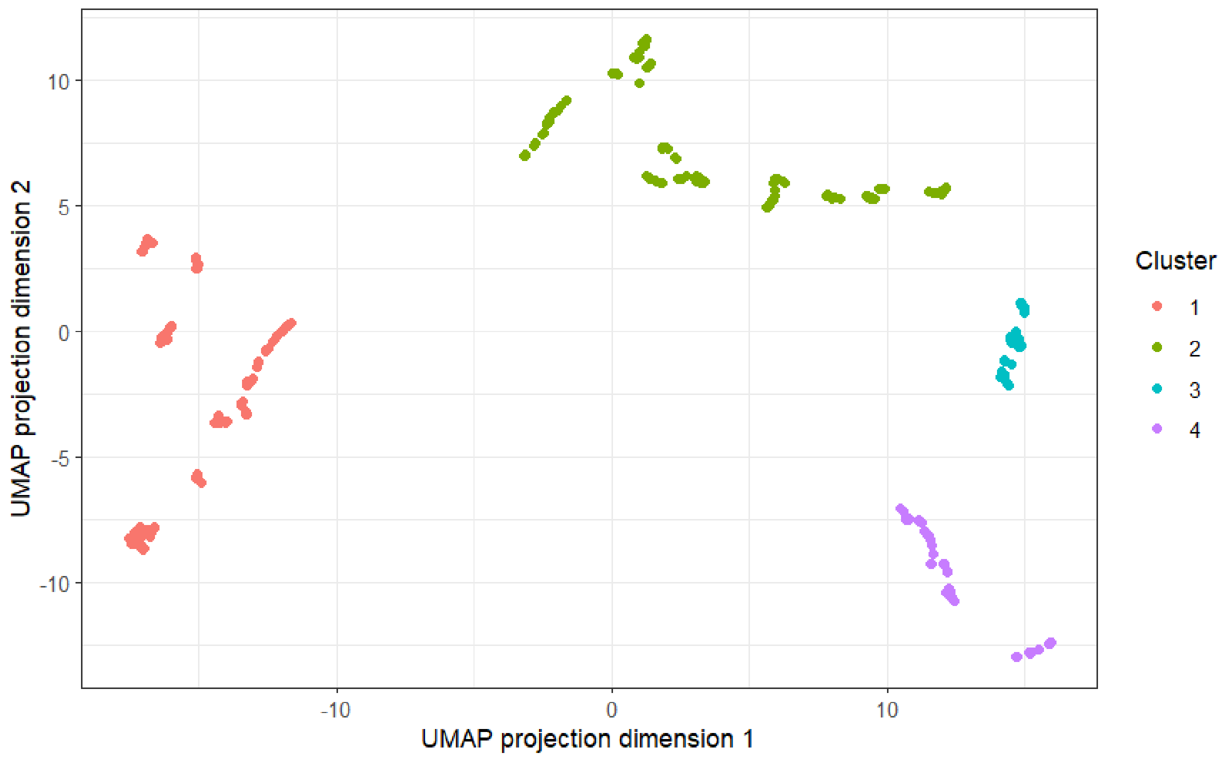


Figure 4.2: UMAP and HDBSCAN results indicating the four performance profiles based on exam performance across three exams in the semester

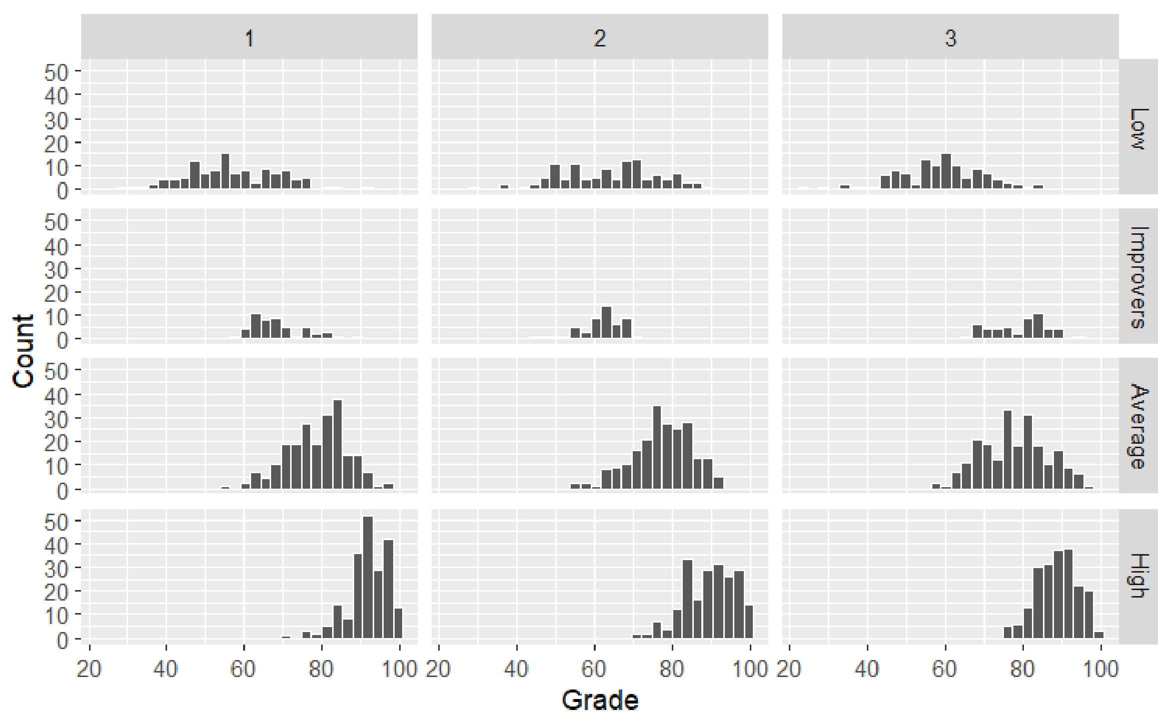


Figure 4.3: Cluster Analysis of Student Performance Based on Exam Grades. This figure presents the results of a cluster analysis using HDBSCAN to group students based on their performance in exams 1, 2, and 3. The horizontal axis represents the grades obtained in the exams, while the vertical axis displays the frequency of students achieving a specific grade for each student profile. Four distinct clusters were identified through HDBSCAN. The Low cluster comprises students who exhibited above-average performance on only one exam at most across the three module exams while performing poorly on the other two exams. The Improvers cluster encompasses students who demonstrated a progressive improvement in their grades over the course of the semester. The Average student profile represents those who maintained consistent performance throughout the semester. Lastly, the High profile consists of students who consistently excelled in all three exams.

A summary of the student profile results is shown in Table 4.6. The student profiles derived from the cluster analysis were categorized as Low, Improvers, Average, and High. Students assigned to the High cluster demonstrated strong performance across all module exams, achieving a high average exam grade of 90%. There were 205 students in that group. The Average profile encompassed 215 students who consistently performed well, with an average exam grade of 78.4%. The Improvers group exhibited a noticeable shift in their module exam grades with 51 students belonging to this group. Figure 4.2 illustrates that most Improvers initially scored around 60% on the first and second exams, but their grades improved and approached 80% on the third exam. The Low profile included students who consistently obtained the lowest scores on at least two out of the three exams, with an average exam grade of 59.7%.

Table 4.6: Student Profile and Exam Results

Profile	Description	Exam 1 (%)	Exam 2 (%)	Exam 3 (%)	Average Grade (%)	No. of Students
Low	Students achieving generally lower grades across the three exams	57	63.8	58.7	59.7	114
Improvers	Students whose grades dropped in the second exam, but then improved in the final exam	68.7	61.8	79.1	69.9	51
Average	Students who achieve consistent grades	78.8	77.2	77.8	77.9	215

High	Top performers in the exams	91.8	89.6	88.8	90.1	205

Tables 4.7 and 4.8 show the average number of strategies or skills that students of different profiles reported for each question that was analyzed and each exam wrapper. This was done to ensure that the data was not skewed by one group mentioning more strategies or skills than another group. To test the null hypothesis that there is no significant difference in the average number of strategies that students mention in exam wrapper responses among different performance profiles, we conducted a one-way ANOVA on both exam wrapper questions. We found that for both questions “Non-Physics Skills” ($F(3, 8) = 0.339$, $p = 0.798$) and “Preparation process” ($F(3, 8) = 0.745$, $p = 0.555$) we failed to reject the null hypothesis. This indicates that each performance profile contributed a similar number of average responses per exam wrapper and exam wrapper question.

Table 4.7: Average Number of Strategies Used by Each Profile Per Exam for “Non-Physics Skills”

	Exam 1	Exam 2	Exam 3
Low	4.5	4.1	3.5
Improvers	4.1	3.9	3.5
Average	4.4	4	3.5
High	4.3	3.8	3.5

Table 4.8: Average Number of Strategies Used by Each Profile Per Exam for “Preparation Process”

	Exam 1	Exam 2	Exam 3
Low	5.6	5.4	4.8
Improvers	5.8	5.5	4.8
Average	6.1	5.9	5.2
High	6.1	5.8	5.3

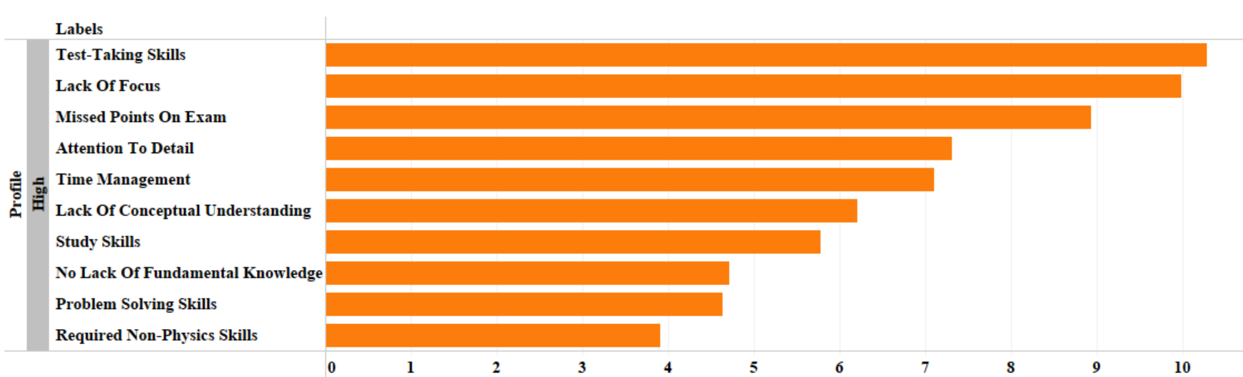
4.5.2 Learning Strategy and SRL Construct Distribution

This section presents the distribution of learning strategies discussed by students in their exam wrappers, with a specific focus on the “Non-Physics Skills” and “Preparation Process” questions. The objective was to examine the prevalence of different strategies among students throughout the semester and gain insights into the SRL constructs they discuss. All the horizontal axes in the figures represent the percentage of students for that graph. The reason for this is to make it easier to make comparisons between different profiles. To provide a comprehensive analysis, the distribution of topics addressed for each question across the entire semester is presented in Figures 4.4 and 4.9. This overview enables a broad understanding of the patterns and trends in student responses. Figures 4.5 and 4.10 show the topics mentioned by different profiles across exams for both exam wrapper questions.

Additionally, we map the individual strategies to SRL constructs in Zimmerman's model. Figures 4.6 and 4.11 display the distribution of these constructs for all students, offering a theoretically grounded picture of student SRL practices. Furthermore, a detailed breakdown of these figures based on student performance profiles is presented in Figures 4.7 and 4.12. Finally, we present graphs illustrating the SRL strategies students employed as stratified by different profiles across the three exams during the semester.

4.5.2.1 Learning Strategies Across Different Student Profiles - Non-Physics Skills

Analysis of student responses to the "Non-Physics Skills" questions in the exam wrapper revealed trends across the High, Average, Improvers, and Low performers. Figure 4.4 provides a visual representation of the top ten topics discussed by students in their responses.



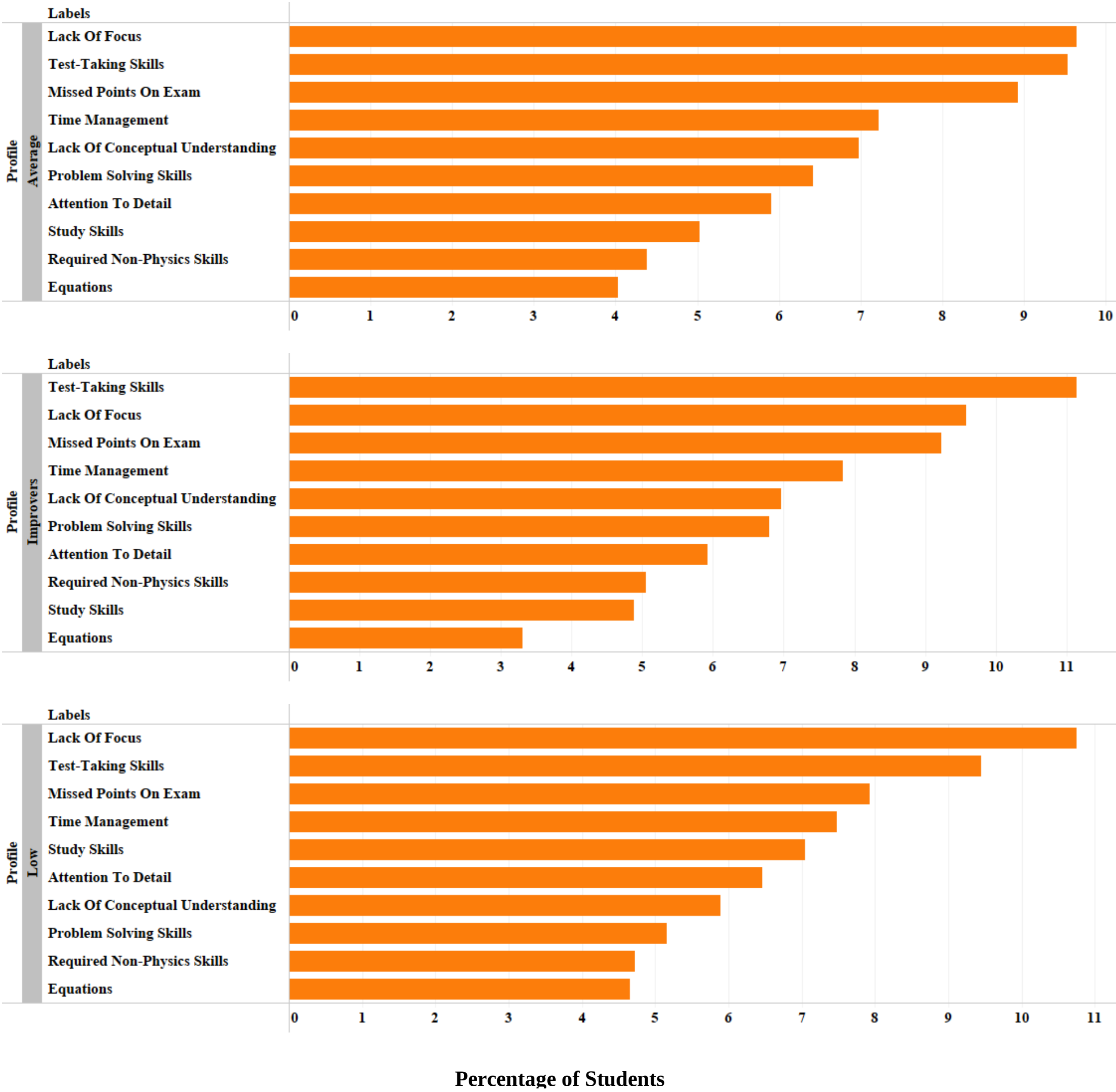


Figure 4.4 Distribution of topics that students of different performance profiles discuss in response to the “Non-Physics Skills” exam wrapper question

While most of the topics discussed by the different profiles were the same, the ranking of the different topics in the top ten differed. Among all profiles except for Improvers, the most frequently mentioned topic was “Lack of Focus,” which refers to skills perceived as weak and obstructing performance in exams. In contrast, the Improvers profile’s primary topic of discussion was “Test-Taking Skills” when responding to the exam wrapper question. For

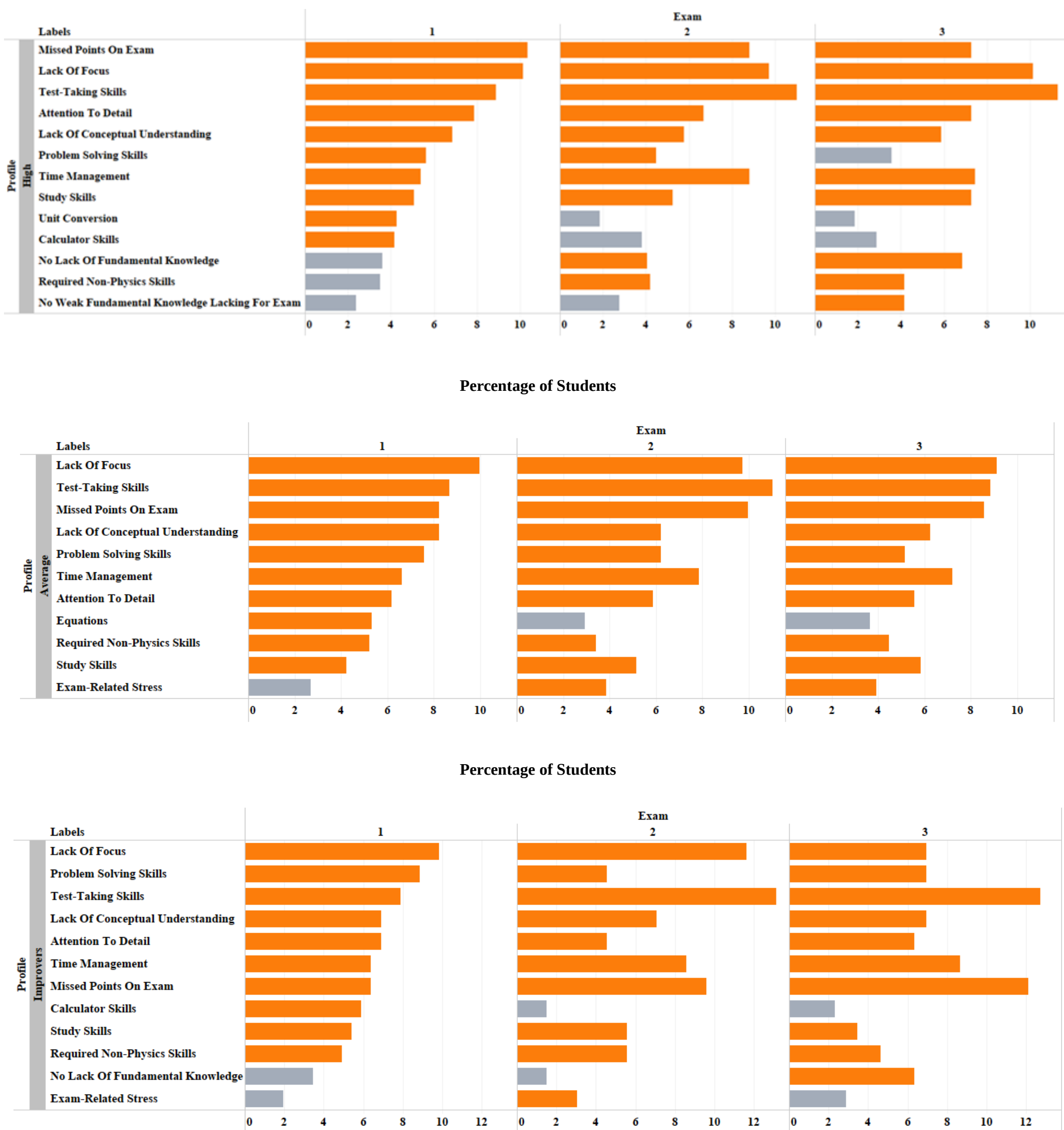
Improvers, “Lack Of Focus” was the second most discussed skill so it was still highly discussed by this group. The topic of “Time Management” emerged as a highly ranked skill across all groups, except for the High profile, which ranked “Time Management” fifth instead of fourth. Another noteworthy finding was the topic “Missed Points on Exams” which ranked third among the topics discussed by all the student profiles. The students highlighted the mistakes or errors they made, which resulted in point deductions on their exams. The skills and their high percentage of discussion across different profiles indicate that there are commonalities in skills that students from different profiles perceive as necessary for successful academic performance.

There is also varied emphasis concerning different performance profiles and the non-physics skills that they perceive as important to demonstrate their physics knowledge in exams. One of the skills that distinguished High performers from the other profiles was “Attention to Detail.” It ranked as the fourth most discussed topic for High performers but ranked lower for all other groups. The Low profiles identified “Attention to Detail” as the sixth most discussed topic, while the Improvers and Average profiles discussed this skill even less, placing it seventh. Another topic that set the High profile apart was their mention that they had “No Lack of Fundamental Knowledge” which makes sense since they were the highest performing group. The other profiles had discussed “Equations” instead, although it was the lowest discussed topic for all those profiles.

Other top ten skills included “Lack of Conceptual Understanding” and “Problem-Solving Skills.” The Average and Improvers group ranked those skills fifth and sixth, respectively, while the Low profile ranked them seventh and eighth. All the profiles ranked “Lack of Conceptual Understanding” higher than “Problem-Solving Skills”. The High profile, however, ranked “Problem-Solving Skills” at ninth and “Lack of Conceptual Understanding” at sixth.

4.5.2.2 Learning Strategies Across Exams – Non-Physics Skills

The graphs in Figure 4.5 present an examination of the top ten skills discussed by students in their responses to the “Non-Physics Skills” questions in the exam wrappers for three module exams throughout the semester. The topics discussed are divided according to the specific exams. Notable patterns emerge when comparing the skills mentioned by students across different exams and student profiles. The skills represented by the grey bars are not in the top ten skills for that particular exam but are the top ten skills mentioned in another exam for that profile.



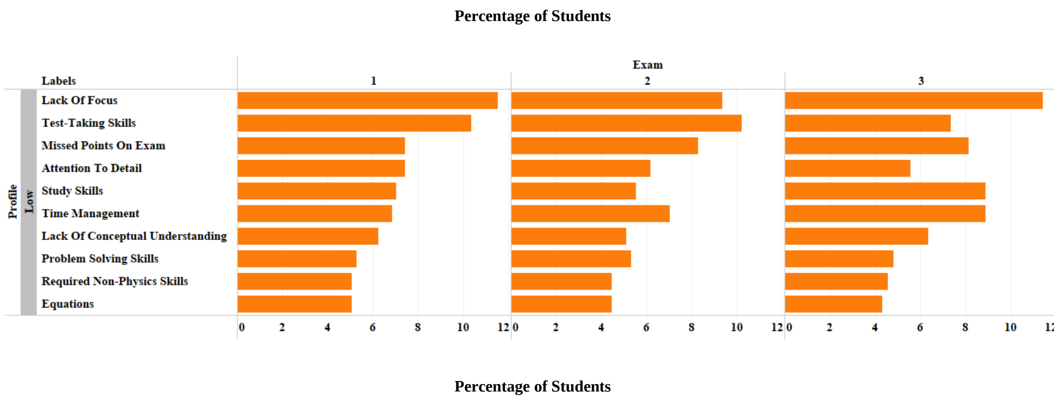


Figure 4.5 Distribution of topics that students of different performance profiles discuss in response to the "Non-Physics Skills" exam wrapper question broken down by exam

Firstly, all the profiles discussed "Time Management" as a skill more prominently in the subsequent exams. For example, "Time Management" went from the seventh discussed topic for High profile students to the third most discussed topic in exams two and three. Similarly, this skill went from sixth in the first exam, to fourth in the subsequent exams for the Average profile. Improvers and Low-profile students also saw similar increases in ranking after exam one where "Time Management" was ranked sixth in the first exam and then increased to third in the third exam.

The prominence of "Lack of Focus" as a discussed skill differed across student profiles and exams. The High profile discussed it consistently and it was always ranked as the second most discussed topic. The Low group consistently identified it as the highest-ranked skill needed to demonstrate their physics knowledge across exams but it ranked second for exam two. Moreover, the first exam saw a higher emphasis on "Lack of Focus" for both Improvers and Average profiles. The prominence of "Lack Of Focus" decreased slightly in the second and third exams for the Average profile but was still consistently discussed. For the Improvers "Lack Of Focus" increased in discussion from exam one to exam two, but then dropped in the rankings for the third exam.

The importance attributed to “Test-taking Skills” also varied across student profiles and exams. The Low group discussed these skills less as the exams proceeded, while the High group increased their emphasis on test-taking skills. The Average group ranked “Test-taking Skills” second in the first exam but saw an increase to first in the second exam before it dropped back to second. The Improvers, however, saw “Test-Taking Skills” go from third-ranked in the first exam to first-ranked in subsequent exams. This suggests that all profiles, except for the Low group, recognized the increasing significance of test-taking skills as the semester advanced.

Another noteworthy observation relates to the topic of “Missed Points on Exams.” High-profile students discussed this topic the most in the first exam, but then it dropped to as low as fifth by the third exam, suggesting that High profile students may have had fewer issues with missing points due to careless errors when compared to other profiles. The Average group displayed less of a pattern with missed points since the ranking of that topic went from third to second, but then went back to third. Interestingly, Improvers increased their discussion of missed points in exams since it ranked seventh in the first exam and then ranked second in the third exam. The grey bars in Figure 4.5 show instances where that particular skill was not in the top ten for the corresponding exam but was in the top ten for a different exam. This was never the case for the Low profile since they always discussed the same top ten skills. The only difference is the ranking of some of the skills discussed.

The High profile had the most difference in topics discussed. Notably, “Problem-Solving Skills” went from the sixth most discussed skill in the first exam to not being in the top ten in the third exam. “Unit Conversions” and “Calculator Skills” were discussed as ninth and tenth in the first exam, respectively, but then dropped out of the top ten for subsequent exams. The topic of “No Lack of Fundamental Knowledge” increased in prominence throughout the exams from not being in the top ten in the first exam. Similarly, the topic of “No Weak Fundamental Knowledge Lacking for Exam” was not in the top ten for High profile students until the third exam. This suggests that the High-profile students were addressing their issues with fundamental knowledge throughout the semester.

For the Average profile, the topic of “Exam-Related Stress” was not in the top ten in the first exam but became more prominent in the subsequent exams. Conversely, the topic of “Equations” was ranked seventh in the first exam, but then was not in the top ten for exams two and three. The Improvers group showed a similar trend to the High-profile group because they also had “Calculator Skills” as a top ten discussed topic which then dropped out of the top ten for exams two and three. The topic “No Lack Of Fundamental Knowledge” became a top ten discussed topic for the Improvers by the third exam. This falls in line with their grades improving by the third exam. Among the Improvers, “Problem-Solving Skills” was nearly the most discussed topic in the first exam wrapper, although it was still ranked relatively high for other profiles. However, its prominence decreased with each subsequent exam for all profiles.

4.5.2.3 Self-Regulated Learning Constructs – Non-Physics Skills

After the analysis of the topics discussed by students in their exam wrapper responses, we categorized them into specific SRL constructs within Zimmerman’s Model, as depicted in Figure 4.6. The construct that emerged as the most frequently discussed was Task Strategies, which encompassed topics such as “Calculator Skills”, “Test-Taking Skills”, “Problem-Solving Skills”, and “Study Skills.” These topics indicated that students recognized the importance of employing effective strategies in tasks related to exams and studying.

Following Task Strategies, the next prominent construct was Causal Attributions, wherein students reflected on non-physics skills that hindered their ability to demonstrate their knowledge. This construct included topics such as “Lack of Focus”, “Missed Points on Exam”, “Careless Errors”, and “Weak Fundamental Skills - Multiple Choice Strategies.” These discussions demonstrated that students acknowledged external factors or personal shortcomings that impacted their exam performance. Additionally, Time Management emerged as a frequently discussed SRL construct across all student profiles. It encompassed topics such as “Time Management”, “Organizational Skills”, and “Exam-Related Stress” The prominence of this construct suggested that students recognized the significance of managing their time effectively and addressing potential anxiety-related issues during exams.

On the other hand, the least discussed SRL constructs from Zimmerman’s model were Adaptive Reactions and Self-efficacy, each comprising only one topic. The topic of “Better Equation Sheet” fell under the Adaptive Reactions construct, indicating that students suggested improvements in the resources provided to them. The topic of “Test Anxiety” represented the construct of Self-efficacy, reflecting students’ concerns about their confidence and belief in their abilities during exams.

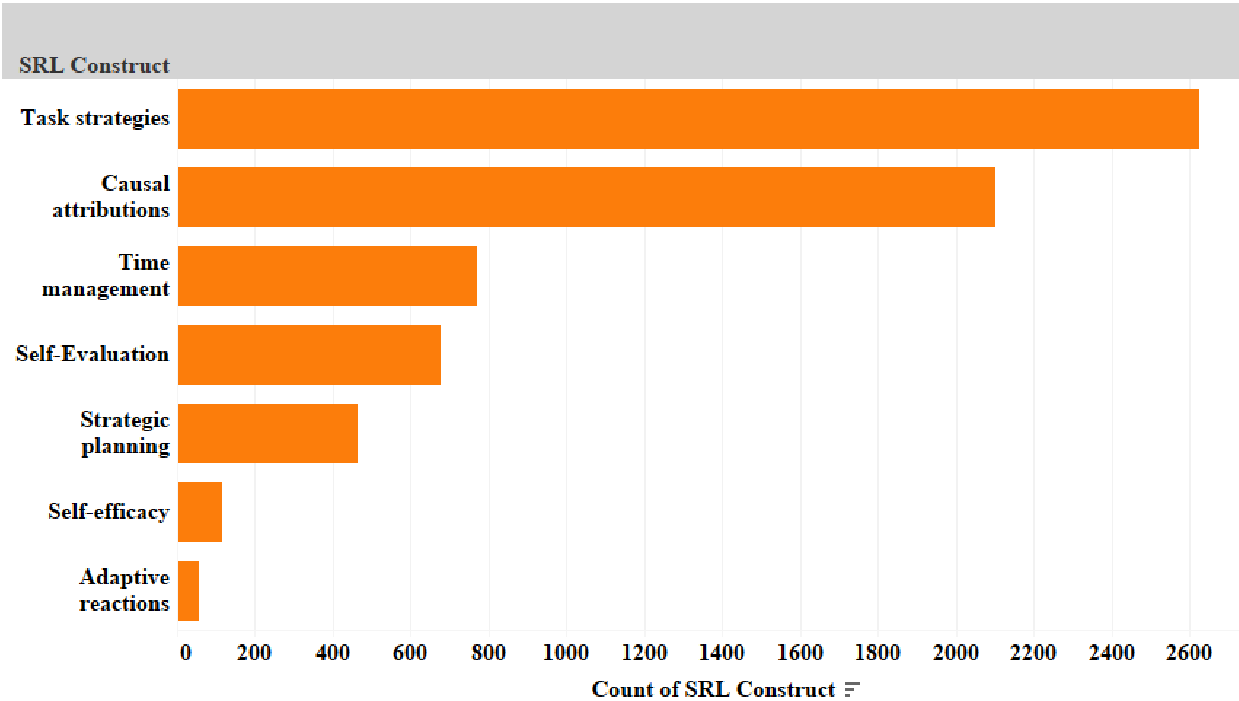


Figure 4.6 Distribution of SRL constructs from Zimmerman's model that students discuss in response to the "Non-Physics Skills" exam wrapper question

Upon examining the distribution of SRL constructs across different groups in Figure 4.6, we observe that the overall distribution of strategies remains consistent when considering all exam wrapper responses combined. However, when analyzing the constructs for each student profile per exam, several notable changes emerge. Figure 4.8 shows that within the High profile, students engage in more discussions related to Self-Evaluation than Time Management in the third exam. Moreover, in the third exam, students in the High profile exhibited a higher frequency of discussions related to Adaptive Reactions compared to Self-efficacy, in contrast to the previous two exams. Another notable aspect of the High profile is the increasing difference between Task Strategies and Causal Attributions in each subsequent exam wrapper throughout the semester. In the case of the Low profile, Self-Evaluation is discussed more than Time management in the second exam wrapper, but then the distribution shifts to the same ranking as in the first exam wrapper. The shifts in emphasis and the varying frequencies of discussions indicate the evolving focus and priorities of the students within each profile as they progress through the semester.

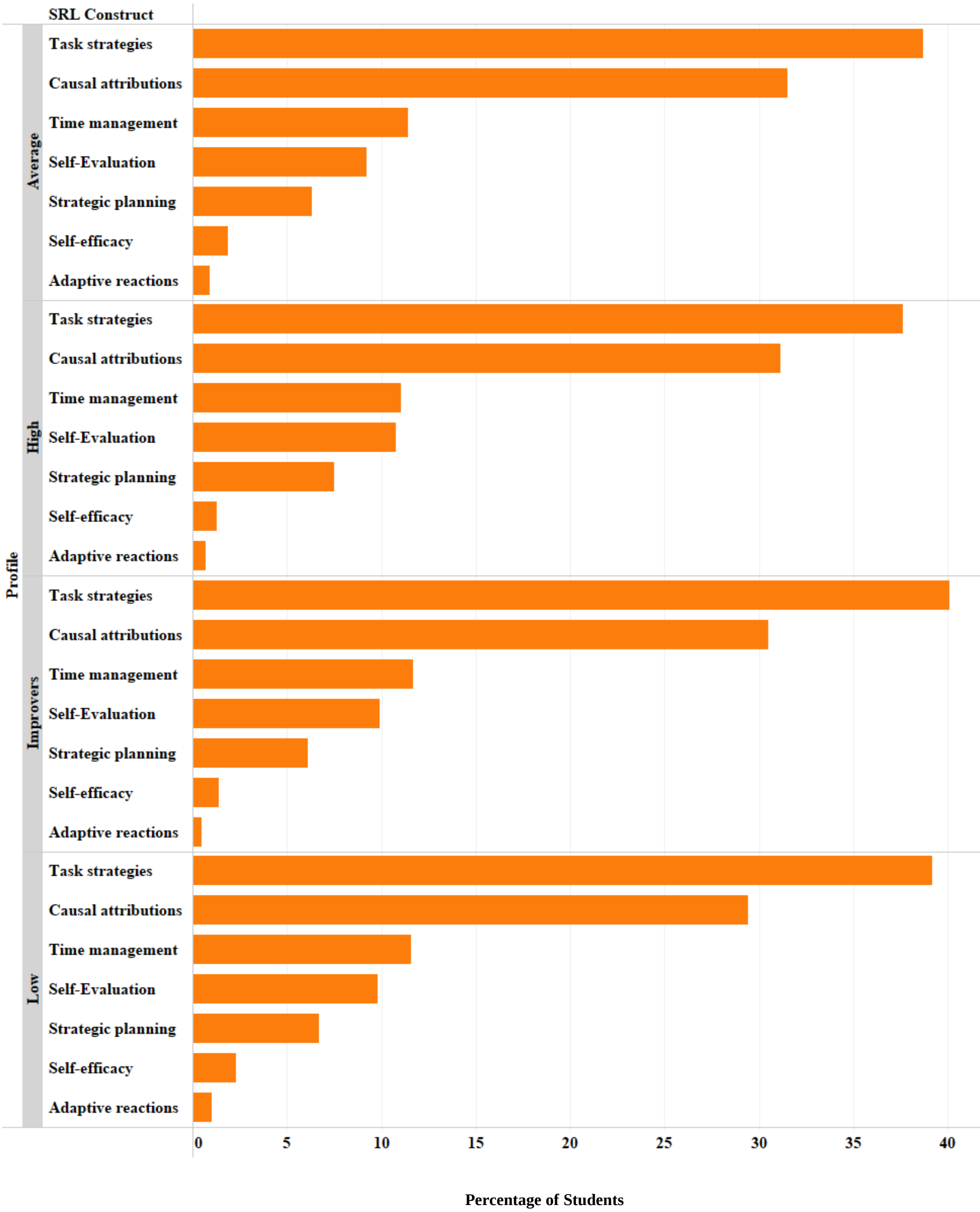
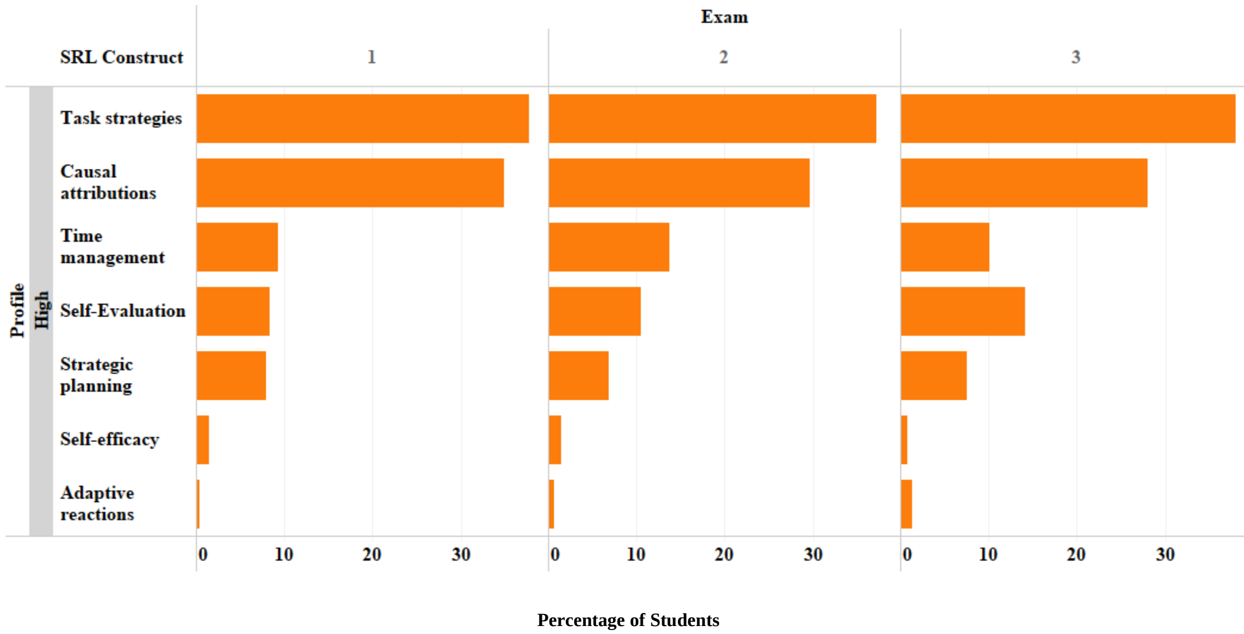
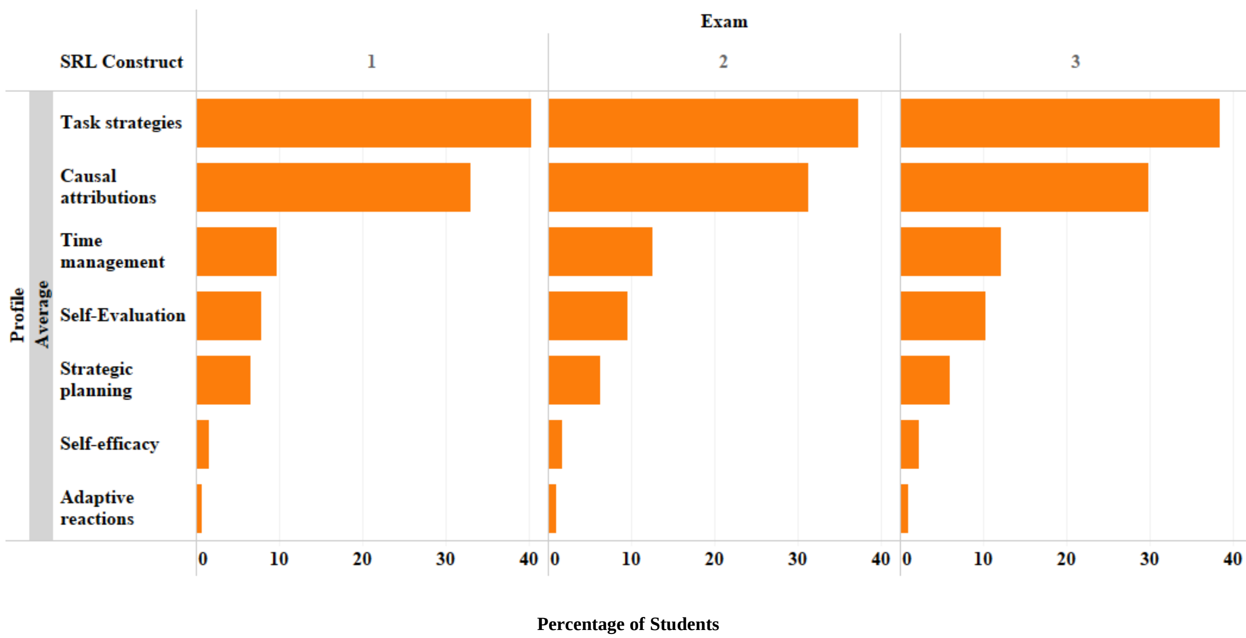


Figure 4.6 Distribution of SRL constructs from Zimmerman's model that students of different exam performance profiles discuss in response to the "Non-Physics Skills" exam wrapper question


Exam
SRL Construct
1
2
3
Profile
High
Task strategies
Causal attributions
Time management
Self-Evaluation
Strategic planning
Self-efficacy
Adaptive reactions
0
10
20
30
Percentage of Students


Exam
SRL Construct
1
2
3
Profile
Average
Task strategies
Causal attributions
Time management
Self-Evaluation
Strategic planning
Self-efficacy
Adaptive reactions
0
10
20
30
40
Percentage of Students

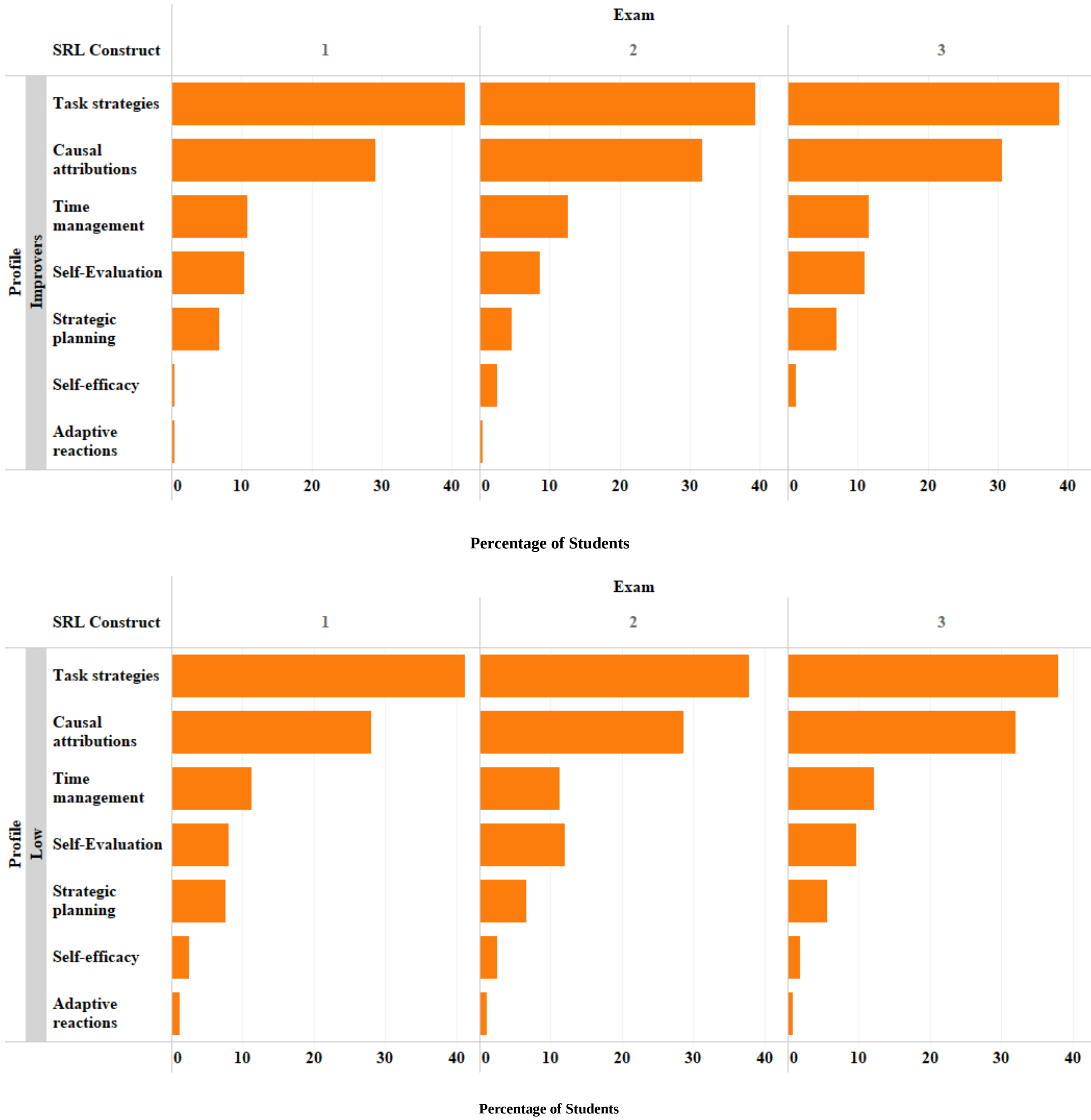
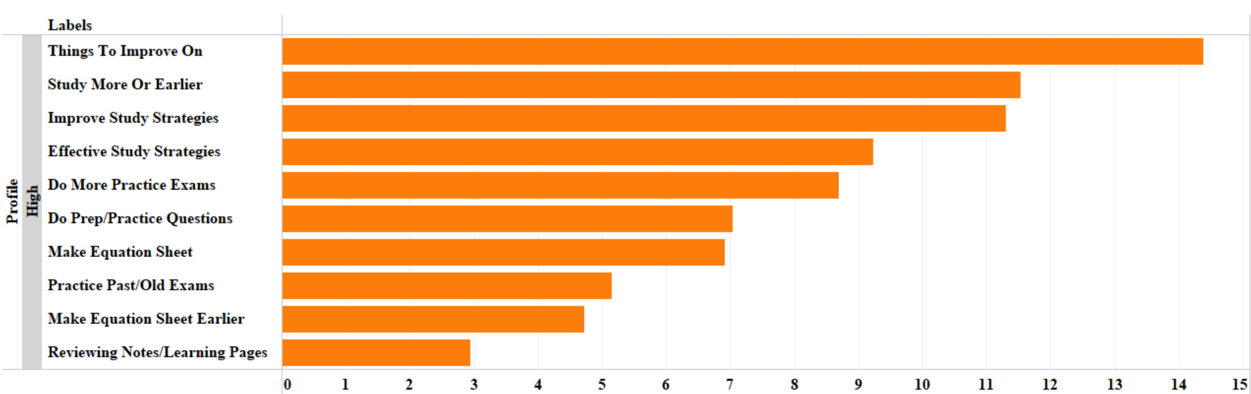


Figure 4.7 Distribution of SRL constructs from Zimmerman's model that students of different exam performance profiles discuss in response to the "Non-Physics Skills" exam wrapper question broken down by exam

4.5.2.4 Learning Strategies Across Different Student Profiles - Preparation Process

Figure 4.9 presents the predominant topics discussed by students in response to the "Preparation Process" question within the exam wrappers. This question aimed to elicit information about students' typical preparation activities for module exams and their suggestions for improving their preparation process. Among the four profiles, the most discussed strategy was “Study More or Earlier” reflecting students’ perceived need to allocate more time or start their preparation earlier. “Do More Practice Exams” was one of the more common topics that had an actual strategy it was discussing. Another prevalent strategy mentioned by students was “Do Prep/Practice Questions” indicating the value they placed on engaging in preparatory exercises. Additionally, the strategy of “Making an Equation Sheet” was frequently discussed as a helpful approach to exam preparation.

It is important to note that more general topics not related to specific preparation or improvement activities were not discussed in detail. However, they provide an understanding of the broader themes addressed by students in their responses. These overarching topics included “Things to Improve On”, “Improve Study Strategies”, and “Effective Study Strategies.” There are slight differences between these topics even though they could be seen as the same topic. “Things to Improve On” was a general cluster where students mentioned something they could improve on. This includes better study habits, improving time management, and confidence. “Improve Study Strategies” refers to instances where students mention that they need to improve their study strategies. “Effective Study Strategies” are instances where students mention or refer to study strategies that have worked well for them in their exam preparation process.



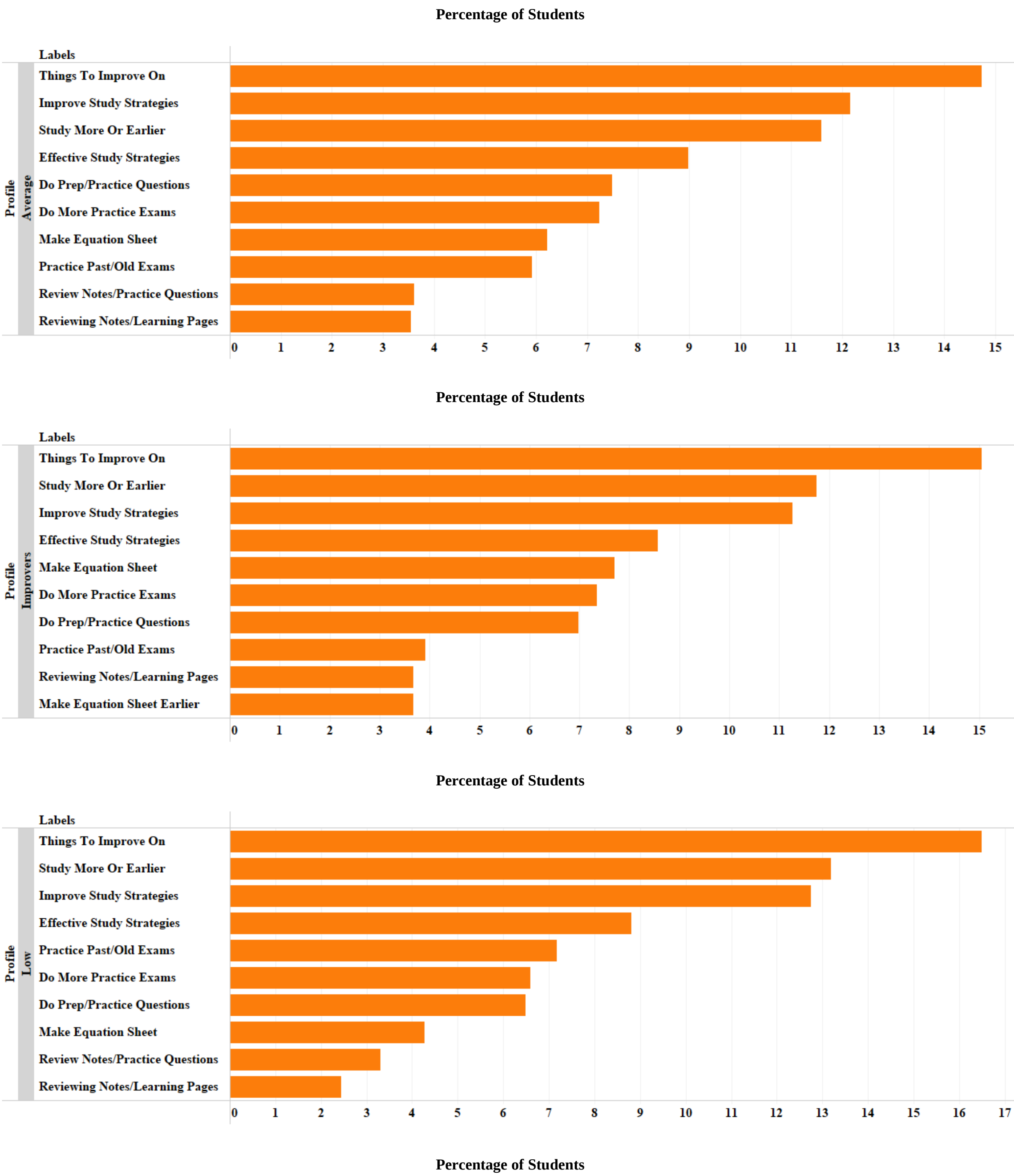


Figure 4.9 Distribution of topics that students of different performance profiles discuss in response to the “Preparation Process” exam wrapper question

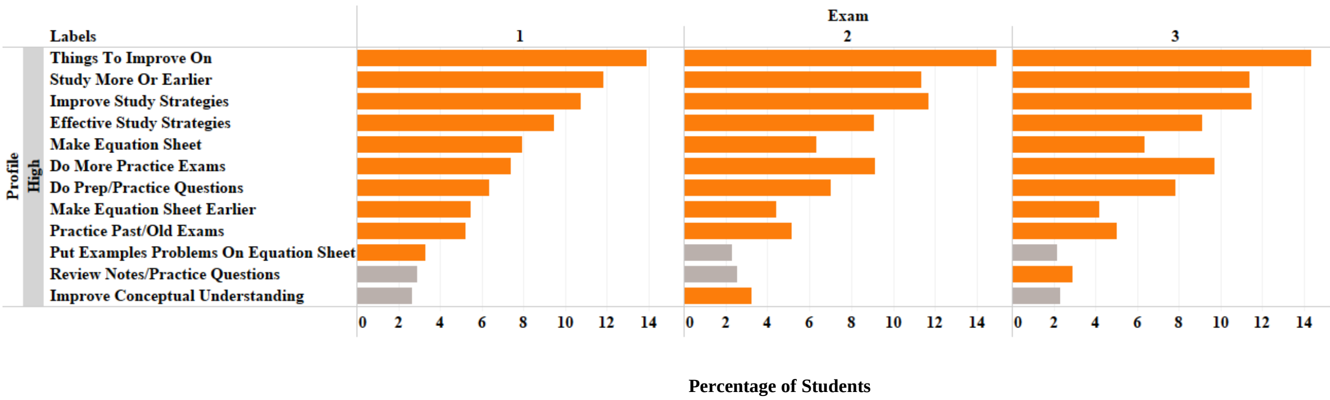
Examining the rankings of topics across different student profiles in Figure 4.9 reveals some variations. The Low profile assigned a higher rank to “Practice Past/Old Exams” compared to the other profiles. Furthermore, the Low profile ranked “Make Equation Sheet” lower in terms of an exam preparation strategy when compared to other profiles. The Improvers ranked “Make Equation Sheet” the highest out of all the profiles. The Low and Improvers profiles also discussed the strategy of “Do Prep/Practice Questions” less frequently than the other profiles, while the other profiles exhibited similar rankings for this learning strategy.

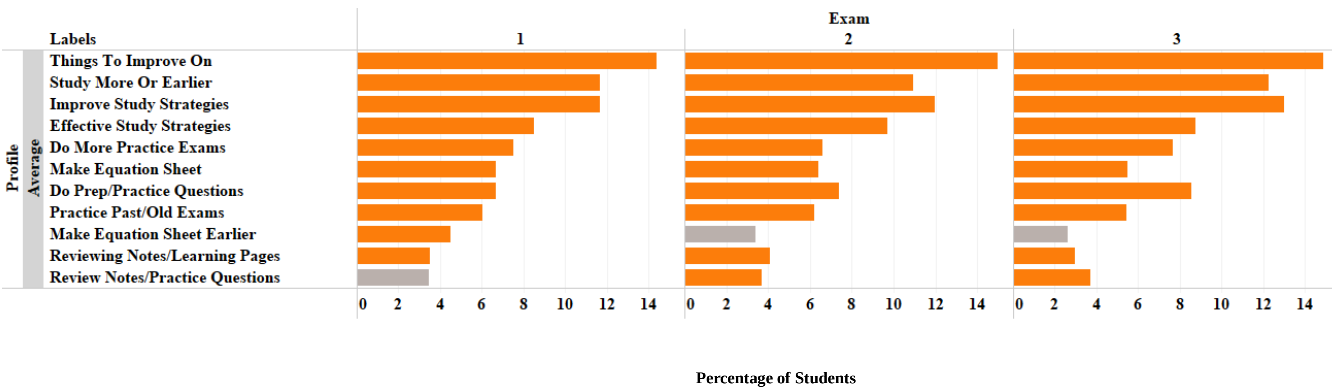
Regarding the High profile, they ranked “Do More Practice Exams” as an improvement point slightly higher than other profiles, with the Low profile discussing this strategy the least. The High profile also noted another improvement point in “Make Equation Sheet Earlier.” The Improvers also noted this in the top ten discussed topics but ranked it lower than the High profile. The other two profiles did not have this as a top-ten strategy. Similarly to the “Non-Physics Skills” question, we have some variation in strategies, but there are mostly commonalities in students’ exam preparation processes and their views to improve them moving forward.

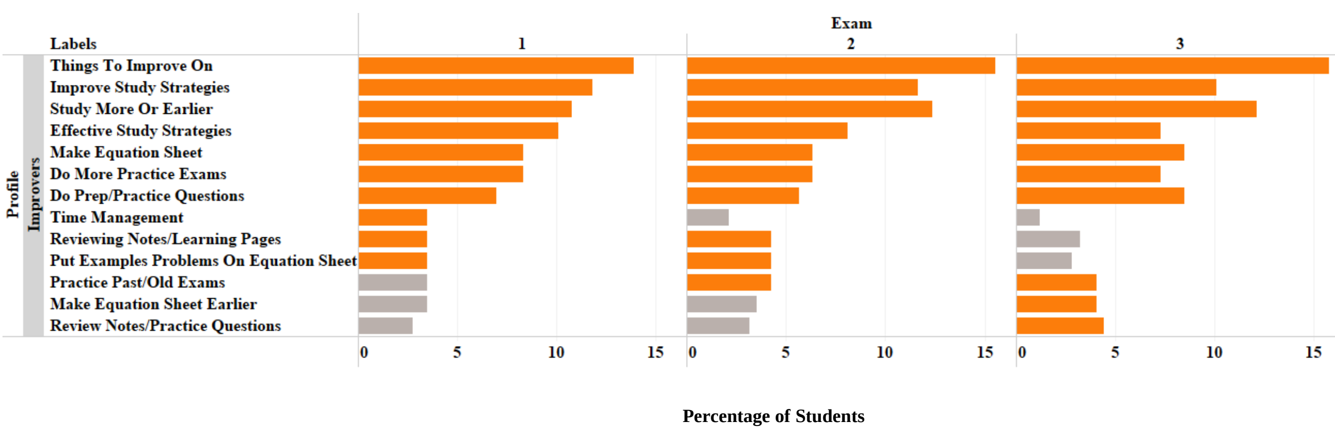
4.5.2.5 Learning Strategies Across Exams – Preparation Process

Figure 4.10 provides an overview of the topics discussed by students in response to the “Preparation Process” question across different student profiles and module exams throughout the semester. Analyzing the data, several trends emerge for each profile and exam. We see that most of the strategies discussed by the different profiles are similar, but there are a few strategies that are not in the top ten for that particular performance profile, which are shaded in grey. For the Low profile, “Time Management” and “Improve Conceptual Understanding” were two strategies that were not in the top ten in the first exam wrapper. This improvement in conceptual understanding was then the tenth most discussed topic in exam two but subsequently fell out of the top ten in the third exam. This is a trend that is observed in High profile students as well.

"Improve Conceptual Understanding" does not show up in the top ten for the Average and Improvers profile.







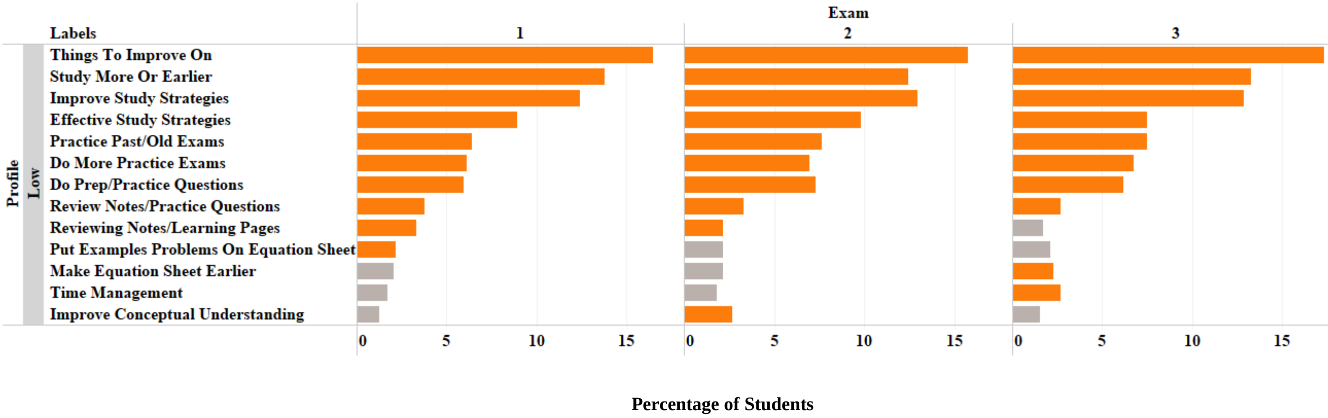


Figure 4.10 Distribution of topics that students of different performance profiles discuss in response to the “Preparation Process” exam wrapper question broken down by exam

Time Management was a topic that became increasingly important for Low profile students as the semester progressed. In contrast, it decreased in importance for the Improvers group by dropping out of the top ten discussed topics after exam one. The High and Average profiles did not discuss Time Management as a top ten strategy or improvement point. One other point of interest in the Low profile is that the “Reviewing Notes/Learning Pages” topic became less discussed for each subsequent exam after exam one. Conversely, the High profile students increased their discussion of this strategy when preparing for exams since it became a top ten strategy by the third exam. A strategy that is only present in the top ten for the High and Improvers profile is “Put Example Problems On Equation Sheet.” While this strategy did fall out of the top ten for the High profile after the first exam, and for Improvers by the third exam, it did not show up for the other profiles as a top ten strategy. This might indicate that the High and Improvers found this to be a useful strategy and did not need to mention it in subsequent exam wrappers when discussing their exam preparation activities.

The Average profile did not display as much variation in different strategies as the other profiles. “Make Equation Sheet Earlier” was a strategy that showed various trends across exams for different profiles. For the Average profile, it was a top ten strategy in the first exam but then fell away after that. For the Improvers, it became increasingly more discussed and became a top ten strategy by the third exam. The same was seen for the Low profile. High-profile students discussed this strategy as a top ten throughout each exam. A final note when looking at the strategy variation across different exams is that the Improvers profile shows the most variation in strategy since there were always three strategies that were in the top ten in one exam that were not present in the top ten in at least one other exam.

4.5.2.6 Self-Regulated Learning Constructs – Preparation Process

Based on the analysis connecting learning strategies to Zimmerman's SRL constructs, as depicted in Figure 4.11, Task Strategies emerged as the most frequently discussed SRL construct among the students. This construct encompassed strategies such as "Make Equation Sheet," "Practice Exam and Check Answers," "Doing Problems That Are Hard," and "Practice Past/Old Exams." The students' focus on these strategies indicated their active engagement in task-related approaches to exam preparation. The second most prevalent SRL construct was Adaptive Reactions, which primarily addressed the improvement aspect of the "Preparation Process" question. Strategies associated with this construct included "Do More Practice Exams," "Improve Conceptual Understanding," and "Improve on Multiple Choice." These strategies reflected the student's awareness of their areas of weakness and their intention to adapt and enhance their learning approaches. Time Management emerged as the third most discussed SRL construct, comprising strategies such as "Study More or Earlier," "Time Management," "Time Practice Exams," and "Make Equation Sheet Earlier."

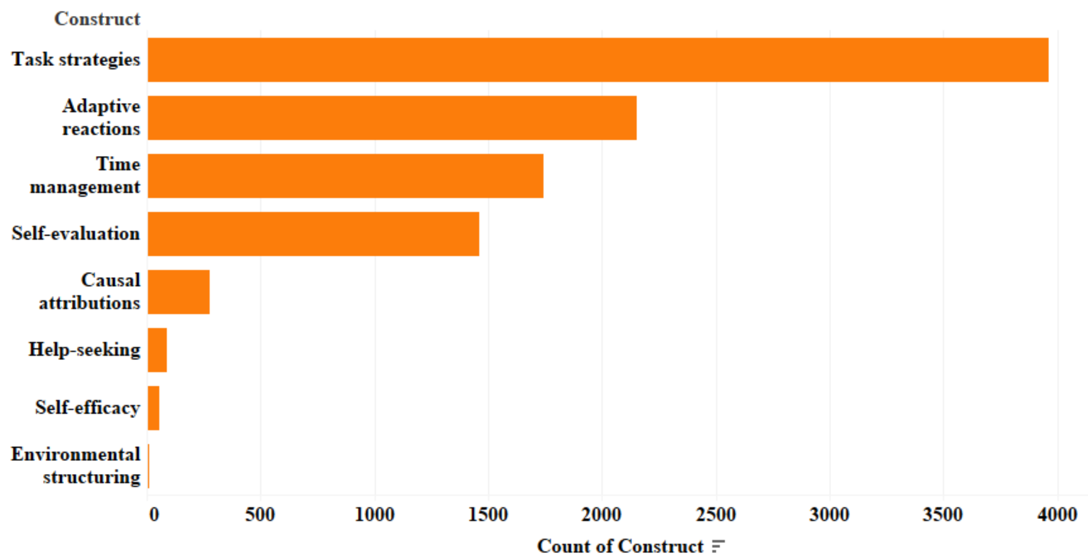
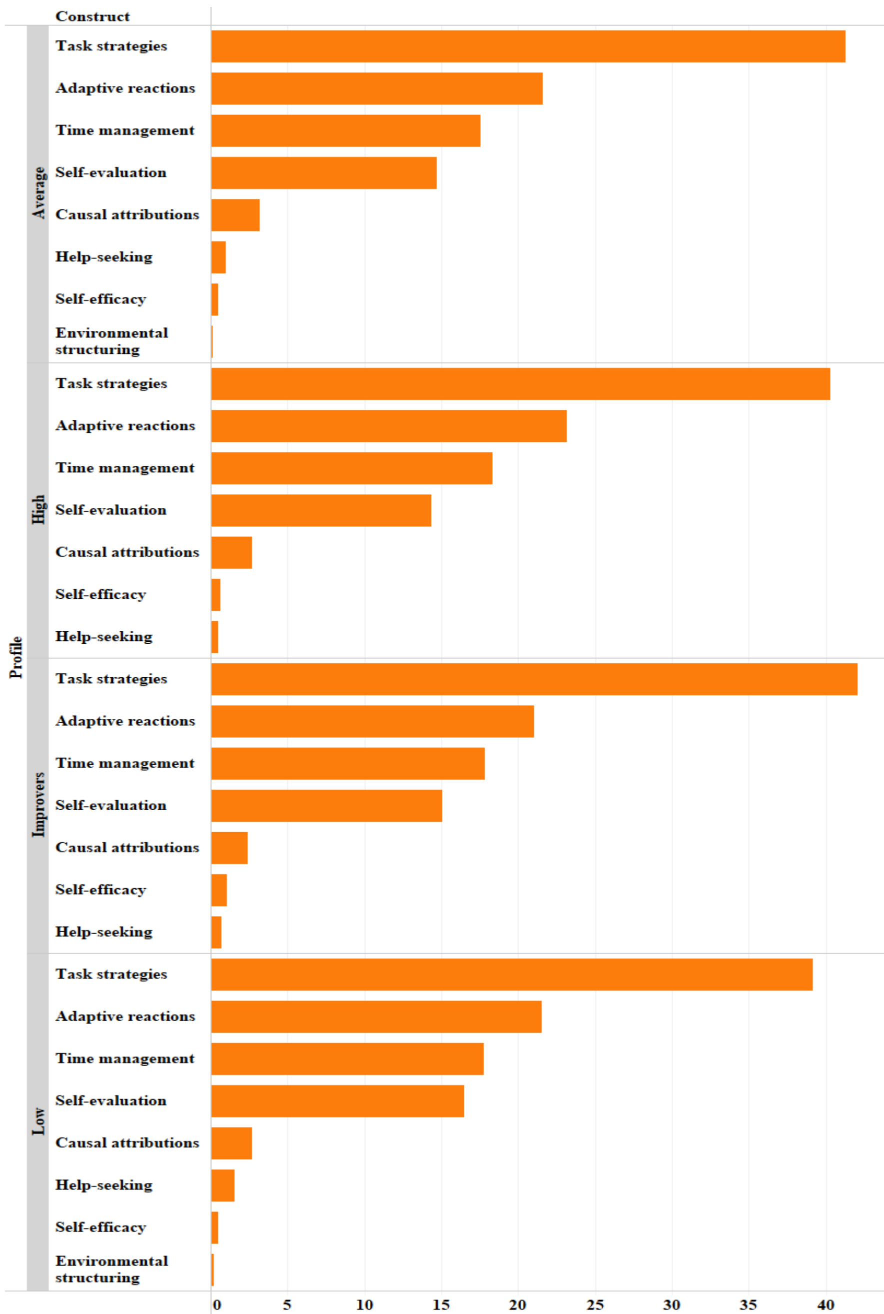


Figure 4.11 Distribution of SRL constructs from Zimmerman's model that students discuss in response to the "Non-Physics Skills" exam wrapper question

On the other hand, Environmental Structuring, Help-seeking, and Self-efficacy were the least discussed SRL constructs as shown in Figure 4.11. Environmental Structuring, represented by the strategy of "Go to EF Study Room," highlighted the students' limited emphasis on creating an optimal learning environment. Help-seeking, exemplified by "Group Study," indicated a lower inclination among students to seek external assistance or collaborate with peers during their exam preparation. Lastly, Self-efficacy, represented by the strategy of "Confidence," suggested a relatively lower level of discussion regarding students' belief in their abilities to succeed in their exam performance.

Figure 4.12 provides an overview of the SRL constructs discussed by different student profiles, highlighting variations in their rankings and frequencies. All of the profiles have the same ranking for the first five SRL constructs. The High and Improvers profiles have the same rankings of SRL constructs discussed and similarly, the Low and Average profiles have the same order of SRL constructs. The divergence among different profiles occurs at rank six onwards. The High and Improvers profiles have Self-efficacy as ranked sixth, whereas the Low and

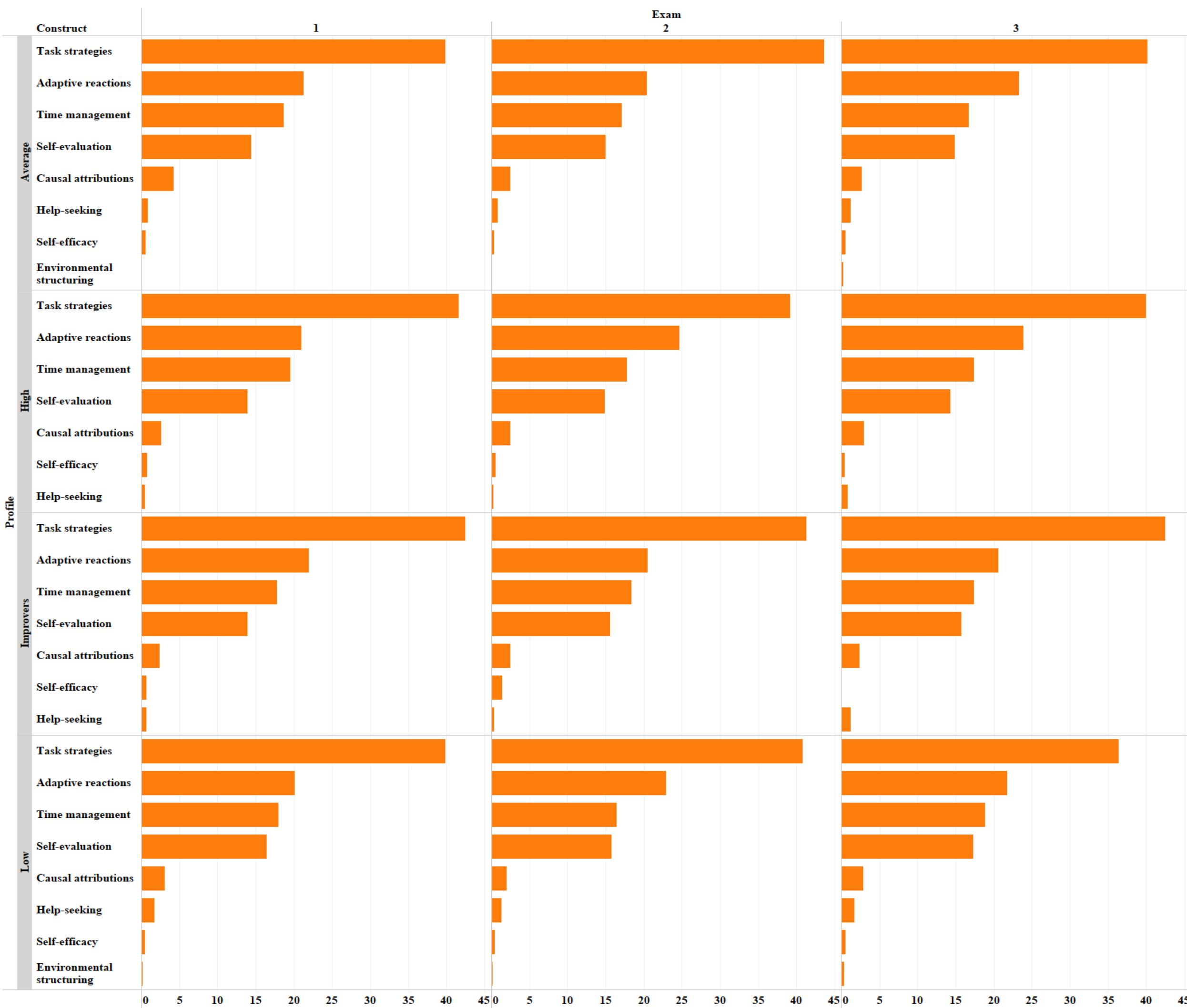
Average profiles have it ranked seventh. Another commonality between the High and Improvers profiles is that Help-seeking is the lowest-ranked SRL construct for both of those profiles, whereas the Average and Low profiles rank it higher. Both the Low and Average profiles have instances of Environmental structuring, which is absent from both the High and Improvers profiles.



Percentage of Students

Figure 4.12 Distribution of proportions of SRL constructs from Zimmerman's model that students of different exam performance profiles discuss in response to the "Preparation Process" exam wrapper question

Figure 4.13 presents a comprehensive analysis of the SRL constructs discussed by students belonging to different profiles in response to the "Preparation Process" question. The Low and Average profiles show no change in the ranking of different SRL constructs across exams. However, it should be noted that Low profile students mention Environmental structuring in each of the three exam wrappers, whereas the Average group only mentions Environmental structuring in the third exam.



Percentage of Students

Figure 4.13 Distribution of SRL constructs from Zimmerman's model that students of different exam performance profiles discuss in response to the "Preparation Process" exam wrapper question broken down by exam

The only shift in SRL constructs discussed by the High profile is that Help-seeking is more discussed than Self-efficacy in the third exam. Improvers also see an increase in Help-seeking. At first, it becomes less discussed from exam one to exam two, but then it is more discussed than the SRL construct of Self-efficacy by the third exam.

4.6 Discussion

In this section, we unpack the findings from this study, which investigated the degree to which differences exist in the SRL strategies that students of varying performance profiles report in their exam wrapper responses. Our findings not only answer this central research question but also provide deeper insights into the dynamics of learning strategies and SRL within engineering education. Moreover, we will address the implications of these findings, exploring how they might inform pedagogical practices, influence curriculum development, and shape future research in this field. Recognizing the limitations of our study, we also acknowledge the opportunities that these present for further refining our understanding of SRL strategies in the context of academic performance.

4.6.1 Time Management

Time management emerged as a prominent theme in students' discussions of exam preparation across all student profiles. It was one of the most frequently discussed topics overall, especially in responses to the question about non-physics skills. Time management was in the top ten topics for all profiles, with mentions by Low and Improver students when responding to the Preparation Process question. This aligns with research on SRL skills that are effective for learning which states that students lack adequate knowledge of time-management strategies

(Britton & Tesser, 1991). The fact that time management was mentioned in both exam wrapper questions – one talking about exam-specific skills (Non-physics skills) and another relating to exam preparation (Preparation Process) speaks to the complexity of this SRL construct and its importance in students' success both during exams and while preparing for them. Time management is not considered a single trait, or skill, but it is characterized as a multidimensional process through which a student will deliberately control what tasks they perform, when they perform those tasks, and for how long they do so (Wolters et al., 2017). Therefore, fostering these skills is crucial for a student's academic success but also a challenging skill to master.

The increase in emphasis on time management with each subsequent exam for each profile also suggests that students are becoming more aware of the role that time management plays in their academic success (Basila, 2014). College presents new challenges regarding time management that most students may be unaware of given that their time is not as structured for them as when they were in high school (van der Meer et al., 2010). College students are expected to manage more hours of their day most of the time without much external assistance (Dembo, 2013). This increase in awareness of time management is especially important for first-year engineering students given that it is their first year out of high school. Many learning strategies require SRL skills like time management. For example, studying more requires that you allocate more time to studying, careless errors in exams could be due to poor time management in exams (Nilson & Zimmerman, 2013), and completing the required tasks in one's preparation for exams requires the management when and how long these tasks will take (Dembo et al., 2007). Overall, time management was the third most discussed SRL construct across profiles and exams for both exam wrapper questions. This prominence highlights students' understanding of managing time effectively to address exam anxiety (Britton & Tesser, 1991; Schraw et al., 2006; Zeidner, 1998).

The findings of this study also indicate that time management is an important aspect of exam preparation and a non-physics skill that could impede a student's ability to perform in exams.

Another factor that is related to time management is procrastination. Time management is an important predictor of procrastination (Wolters et al., 2017) which is something many students discussed while referring to time management. O'Brien (2002) estimated that between 80-95% of college students engage in procrastination which is the delaying or putting off of task to a later stage (Kim & Seo, 2015). In a meta-analysis conducted by Kim & Seo (2015), the authors found a negative correlation between procrastination and academic performance. This could mean that the students' struggles with procrastination could be contributing to the prevalence of time management, particularly among Low performing students.

Specific time management strategies like starting preparation earlier and spending more overall time studying were emphasized across profiles, aligning with research on effective exam preparation through proper time allocation and avoiding cramming (Dunlosky et al., 2013). Low performing students also recognized time management's role in improving their preparation activities, through approaches like scheduling and prioritizing tasks (Britton & Tesser, 1991; Schraw et al., 2006).

While "Make Equation Sheet" was mentioned by all profiles, only Improvers and High performers consistently mentioned "Make Equation Sheet Earlier" as an improvement strategy. The mention of making the equation sheet ahead of time could be seen as a time management strategy that could be a reason that differentiates High and Improver students from Low students. For Low performers, increased discussions of time management reflect growing but still emerging recognition of the value of planning and deep learning (Zimmerman, 2002). Their continued challenges with time management point to issues commonly faced by struggling students in applying SRL strategies (Zimmerman & Schunk, 2001).

4.6.2 Attention and Lack of Focus

Another prevalent topic discussed by most profiles was “Lack of Focus.” This finding is consistent with the literature on attention and concentration as essential skills for academic success (e.g., Duckworth et al., 2007). One factor that could contribute to this mention of lack of focus is the SRL construct of Environment Structuring. Bellur et al. (2015) examined the multitasking behaviors of college students both outside and within the classroom and found that students who multitask frequently within the classroom have lower college GPAs. The authors also found that texting emerged as a dominant activity both while attending class and while doing homework. Students who reported multitasking while doing homework spent more time studying outside of class, thereby contributing to inefficient study habits. Furthermore, distractions such as a tactile or auditory notification from a mobile device are associated with poor performance on attention-demanding tasks such as exam preparation (Stothart et al.,2015). This finding contributes to the emphasis on the importance of environmental structuring as a potential challenge to first-year engineering students. Additionally, “Lack Of Focus” was mentioned increasingly by Low performers in subsequent exams as a non-physics skill that impeded their ability to perform well in exams. In contrast, Improvers mentioned “Lack Of Focus” less in each subsequent exam to the point where it was the most discussed topic in exam one but then fell to sixth by the third exam.

Further to that, Zimmerman & Schunk (2001) posit that students who take control of their learning process, including maintaining focus, are more likely to perform well academically. Lack of focus can impair a student’s ability to effectively process and comprehend the complex concepts typically presented in engineering physics. Turning back to SRL specifically, Pintrich (2000) suggested that a student's goal orientation significantly impacts their ability to self-regulate their learning. For example, if a student loses focus or becomes easily distracted, it may indicate that they are not goal-oriented or motivated, which could negatively affect their exam performance. Lack of focus can also be seen as a motivational component of SRL that is associated with self-monitoring (Zimmerman & Moylan, 2009). For example, Chi et al. (1989) found that students who engaged in more self-monitoring tended to be better problem solvers. Related to the code of "Lack of Focus" was the increased prominence of the discussions of "Missed Points on Exams" in subsequent exams by all profiles except the Low profile, reflecting their growing awareness of mistakes leading to point deductions. This aligns with the literature on the significance of error analysis for performance improvement (Dunlosky et al., 2013) and the need for attention to detail to avoid careless errors. This attention to detail for error analysis is also an instance of self-monitoring skill that can assist students in reducing missing points on exams due to common errors made by students (Du Bois & Staley, 1997). The missed points on exam code, as well as the reference to lack of focus, align with what Chew et al. (2016) found that exam wrappers can be useful in assisting students' SRL by having them reflect on their missed points through error analysis and make students more aware of what they need to pay attention to. Conversely, the Low profile exhibited decreased focus on this topic, suggesting potential gaps in their use of error analysis. This contrast between Low performers and other profiles may point to differences in self-evaluation approaches and could be one of the contributing factors to Low performers' exam grades. The Improvers profile's attention to "Missed Points on Exams" aligns with literature on metacognition, which emphasizes the role of self-awareness in learning processes (Flavell, 1979). Acknowledging mistakes was a more common strategy among Improvers and could be why they showed an increase in their third exam scores.

On the other hand, the High and Improvers profiles demonstrated confidence in their foundational knowledge, as shown by the frequent discussion of “No Lack of Fundamental Knowledge.” “Attention to Detail,” was particularly prevalent among High performers, which is consistent with studies linking conscientiousness, including meticulousness, to academic success (Duckworth et al., 2007; Komarraju et al., 2009). Attention to detail may also be related to “Missed Points On Exams” in the sense that carelessness could lead to missing exam points. In terms of the overlaps between attention and focus and time management which we discuss in the previous section, Self-Observation is an overarching SRL construct that is associated with both these areas of students’ struggles (Du Bois & Staley, 1997). The authors suggest that Self-Observation requires a student to be aware of themselves and their actions in a self-diagnostic way. This could then lead to a self-motivating function that occurs when a student begins to self-evaluate. For example, if a student is procrastinating, and they want to reduce their procrastination, it requires them to manage time and be aware of their behavior. This requires time management as a skill as well as attention. Furthermore, Passow & Passow (2017) suggested that attributes including time management, planning, and focusing on goals were identified by engineering practitioners and faculty as the most important skills to contribute to the future engineering workplace. Based on our findings about these non-physics skills, and the literature on SRL, we believe that fostering these attributes is crucial in developing future engineers, especially considering the neglect of developing a more holistic skillset in engineering education (Clough, 2005).

4.6.3 Exam Preparation Tasks

While developing an engineering student's non-physics skills is a crucial component to being successful in exams, there are also domain-specific preparation tasks that students have discussed in their exam wrapper responses. These exam preparation tasks have implications for increasing our awareness of what students are doing to prepare for exams and to what extent there may be differences in the exam preparation tasks of students in different performance profiles. The analysis of the exam wrapper responses to the "Preparation Process" question provides insights into the learning strategies employed by different student profiles within each performance profile and across different profiles as they prepare for the module exam. Common strategies, such as "Study More or Earlier", "Do More Practice Exams", and "Making an Equation Sheet" were universally recognized across all student profiles. This recognition signifies a shared understanding of the importance of time management, active engagement with course material, and effective resource creation in exam preparation.

All the profiles ranked "Do More Practice Exams" as an improvement point which could indicate their ongoing commitment to active learning strategies and their realization that taking practice exams not only works on their application of physics concepts but also their test-taking skills. An exam preparation strategy that was common to all profiles with the Low-profile students assigning a higher rank, was "Practice Past/Old Exams." Engaging in practice exams is supported by the literature as an effective strategy for exam preparation since it helps students familiarize themselves with the format, content, and timing of the actual exams (Balch, 1998). Furthermore, this strategy aligns with the literature emphasizing the benefits of repeated retrieval practice and self-testing to enhance learning and performance (Dunlosky et al., 2013; Roediger & Karpicke, 2006). While each performance profile mentions doing past exams as a strategy to prepare for exams, little is known about how students from different performance profiles use these past exams to prepare for the exams. For example, if solutions to past exams are available, the memorization thereof does not promote deep conceptual understanding (Litzinger, 2011).

The Average profile discussed “Do Prep/Practice Questions” less frequently compared to other profiles, except for the High profile in the first exam. They also demonstrate a shift in emphasis from studying more or earlier in the first exam to improving study strategies in the second and third exams. This shift indicates that students in this profile recognized the need to optimize their study approaches and adopt effective strategies to enhance their learning outcomes instead of simply increasing the time they need to study. It is worth noting that the topic of “Do Prep/Practice Questions” differs from “Practice Old/Past Exams” because it refers to questions given to students during the module that focus on specific topics and concepts during the module that is completed at the time the concepts are taught. Students review these “Prep” and “Practice” questions as an additional preparation activity to practice past exams.

The prominence of making an equation sheet makes sense since this allows students to write down notes that could aid them in taking the exam. The Low and Improvers profile highlights the strategy of creating an equation sheet earlier more prominently in the third exam compared to the first and second exams. This suggests that students in these profiles started to recognize the value of organizing key information and formulas as a study aid later on in the semester which could be why their exam averages are lower than the High profile. Making one’s equation sheet earlier, along with the strategy of putting examples on the formula sheet, could be a key reason for the increase in the Improvers exam performance.

Putting examples of problems and solutions appears to be unique to High and Improvers profiles but drops out of the top ten for both after the first and third exams, respectively. The lack of its mention in the Low and Average profiles could indicate different methods of exam preparation or the perceived effectiveness of this strategy. This strategy aligns with the literature on worked examples, which demonstrates their effectiveness in promoting learning and problem-solving skills (Atkinson et al., 2000; Sweller et al., 1998). Therefore, being able to reference worked examples in an exam could significantly improve their exam performance as it could provide a framework for solving complex physics problems in the exam. McCaskey (2014) found that equation sheet strategies differ for different students from an epistemological perspective. Having equations that do not need memorization attempts to remove the emphasis on memorization from the course. In addition to not memorizing concepts, the author also found that students used equation sheets to assist them with recalling concepts, definitions of symbols, and using examples as a framework to solve problems.

The Improvers recognized the need to engage in additional practice exams to improve their exam preparation. This strategy reflects their proactive approach to enhancing their understanding and test-taking skills. This could also be an indication of why they were able to improve their exam performance since Balch (1998), Rowland (2014), and Karpicke & Aue (2015) demonstrated the benefits of the testing effect on exam success.

It is essential to consider that these learning strategies may interact and complement each other. For example, combining the strategy of practicing past exams with effective time management and utilizing study resources like equation sheets can lead to more efficient and comprehensive exam preparation.

4.6.2 Limitations

Several limitations exist for this study which may impact the quality of this research. The limitations we address are related to NLP, the ZSL classifier, and LLMs in general. We also address the study design limitations in terms of generalizability. Finally, we address limitations in the interpretation of the results concerning profile-based analysis and the influence of other variables in the study.

4.6.2.1 Limitations of Zero-shot Learning Natural Language Processing

The first limitation is the potential for inherent bias that could be carried over from the training data for the LLM (Bender et al., 2021). These models are trained on web data, so any harmful and toxic data found on the internet could present itself in how the LLM classifies sentences (Tamkin et al., 2021). Since the responses are classified into a group of codes and themes set by the researcher, the bias limitation is not a major concern given the context of this study. Another limitation of the ZSL classifier is the dependence on pre-defined classes. The accuracy of the classifier is dependent on the predefined labels and how well those labels represent the data. In cases where the classes do not represent the data well, there would be suboptimal results. This dependency on predefined classes also leads to the next limitation which is the difficulty in assessing the accuracy of the model. Since there is no "ground truth" or labeled data, assessing the accuracy of the ZSL classifier can be challenging. We attempted to minimize the concerns of accuracy and potentially suboptimal results in labeling in a previous study where we evaluated a subset of the data classified using ZSL to determine the level of matching codes between the classifier and a human researcher (Gamieldien et al., 2023).

4.6.2.2 Self-report Bias in Exam Wrappers

The study relies on students to report their learning strategies in their exam wrapper responses. This self-reporting may not accurately reflect their actual learning strategies or SRL constructs. A reason for this could be that students may be reporting what they believe to be "ideal" strategies rather than what they do. While the first-year engineering students are prompted to provide some detail to their descriptions in their exam wrapper responses, it may not be clear whether they use those SRL strategies or to what extent they may have used them. Schunk & Zimmerman (2011) also note a limitation in self-reflections since they are offline - meaning that it is not happening while students are engaged in the task - and offer solutions such as live measures of SRL such as computerized tasks, trace evidence, and observations, although these methods come with their own set of limitations. Another limitation of this study is that self-reflections are subjective, and this could be the reason why we see similarities in the strategies and impediments that students discuss. For example, practicing more exams was frequently discussed by all profiles but could differ between profiles in many ways including how many exams they currently practice, what environments they are practicing in, and the intentions behind practicing more.

4.6.2.3 Generalizability

This study has several limitations in terms of generalizability. Firstly, the research focuses primarily on an engineering physics course, which inherently imposes boundaries on the generalizability of the findings. The learning strategies and SRL constructs that students discuss may differ for other courses and in different educational contexts. Furthermore, the study population comprised engineering students which imposes a further limitation since it limits the generalizability of the results to students in other fields of study. Finally, while this study does capture the learning strategies and SRL constructs of students across the three module exams during three semesters, it may not account for any changes in these periods over more extended periods. This element of the study could mask some longer-term effects we may see which could affect the interpretation and applications of these findings.

4.6.2.4 Interpretational Limitations

While the findings of this study are promising, there are several interpretational limitations to consider. First, the practice of categorizing students into discrete profiles (High, Low, Average, and Improvers) for analysis may oversimplify the complexity and diversity of individual learning experiences. It may not fully account for the nuanced variations within each category and how these individual differences might influence learning strategies or SRL constructs. Furthermore, this approach overlooks potential intra-individual variability in these strategies and constructs over time, which could provide richer insights into students' learning processes. The study does not incorporate potential unobserved variables, such as personal traits or past academic experiences, which may significantly influence the adoption and efficacy of specific learning strategies and SRL constructs. Therefore, while the study provides valuable findings, these limitations should be kept in mind when interpreting the results.

This study employed a simplified approach by assigning each student-reported learning strategy to a single SRL construct. While this facilitated a clearer analysis, it may not fully encapsulate the multifaceted nature of learning strategies, as many could intersect with multiple SRL constructs. For example, creating an equation sheet, while treated as a task strategy, also embodies elements of time management and potentially fosters self-efficacy. Consequently, this methodology may not comprehensively capture the complex dynamics of SRL, underscoring the need for more nuanced methods in future research that can account for the interconnectedness of learning strategies across multiple SRL constructs.

4.6.3 Study Implications

The results of this research have implications for a variety of stakeholders engaged in educational practice and research in both engineering education and SRL. The study places great emphasis on SRL constructs such as Task Strategies, Adaptive Reactions, Causal Attributions, and Time Management, underscoring their importance in academic success (Zimmerman & Schunk, 2001). It calls for a deeper exploration of these constructs, especially those that have been previously underemphasized. This study paves the way for using NLP and LLMs as a means to scaling up studies on SRL to large classrooms which could give us better insights into the best SRL strategies in different engineering education contexts and for different student profiles. Unique learning approaches have been identified, along with the value of early preparation, practice exams, equation sheet creation, focus on tasks and exams, and test-taking strategies for students across all profiles.

4.6.3.1 Implications for Instructors and Students

Regarding improving students' performance, we have a few suggestions that instructors and students could implement. The first SRL skill that students could focus on is time management. This skill was most frequently discussed by students of all profiles, especially the Improvers group. Time management skills are foundational in a student's ability to perform tasks to be successful in a first-year engineering physics course. Additionally, the first year of college comes with added responsibilities and time management is a different challenge in college where students have more responsibility to manage their time (Dembo & Seli, 2008). Instructors could provide students with guidance on planning their study schedules, strategies to avoid procrastination, and practicing skills to keep track of time spent on different tasks throughout their day. Turning to specific strategies for managing time, instructors could guide students in exam preparation by encouraging students to prepare for exams earlier through practice questions and exams, emphasize spaced study sessions during the semester instead of cramming.

A related area of SRL that also forms part of a student's ability to perform their tasks is focus and attention which could also be seen as metacognitive monitoring (Zimmerman & Moylan, 2009). Instructors could inform students of the importance and benefits of improved focus and deep attention to being successful in engineering. Furthermore, instructors could also suggest concentration strategies, and promote a culture of minimizing distractions potentially caused by phones or other devices when preparing for exams. Instructors could also discuss the impacts of divided attention with students and how it may affect their academic lives. Discussing new strategies and having students engage in strategies they were not familiar with can be useful as we saw with Improvers who had the most diverse mentioning of different topics across the three module exams during the semester. This adjustment of strategies is in keeping with findings by Grohs et al. (2018) who suggested that students should engage in adjustment of strategies earlier in the semester, an approach that is in keeping with Zimmerman's model of the SRL cycle.

The Low-profile students showed significant concern for "Exam-Related Stress," aligning with research indicating a correlation between heightened anxiety and lower academic performance (Zeidner, 1998). This suggests the necessity for interventions to help these students manage exam-related stress for improved outcomes. In the "Equations" topic, Low-profile students found the knowledge and application of equations challenging, consistent with studies indicating difficulties with applying complex equations in physics (Kuo et al., 2013; Vaiyavutjamai & Clements, 2006). This hints at the need for targeted support to help Low performers comprehend and apply equations effectively. Understanding the concepts behind equations rather than simply having them on the equation sheet was also something highlighted by McCaskey (2014), who suggested that being more conceptually driven when writing down equations could lead to more success in exams.

Peer learning and tutoring support could be used to encourage help-seeking early in the semester instead of at a later stage as seen in the exam wrapper results. Facilitating study groups where High and Average performers could voluntarily work with Low performers on their application of equations conceptually, learning from errors, test-taking skills, and other fundamental skills can help students from each profile in diverse ways. The Low performers can learn and practice these fundamental skills that are expected of them, and the High and Average performers can re-enforce their fundamental skills like equations, unit conversions, and calculator skills that are important in being successful in exams.

Finally, instructors could assist students with equation sheet creation by showing them how worked examples can be used as a reference during the exams (Zimmerman, 2013). Another equation sheet strategy would be to have students make their equation sheet during the semester in a spaced way so that they learn what is on their equation sheet as they make it. This will help students become familiar with their equation sheets in the exam, which will improve their efficiency. Experiments with a course-provided equation sheet could also be used that minimize any errors that students could make while making their equation sheet, such as copying down formulae incorrectly or not having the appropriate equations to answer a question (Paquin et al., 2020).

4.6.4 Future Work

One opportunity for future research lies in the application of other AI techniques. While our study used ZSL for text analysis, other AI techniques may provide additional insights and mitigate some of the limitations of ZSL. Exploring these methodologies could enhance the richness and accuracy of the data analysis, leading to a more nuanced understanding of student learning strategies. Additionally, integrating other data sources could further enrich our findings. The use of observational data or interviews could complement the exam wrappers we used in this study. Finally, we could further analyze the other questions in the exam wrappers that we have access to such as student confidence, reflections from previous exam wrappers, the learning process in the regular work-week, and behaviors that students want to start or stop doing that could lead to their success in exams. These additional data sources could provide more context and depth, allowing for a more comprehensive understanding of students' SRL behaviors and strategies.

4.7 Conclusion

To answer our research question, "To what extent are there differences in the SRL strategies that students of different performance profiles report that they use in their exam wrapper responses?", this study delved into student learning strategies and SRL in the context of an engineering physics course. By applying dimension reduction and clustering on student exam grades, we categorized students into distinct performance profiles - Low, Average, High, and Improvers. The findings of our investigation shed light not only on the diversity of study strategies and the differential emphasis placed on various SRL constructs across these profiles but also on the strategies and SRL constructs that are important for students across performance profiles.

Task-oriented strategies, including practicing past exams, enhancing conceptual understanding, and developing equation sheets, were universally recognized as critical. Additionally, non-physics skills such as focus, time management, test-taking skills, and missing points were the most discussed skills in students' exam wrappers. The adaptive nature of learners was evident, with students continually adapting their learning approaches across different exams.

We found that the SRL constructs of Task Strategies, Time Management, Adaptive Reactions, and Causal Attributions were the most discussed constructs in the "Non-Physics Skills" and "Preparation Process" questions. However, elements such as Environmental Structuring, Help-Seeking, and Self-Efficacy were less prioritized, marking potential areas for enhancement.

Our study is not devoid of limitations. The inherent subjectivity of self-reported data, the specific context of the course and population, and the potential loss of nuanced information due to the ZSL analysis challenge the generalizability and full interpretation of our findings. Nonetheless, in answering our research question, this study provides critical insights into the multifaceted nature of student learning in engineering education, while also advancing the application of AI techniques in educational research. As it underscores the need for a more holistic approach to SRL, it adds valuable knowledge to efforts directed at refining educational practices and curriculum design, ultimately aiming to enhance student performance.

References

Adams, R., & Felder, R. (2008). Reframing Professional Development: A Systems Approach to Preparing Engineering Educators to Educate Tomorrow's Engineers. *Journal of Engineering Education*, *97*. https://doi.org/10.1002/j.2168-9830.2008.tb00975.x

Amann, E., & Rzepka, S. (2023). The effect of goal-setting prompts in a blended learning environment—Evidence from a field experiment. *Economics of Education Review*, *92*(C). https://econpapers.repec.org/article/eeeecoedu/v_3a92_3ay_3a2023_3ai_3ac_3as0272775722001042.htm

Anakok, I., Woods, J., Huerta, M., Schoepf, J., Murzi, H., & Katz, A. (2022). Students' Feedback About Their Experiences in EPICS Using Natural Language Processing. *2022 IEEE Frontiers in Education Conference, FIE 2022, October 8, 2022 - October 11, 2022*, *2022-October*. https://doi.org/10.1109/FIE56618.2022.9962557

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*(2), 181–214.

Autrey, J. L., Sieber, J. M., Siddique, Z., & Mistree, F. (2017). *Work in progress: Quantification of learning through learning statements and text mining in an engineering design, build, and test course*. *2017-June*. Scopus. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85030528378&partnerID=40&md5=bcdcbbc069df824dc09f9ecaaac7273e

Balch, W. R. (1998). Practice versus review exams and final exam performance. *Teaching of Psychology*, *25*(3), 181–185.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory* (pp. xiii, 617). Prentice-Hall, Inc.

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, *50*(2), 248–287. https://doi.org/10.1016/0749-5978(91)90022-L

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*, *41*, 586–598. https://doi.org/10.1037/0022-3514.41.3.586

Bandura, A., & Wessels, S. (1994). *Self-efficacy* (Vol. 4). na.

Basila, C. (2014). Good Time Management and Motivation Level Predict Student Academic Success in College On-Line Courses. *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL)*, *4*(3), 45–52. https://doi.org/10.4018/ijcbpl.2014070104

Bellur, S., Nowak, K. L., & Hull, K. S. (2015). Make it our time: In-class multitaskers have lower academic performance. *Computers in Human Behavior*, *53*, 63–70. https://doi.org/10.1016/j.chb.2015.06.027

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bergin, S., Reilly, R., & Traynor, D. (2005). Examining the role of self-regulated learning on introductory programming performance. *Proceedings of the First International Workshop on Computing Education Research*, 81–86. https://doi.org/10.1145/1089786.1089794

Bhaduri, S. (2018). *NLP in Engineering Education-Demonstrating the Use of Natural Language Processing Techniques for Use in Engineering Education Classrooms and Research* [Ph.D. Thesis]. Virginia Tech.

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

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, *31*(6), 445–457. https://doi.org/10.1016/S0883-0355(99)00014-2

Borrego, M., & Bernhard, J. (2011). The Emergence of Engineering Education Research as an Internationally Connected Field of Inquiry. *Journal of Engineering Education*, *100*(1), 14–47. https://doi.org/10.1002/j.2168-9830.2011.tb00003.x

Bouffard-Bouchard, T., Parent, S., & Larivée, S. (1991). Influence of self-efficacy on self-regulation and performance among junior and senior high-school age students. *International Journal of Behavioral Development*, *14*(2), 153–164. https://doi.org/10.1177/016502549101400203

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How People Learn: Brain, Mind, Experience, and School. Expanded Edition*. National Academies Press, 2102 Constitution Avenue N.

Britton, B. K., & Tesser, A. (1991). Effects of time-management practices on college grades. *Journal of Educational Psychology*, *83*, 405–410. https://doi.org/10.1037/0022-0663.83.3.405

Bui, T. T. H., Kaur, A., & Trang Vu, M. (2021). Effectiveness of technology-integrated project-based approach for self-regulated learning of engineering students. *European Journal of Engineering Education*, *0*(0), 1–15. https://doi.org/10.1080/03043797.2021.2016636

Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 160–172). Springer. https://doi.org/10.1007/978-3-642-37456-2_14

Carpenter, T. S., Beall, L. C., & Hodges, L. C. (2020). *Using the LMS for Exam Wrapper Feedback to Prompt Metacognitive Awareness in Large Courses*. https://doi.org/10.14434/jotlt.v9i1.29156

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, *13*(2), 145–182. https://doi.org/10.1207/s15516709cog1302_1

Chew, K. J., Chen, H. L., Rieken, B., Turpin, A., & Sheppard, S. (2016). Improving students' learning in statics skills: Using homework and exam wrappers to strengthen self-regulated learning. *123rd ASEE Annual Conference and Exposition, June 26, 2016 - June 29, 2016*, *2016-June*.

Chowdhury, T., & Murzi, H. (2019). *Literature Review: Exploring Teamwork in Engineering Education*.

Cleary, T. J., & Zimmerman, B. J. (2004). Self-Regulation Empowerment Program: A School-Based Program to Enhance Self-Regulated and Self-Motivated Cycles of Student Learning. *Psychology in the Schools*, *41*(5), 537–550. https://doi.org/10.1002/pits.10177

Clough, G. W. (2005). *The engineer of 2020: Visions of engineering in the new century*. National Academy of Engineering

Cunningham, P., Matusovich, H. M., Hunter, D.-A. N., Blackowski, S. A., & Bhaduri, S. (2017, June 24). *Board # 28: Beginning to Understand Student Indicators of Metacognition*. 2017 ASEE Annual Conference & Exposition. https://peer.asee.org/board-28-beginning-to-understand-student-indicators-of-metacognition

De Lin, O., Gottipati, S., Ling, L. S., & Shankararaman, V. (2021). Mining Informal amp; Short Student Self-Reflections for Detecting Challenging Topics – A Learning Outcomes Insight Dashboard. *2021 IEEE Frontiers in Education Conference (FIE)*, 1–9. https://doi.org/10.1109/FIE49875.2021.9637181

Du Bois, N. F., & Staley, R. K. (1997). A Self-Regulated Learning Approach to Teaching Educational Psychology. *Educational Psychology Review*, *9*(2), 171–197.

Dembo, M. H., & Seli, H. (2008). *Motivation and learning strategies for college success: A self-management approach* (3rd ed). Lawrence Erlbaum Associates.

Dembo, M. H. (2013). *Motivation and Learning Strategies for College Success: A Self-Management Approach*. Routledge.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. https://doi.org/10.48550/arXiv.1810.04805

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Eissa, M. A. (2015). Fathi Abdul Hamid Abdul Kader , PhD. *Educational Sciences*, 8.

El Bojairami, I., Kouthouridis, S., Sobhanigavgani, Z., Jacques, M., Moraes, C., Omelon, S., Driscoll, M., & Chen, L. R. (2019). *Utility and Implementation of Reflection in Distinct Types of Engineering Courses*.

Ellestad, R. M. (2016). The Impact of Teaching Self-Regulated Learning Skills to First-Year Engineering Students. *2016 ASEE Annual Conference & Exposition*.

Falkner, K., Vivian, R., & Falkner, N. J. G. (2014). Identifying computer science self-regulated learning strategies. *Proceedings of the 2014 Conference on Innovation & Technology in Computer Science Education*, 291–296. https://doi.org/10.1145/2591708.2591715

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Gamieldien, Y., Case, J. M., & Katz, A. (2023). *Advancing Qualitative Analysis: An Exploration of the Potential of Generative AI and NLP in Thematic Coding* (SSRN Scholarly Paper No. 4487768). https://doi.org/10.2139/ssrn.4487768

Gamieldien, Y., McCord, R., & Katz, A. (2023). *Utilizing Natural Language Processing to Examine Self-Reflections in Self-Regulated Learning* (SSRN Scholarly Paper No. 4487795). https://doi.org/10.2139/ssrn.4487795

Ganesh, A., Scribner, H., Singh, J., Goodman, K., Hertzberg, J., & Kann, K. (2022). Response Construct Tagging: NLP-Aided Assessment for Engineering Education. *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*, 250–261. https://aclanthology.org/2022.bea-1.29

Gezer-Templeton, P. G., Mayhew, E. J., Korte, D. S., & Schmidt, S. J. (2017). Use of Exam Wrappers to Enhance Students' Metacognitive Skills in a Large Introductory Food Science and Human Nutrition Course. *Journal of Food Science Education, 16*(1), 28–36. https://doi.org/10.1111/1541-4329.12103

Grandoit, E., Bergdoll, R., Rosales, E., Turbeville, D., Mayer, S., & Horowitz, G. (2020). Exploring Organic Chemistry I Students' Responses to an Exam Wrappers Intervention. *Journal of the Scholarship of Teaching and Learning*, *20*(1), 105–129.

Grohs, J. R., Knight, D. B., Young, G. D., & Soledad, M. M. (2018). Exploring Academic Performance Paths and Student Learning Strategies in a Large Foundational Engineering Course. *International Journal of Education in Mathematics, Science, and Technology*, 241–253. https://doi.org/10.18404/ijemst.428175

Guimond, S., Begin, G., & Palmer, D. L. (1989). Education and Causal Attributions: The Development of "Person-Blame" and "System-Blame" Ideology. *Social Psychology Quarterly*, *52*(2), 126–140. https://doi.org/10.2307/2786912

Gynnild, V., Holstad, A., & Myrhaug, D. (2008). Identifying and promoting self-regulated learning in higher education: Roles and responsibilities of student tutors. *Mentoring & Tutoring: Partnership in Learning*, *16*(2), 147–161. https://doi.org/10.1080/13611260801916317

Habib, A., Abdullatif, M., & Alzayani, N. J. (2021). Use of Rubric and Assessment to Encourage Self-Regulated Learning. *2021 IEEE Integrated STEM Education Conference (ISEC)*, 195–200. https://doi.org/10.1109/ISEC52395.2021.9763942

Haller, S., Aldea, A., Seifert, C., & Strisciuglio, N. (2022). *Survey on Automated Short Answer Grading with Deep Learning: From Word Embeddings to Transformers*.

Harding, S.-M. (2018). *Self-regulated learning in the classroom*.

He, H., He, Y., Wang, F., & Zhu, W. (2022). Improved K-means algorithm for clustering non-spherical data. *Expert Systems*, *39*(9), e13062. https://doi.org/10.1111/exsy.13062

Higgins, N. L., Rathner, J. A., & Frankland, S. (2021). Development of self-regulated learning: A longitudinal study on academic performance in undergraduate science. *Research in Science & Technological Education*, *0*(0), 1–25. https://doi.org/10.1080/02635143.2021.1997978

Hodges, L. C., Beall, L. C., Anderson, E. C., Carpenter, T. S., Cui, L., Feeser, E., Gierasch, T., Nanes, K. M., Perks, H. M., & Wagner, C. (2020). Effect of Exam Wrappers on Student Achievement in Multiple, Large STEM Courses. *Journal of College Science Teaching*, *50*(1), 69–79.

Hofer, B. K., & Yu, S. L. (2003). Teaching Self-Regulated Learning Through a "Learning to Learn" Course. *Teaching of Psychology*, *30*(1), 30–33. https://doi.org/10.1207/S15328023TOP3001_05

Jonassen, D., Strobel, J., & Lee, C. B. (2006). Everyday Problem Solving in Engineering: Lessons for Engineering Educators. *Journal of Engineering Education*, *95*(2), 139–151. https://doi.org/10.1002/j.2168-9830.2006.tb00885.x

Karpicke, J. D., & Aue, W. R. (2015). The Testing Effect Is Alive and Well with Complex Materials. *Educational Psychology Review*, *27*(2), 317–326. https://doi.org/10.1007/s10648-015-9309-3

Katz, A., Norris, M., Alsharif, A. M., Klopfer, M. D., Knight, D. B., & Grohs, J. R. (2021, July 26). *Using Natural Language Processing to Facilitate Student Feedback Analysis*. 2021 ASEE Virtual Annual Conference Content Access. https://peer.asee.org/using-natural-language-processing-to-facilitate-student-feedback-analysis

Katz, A., Shakir, U., & Chambers, B. (2023). *The Utility of Large Language Models and Generative AI for Education Research* (arXiv:2305.18125). arXiv. https://doi.org/10.48550/arXiv.2305.18125

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2022). Natural language processing: State of the art, current trends, and challenges. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-022-13428-4

Kim, K. R., & Seo, E. H. (2015). The relationship between procrastination and academic performance: A meta-analysis. *Personality and Individual Differences*, *82*, 26–33. https://doi.org/10.1016/j.paid.2015.02.038

Komarraju, M., Karau, S. J., & Schmeck, R. R. (2009). Role of the Big Five personality traits in predicting college student's academic Motivation and Achievement. *Learning and Individual Differences*, *19*(1), 47–52. https://doi.org/10.1016/j.lindif.2008.07.001

Kosnin, A. M. (2007). Self-regulated learning and academic achievement in Malaysian undergraduates. *International Education Journal*, *8*(1), 221–228. Scopus.

Kuo, E., Hull, M. M., Gupta, A., & Elby, A. (2013). How students blend conceptual and formal mathematical reasoning in solving physics problems. *Science Education*, *97*(1), 32–57. https://doi.org/10.1002/sce.21043

Lakretz, Y., Dehaene, S., & King, J.-R. (2020). What Limits Our Capacity to Process Nested Long-Range Dependencies in Sentence Comprehension? *Entropy*, *22*(4), Article 4. https://doi.org/10.3390/e22040446

Lawanto, O., Santoso, H., Lawanto, K., & Goodridge, W. (2014). Self-Regulated Learning Skills and Online Activities Between Higher and Lower Performers on a Web Intensive

Undergraduate Engineering Course. *The Journal of Educators Online, 11*(3). https://doi.org/10.9743/JEO.2014.3.2

Lay, C. H., & Schouwenburg, H. C. (1993). Trait procrastination, time management, and academic behavior. *Journal of Social Behavior & Personality, 8,* 647–662.

Liao, J., Kunberger, T., Papkov, G. I., Badir, A., O'Neill, R., & Nguyen, L. D. (2018, June 23). *Exam Wrappers, Reflection, and Student Performance in Engineering Mechanics*. 2018 ASEE Annual Conference & Exposition. https://peer.asee.org/exam-wrappers-reflection-and-student-performance-in-engineering-mechanics

Litzinger, T. A., Lattuca, L. R., Hadgraft, R. G., Newstetter, W. C., Alley, M., Atman, C., DiBiasio, D., Finelli, C., Diefes-Dux, H., Kolmos, A., Riley, D., Sheppard, S., Weimer, M., & Yasuhara, K. (2011). Engineering education and the development of expertise. *Engineering Education*, *100*(1), 123–150. https://doi.org/10.1002/j.2168-9830.2011.tb00006.x

Lovett, M. C. (2013). Make exams worth more than the grade: Using exam wrappers to promote metacognition. In: Kaplan, M., Silver, N., LaVague-Manty, D., Meizlish, D. (eds) Using reflection and metacognition to improve student learning: Across the disciplines, across the academy. In *Using reflection and metacognition to improve student learning: Across the disciplines, across the academy,* (pp. 18–52).

Magooda, A., Litman, D., Ashraf, A., & Menekse, M. (2022). Improving the Quality of Students' Written Reflections Using Natural Language Processing: Model Design and Classroom Evaluation. In M. M. Rodrigo, N. Matsuda, A. I. Cristea, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (pp. 519–525). Springer International Publishing. https://doi.org/10.1007/978-3-031-11644-5_43

McCaskey, T. L. (2014). Understanding Student Preparation of Exam Note Sheets. *2013 Physics Education Research Conference Proceedings*, 249–252. https://doi.org/10.1119/perc.2013.pr.050

McInnes, L., & Healy, J. (2017). Accelerated Hierarchical Density Based Clustering. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 33–42. https://doi.org/10.1109/ICDMW.2017.12

McInnes, L., Healy, J., & Melville, J. (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* (arXiv:1802.03426). arXiv. http://arxiv.org/abs/1802.03426

Menekse, M. (2020). The Reflection-Informed Learning and Instruction to Improve Students' Academic Success in Undergraduate Classrooms. *Journal of Experimental Education*, *88*(2), 183–199. https://doi.org/10.1080/00220973.2019.1620159

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space* (arXiv:1301.3781). arXiv. https://doi.org/10.48550/arXiv.1301.3781

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook, 2nd ed* (pp. xiv, 338). Sage Publications, Inc.

Nager, A., & Atkinson, R. D. (2016). *The Case for Improving U.S. Computer Science Education* (SSRN Scholarly Paper No. 3066335). https://doi.org/10.2139/ssrn.3066335

Nehyba, J., & Štefánik, M. (2022). Applications of deep language models for reflective writings. *Education and Information Technologies,* 1–39.

Nelson, K. G., Shell, D. F., Husman, J., Fishman, E. J., & Soh, L.-K. (2015). Motivational and Self-Regulated Learning Profiles of Students Taking a Foundational Engineering Course. *Journal of Engineering Education*, *104*(1), 74–100. https://doi.org/10.1002/jee.20066

Nilson, L. B., & Zimmerman, B. J. (2013). *Creating Self-Regulated Learners: Strategies to Strengthen Students' Self-Awareness and Learning Skills* (1st edition). Stylus Publishing.

Nota, L., Soresi, S., & Zimmerman, B. J. (2004). Self-regulation and academic achievement and resilience: A longitudinal study. *International Journal of Educational Research*, *41*(3), 198–215. https://doi.org/10.1016/j.ijer.2005.07.001

O'Brien, W. K. (2002). *Applying the transtheoretical model to academic procrastination* (Vol. 62, Issues 11-B, p. 5359). ProQuest Information & Learning.

Paechter, M., Maier, B., & Macher, D. (2010). Students' expectations of, and experiences in e-learning: Their relation to learning achievements and course satisfaction. *Computers & Education*, *54*(1), 222–229. https://doi.org/10.1016/j.compedu.2009.08.005

Panadero, E. (2017). A Review of Self-regulated Learning: Six Models and Four Directions for Research. *Frontiers in Psychology*, *8*, 422. https://doi.org/10.3389/fpsyg.2017.00422

Panadero, E., & Alonso-Tapia, J. (2014). ¿Cómo autorregulan nuestros alumnos? Modelo de Zimmerman sobre estrategias de aprendizaje. *Anales de Psicología*, *30*(2), 450–462. https://doi.org/10.6018/analesps.30.2.167221

Paquin, J., Barron, J., & Miller, M. (2020). Efficacy of Learning with Course-provided Equation Reference Sheets in Engineering Education. *2020 Fall ASEE Mid-Atlantic Section Meeting*, 1–15. https://www.jee.org/36046

Passow, H. J., & Passow, C. H. (2017). What Competencies Should Undergraduate Engineering Programs Emphasize? A Systematic Review. *Journal of Engineering Education*, *106*(3), 475–526. https://doi.org/10.1002/jee.20171

Pedrosa, D., Cravino, J., Morgado, L., & Barreira, C. (2017). Self-regulated learning in higher education: Strategies adopted by computer programming students when supported by the SimProgramming approach. *Production*, *27*(Specialissue). Scopus. https://doi.org/10.1590/0103-6513.225516

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist*, *37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Perry, N. E., Hutchinson, L., & Thauberger, C. (2008). Talking about teaching self-regulated learning: Scaffolding student teachers' development and use of practices that promote self-regulated learning. *International Journal of Educational Research*, *47*(2), 97–108. https://doi.org/10.1016/j.ijer.2007.11.010

Pintrich, P. R. (1995). Understanding self-regulated learning. *New Directions for Teaching and Learning*, *1995*(63), 3–12. https://doi.org/10.1002/tl.37219956304

Pintrich, P. R. (2000). The Role of Goal Orientation in Self-Regulated Learning. In *Handbook of Self-Regulation* (pp. 451–502). Elsevier. https://doi.org/10.1016/B978-012109890-2/50043-3

Pressley, M. (1990). *Cognitive strategy instruction improves children's academic performance* (p. 203). Brookline Books.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). *Learning Transferable Visual*

Models From Natural Language Supervision (arXiv:2103.00020). arXiv. https://doi.org/10.48550/arXiv.2103.00020

Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting Self-Regulation in Science Education: Metacognition as Part of a Broader Perspective on Learning. *Research in Science Education*, *36*(1), 111–139. https://doi.org/10.1007/s11165-005-3917-8

Schunk, D. H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist*, *25*(1), 71–86.

Schunk, D. H. (2008). Metacognition, Self-Regulation, and Self-Regulated Learning: Research Recommendations. *Educational Psychology Review*, *20*(4), 463–467. https://doi.org/10.1007/s10648-008-9086-3

Schunk, D. H., & Zimmerman, B. (2011). *Handbook of Self-Regulation of Learning and Performance*. Taylor & Francis.

Schunk, D. H., & Zimmerman, B. J. (1998). *Self-regulated Learning: From Teaching to Self-reflective Practice*. Guilford Press.

Schunk, D. H., & Zimmerman, B. J. (2008). *Motivation and Self-regulated Learning: Theory, Research, and Applications*. Routledge.

Silverajah, V. S. G., Wong, S. L., Govindaraj, A., Khambari, M. N. Md., Rahmat, R. W. B. O. K., & Deni, A. R. M. (2022). A Systematic Review of Self-Regulated Learning in Flipped Classrooms: Key Findings, Measurement Methods, and Potential Directions. *IEEE Access*, *10*, 20270–20294. https://doi.org/10.1109/ACCESS.2022.3143857

Stothart, C., Mitchum, A., & Yehnert, C. (2015). The attentional cost of receiving a cell phone notification. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(4), 893–897. https://doi.org/10.1037/xhp0000100

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). *Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models* (arXiv:2102.02503). arXiv. https://doi.org/10.48550/arXiv.2102.02503

Tilfarlioglu, F. Y., & Delbesoglugil, A. B. Ö. (2014). Questioning academic success through self-regulation, self-esteem, and attitude in foreign language learning (a case study). *Theory and Practice in Language Studies*, *4*(11), 2223.

Topping, K. J. (1996). The Effectiveness of Peer Tutoring in Further and Higher Education: A Typology and Review of the Literature. *Higher Education*, *32*(3), 321–345.

Tsai, D. C. L., Huang, A. Y. Q., Lu, O. H. T., & Yang, S. J. H. (2021). Automatic Question Generation for Repeated Testing to Improve Student Learning Outcome. *2021 International Conference on Advanced Learning Technologies (ICALT)*, 339–341. https://doi.org/10.1109/ICALT52272.2021.00108

Vaiyavutjamai, P., & Clements, M. (2006). Effects of Classroom Instruction on Students' Understanding of Quadratic Equations. *Mathematics Education Research Journal*, *18*. https://doi.org/10.1007/BF03217429

Van Der Meer, J., Jansen, E., & Torenbeek, M. (2010). 'It's almost a mindset that teachers need to change': First-year students' need to be inducted into time management. *Studies in Higher Education*, *35*(7), 777–791. https://doi.org/10.1080/03075070903383211

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Wallin, P., & Adawi, T. (2018). The reflective diary is a method for the formative assessment of self-regulated learning. *European Journal of Engineering Education*, *43*(4), 507–521. https://doi.org/10.1080/03043797.2017.1290585

Wang, R., Wei, S., Ohland, M., & Ferguson, D. (2019, October 25). *Natural language processing system for self-reflection and peer-evaluation*.

Wolters, C. A., Won, S., & Hussain, M. (2017). Examining the relations of time management and procrastination within a model of self-regulated learning. *Metacognition and Learning*, *12*(3), 381–399. https://doi.org/10.1007/s11409-017-9174-1

Wulff, P., Mientus, L., Nowak, A., & Borowski, A. (2022). Utilizing a Pretrained Language Model (BERT) to Classify Preservice Physics Teachers' Written Reflections. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-022-00290-6

Xian, Y., Schiele, B., & Akata, Z. (2017). Zero-shot learning-the good, the bad, and the ugly. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4582–4591.

Yin, W., Hay, J., & Roth, D. (2019). *Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach* (arXiv:1909.00161). arXiv. https://doi.org/10.48550/arXiv.1909.00161

Zeidner, M. (1998). *Test anxiety: The state of the art* (pp. xxi, 440). Plenum Press.

Zhang, Y., Gong, L., & Wang, Y. (2005). An improved TF-IDF approach for text classification. *Journal of Zhejiang University SCIENCE*, *6*(1), 49–55. https://doi.org/10.1631/jzus.2005.A0049

Zimmerman, B. J. (1995). Self-regulation involves more than metacognition: A social cognitive perspective. *Educational Psychologist*, *30*(4), 217–221. https://doi.org/10.1207/s15326985ep3004_8

Zimmerman, B. J. (2000). Chapter 2 - Attaining Self-Regulation: A Social Cognitive Perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Zimmerman, B. J. (2013). From Cognitive Modeling to Self-Regulation: A Social Cognitive Career Path. *Educational Psychologist*, *48*(3), 135–147. https://doi.org/10.1080/00461520.2013.794676

Zimmerman, B. J., & Campillo, M. (2003). Motivating Self-Regulated Problem Solvers. In J. E. Davidson & R. J. Sternberg (Eds.), *The Psychology of Problem Solving* (pp. 233–262). Cambridge University Press. https://doi.org/10.1017/CBO9780511615771.009

Zimmerman, B. J., & Moylan, A. R. (2009). Self-Regulation: Where Metacognition and Motivation Intersect. In *Handbook of Metacognition in Education*. Routledge.

Zimmerman, B. J., & Schunk, D. H. (2001). *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*. Routledge.

Zumbrunn, S., Tadlock, J., & Roberts, E. (2011). Encourage Self Regulated Learning in the Classroom. *MERC Publications*. https://scholarscompass.vcu.edu/merc_pubs/18

Chapter 5: Discussion and Conclusions

5.1 Introduction

In this final chapter, I aim to situate the findings from my research with existing knowledge in the domain, thereby addressing the pivotal research questions that initially inspired this study. I shall outline the implications of these discoveries for both research and practice within the realms of SRL and NLP. The overarching purpose of this dissertation revolved around an investigation into how the state-of-the-art, transformer-based NLP can illuminate the tactics and SRL constructs employed by first-year engineering physics students in their exam wrappers. This endeavor aimed to bridge a methodological gap evident in existing literature, focusing specifically on the analysis of large corpora of qualitative datasets and probing the potential of NLP and LLMs in mitigating this challenge. The investigation further aimed at using this NLP method to explore the distribution of students' learning strategies and SRL constructs from different performance profiles across different exams. To achieve these objectives, I undertook three distinct studies, each involving an investigation of first-year engineering physics students' responses to exam wrappers across three varied exams distributed over several semesters.

Manuscript 1 was designed to advance qualitative analysis by exploring how NLP can be utilized in creating a qualitative codebook. I achieved this by using first-year engineering students' responses to two exam wrapper questions and comparing three methods of codebook generation: (1) traditional qualitative analysis, (2) NLPCA, and (3) NLPGPT. The results of this manuscript offer readers an understanding of how NLP can expedite the process of qualitative analysis and assist in uncovering more intricate insights from the data under scrutiny. In Manuscript 2, the focus was on evaluating a method for categorizing student exam wrapper responses using ZSL, an innovative NLP technique. This study provided a strategy for evaluating the accuracy of an NLP workflow, boosting a researcher's confidence in the classification results produced by the ZSL NLP workflow for the analysis of exam wrapper responses in terms of strategies and SRL constructs. Reassured by the potential of transformer-based NLP for qualitative analysis, as demonstrated in Manuscripts 1 and 2, I combined these findings with student grades to investigate their strategies and SRL constructs in responses to first-year engineering physics exam wrappers in Manuscript 3. I employed student grades in the three module exams to develop exam performance profiles and to identify any correlations between these profiles and students' strategy usage. This study revealed differences in the challenges faced by students from different performance profiles, as well as their exam preparation strategies across exams. However, the distribution of these challenges and exam preparation strategies remained relatively consistent across all performance profiles. These findings further informed the discussion on the SRL constructs they exhibited. More implications for these students can be found in the following sections.

5.2 Implications

Each study I conducted as part of this dissertation resulted in implications for research and practice. These implications are discussed in the following sections. I provide an overview of the problems being addressed in this dissertation, the research focus, the aim of three manuscripts, and the research outcomes of each manuscript in Figure 5.1.

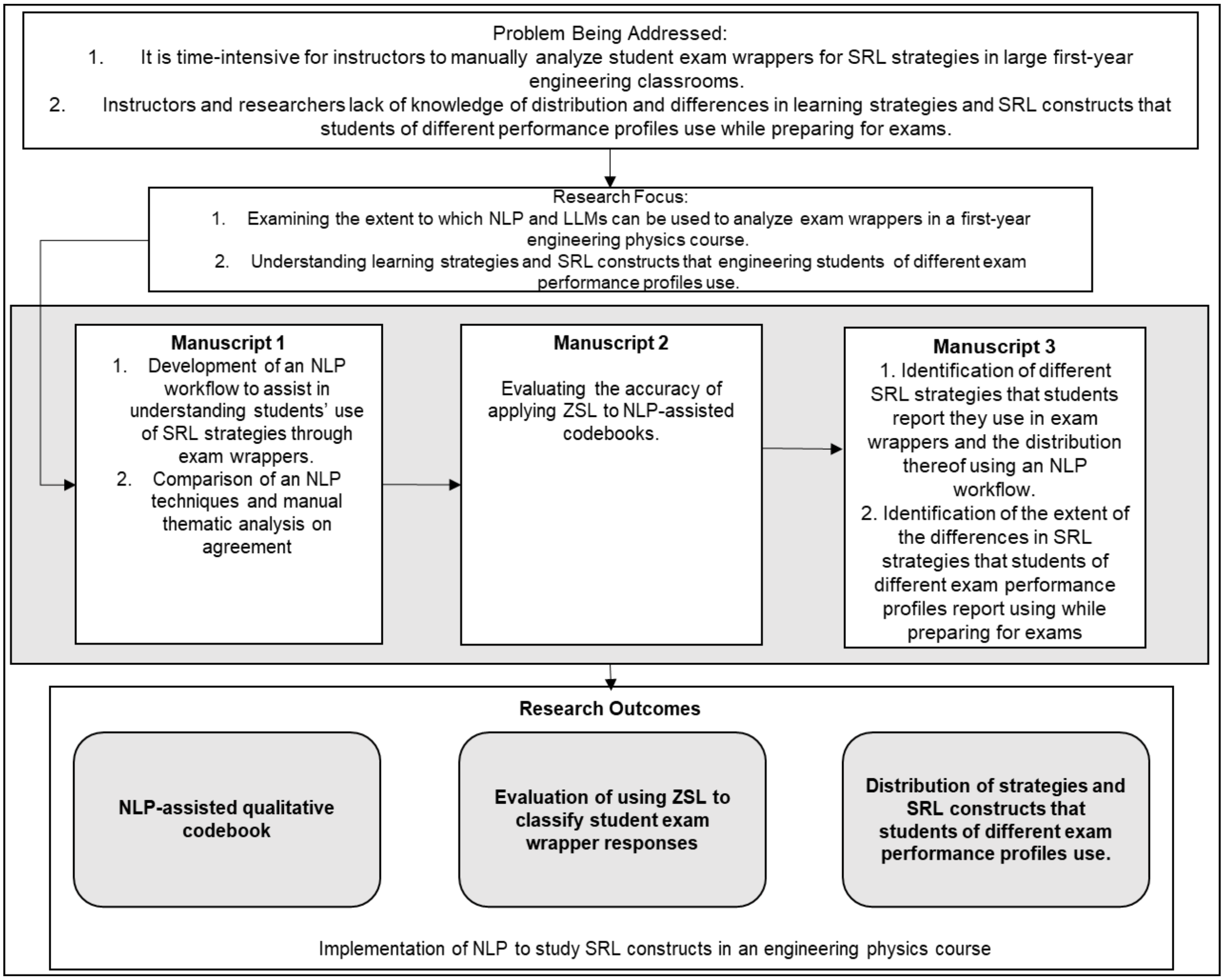


Figure 5.1 Dissertation research focus with an overview of problems being addressed and research outcomes

5.2.1 Implications for Practice

Manuscript 1 offered valuable insights into the potential benefits of harnessing NLP and LLMs for qualitative codebook generation. Upon comparing manual techniques with NLP methods, I discovered that NLP-assisted approaches were capable of delivering a high degree of detail, which could lead to more nuanced patterns in the data when compared to a traditional manual qualitative analysis approach. This level of granularity might encourage instructors to further investigate complex phenomena when examining SRL and possibly implement findings in their classrooms (Borrego & Henderson, 2014). Moreover, automating certain aspects of qualitative data analysis could improve efficiency and scalability. This could allow practitioners to concentrate more on a deeper analysis of students' strategies, struggles, and SRL constructs - tasks uniquely suited to human cognition (Shaddock, 2014). The advent of ChatGPT, built on the same technology as GPT-3.5, has made the use of LLMs accessible to practitioners without a background in coding. While this is a major benefit to ChatGPT, the concerns include data privacy issues, and the risk of biased and inaccurate content (Sallam et al., 2023).

Manuscript 2 revealed that an extensive corpus of exam wrapper data can be analyzed accurately using the ZSL NLP workflow. This methodology provides a viable solution for instructors of large classes who wish to identify the most and least common learning strategies employed by their students. Given the considerable time investment traditionally required for instructors to read and assess qualitative student submissions (Al Yahmady & Al Abri, 2013), the ZSL method significantly alleviates this challenge. Additionally, the potential fatigue associated with analyzing lengthy and often repetitive student work can negatively impact an instructor's ability to consistently assess the full breadth of student data (Bergin, 2011). The ZSL approach can assist instructors in analyzing all the sections they teach, providing insights into pedagogical adjustments, identifying areas for improvement, and illuminating which SRL strategies students need to adopt while preparing for exams. Another advantage of the NLP ZSL workflow is its potential to expedite feedback to students. Traditional feedback can be delayed due to the time an instructor requires to grade the exam wrapper, while the NLP ZSL workflow can provide more timely responses that can be beneficial to students since the process of classifying students' textual data can be expedited to provide more on-time feedback. Timely feedback is associated with a better perception of the constructiveness of the feedback (Bayerlein, 2014). This not only benefits students but could also allow instructors to adjust their courses based on insights gleaned from the exam wrappers.

Drawing from the outcomes of Manuscripts 1 and 2, Manuscript 3 identified unique learning strategies exhibited by students when discussing the non-physics skills required for optimal exam performance and the preparation process leading up to the exam. Utilizing the NLP ZSL method, I found evidence that students primarily attributed a lack of focus as the main non-physics skill hindering their exam performance, along with test-taking skills and carelessness leading to missed points on exams. In terms of exam preparation strategies, the study revealed that effective time management through early preparation, increased practice exam participation, and equation sheet strategies, such as incorporating example problems, were key strategies used by high-performing students. Not only does this study corroborate insights from existing literature on crucial SRL strategies for exam preparation and performance (Chew et al., 2016; Sebesta & Bray Speth, 2017), but it also provides a distribution of strategies and skills across 3,800 participants. The analysis of this large number of responses provides more confidence in the generalizability of the importance of these strategies within the context of the engineering physics course – something that would be time-consuming and resource-intensive for one instructor to do without the aid of NLP or LLMs. Instructors could incorporate these strategies into their teaching, such as providing focus-enhancing techniques, aiding students with test-taking skills, and promoting effective time management. Finally, while nuances across different performance profiles were noted as found by Grohs et al. (2018), similar strategies emerged in my study as primary ones for instructors to focus on, applicable to students across all performance profiles.

5.2.2 Implications for Research

Natural language processing and LLMs hold potential for research across diverse fields, including SRL research. Manuscript 1 enhances our understanding of the utility of LLMs and NLP in qualitative data analysis. As demonstrated through the comparison of manual qualitative analysis, NLPCA, and NLPGPT, each method was capable of capturing similar concepts. Notably, NLPGPT offered a high degree of code granularity. Granularity refers to how broad or specific a code is when referring to a topic. For example, a highly granular code would be related to a very specific idea whereas a code with low granularity would be a broader theme. Large language models, such as GPT-3.5, boast flexibility, allowing for research in qualitative analysis across various contexts, with Manuscript 1 representing just one potential application. For instance, LLMs could be employed to analyze other forms of unstructured texts, such as teacher notes, interview transcripts, and student essays, thereby facilitating codebook creation (Katz et al., 2023). These LLMs can interpret complex statements with more nuance than older, rule-based approaches. Much like practitioners, researchers can also benefit from the time and resources saved using NLP techniques, allowing them to focus on extracting deeper insights that necessitate human cognition.

Manuscript 2 showcased that a large corpus of exam wrapper data can be effectively analyzed using the NLP ZSL workflow, yielding an acceptable level of accuracy which is above 85% of exact matches between the codes generated in Manuscript 1 via the NLPCA approach and student responses (Miles & Huberman, 1994). Furthermore, the findings from other literature in terms of learning strategies match the strategies I found in Manuscript 2 (Chew et al., 2016; Stephen et al., 2020). This provides further confidence in the ZSL NLP method. The key advantage of the ZSL NLP method, however, lies in its ability to facilitate SRL studies with considerably larger sample sizes than those in previous studies. For instance, Chew et al. (2016) utilized a sample of only 69 students, a size that does not reflect an accurate representation of larger class populations. Employing ZSL NLP could incentivize researchers to gather and scrutinize larger data samples, potentially leading to more generalizable findings. This could have implications not only for SRL in first-year engineering physics students preparing for exams but also in various other contexts using exam wrappers including computer science (Carpenter et al., 2020; Craig et al., 2016; Davis, 2021), chemistry (Grandoit et al., 2020), food sciences (Gezer-Templeton et al., 2017) and other STEM courses (Greco, 2012; Hodges et al., 2020; Liao et al., 2018).

Manuscript 3 has research implications for a variety of stakeholders engaged in educational research in engineering education and SRL. The study suggested that students mostly focus on the SRL constructs of Task Strategies, Adaptive Reactions, Causal Attributions, and Time Management. Other learning approaches highlighted by the study are the value of early preparation, practice exams, equation sheet strategies, focusing strategies, and test-taking strategies. Each of these forms part of the aforementioned SRL constructs in Zimmerman's model (Zimmerman & Moylan, 2009). This study paves the way for studying SRL constructs in other educational settings, including the use of exam performance profiles and observing the evolution of SRL constructs students use and discuss through time.

5.3 Future Work

The outcomes from each manuscript, and the overarching implications of this dissertation, provide opportunities for future work in the space of SRL, NLP, and LLMs in many contexts including engineering education. Manuscript 1 highlights potential avenues for future research from a methodological standpoint. The refinement and application of NLP tools in qualitative research still have many open questions to explore. For instance, examining the potential uses of NLP in domains beyond the context of first-year engineering physics and SRL could offer invaluable insights (Liu et al., 2019). Another promising area for investigation lies in the field of prompt engineering, specifically testing the effects different prompts may have on the qualitative codes generated (Liu et al., 2023). How LLMs like GPT-3.5 are prompted can impact their outputs and potentially alter the codes or summaries the models produce (White et al., 2023; Xiao et al., 2023). Alternatively, exploring different combinations of NLP-assisted approaches might be an interesting direction for future work.

In Manuscript 1, we utilized NLP for clustering and manually labeling codes, while also testing a method of having GPT-3.5 label the responses for codes. These represent just two approaches for leveraging NLP in codebook generation. Another potential approach could involve using GPT-3.5-derived codes as the initial level of codes, and then inputting these codes back into GPT-3.5 to further cluster them into broader codes or themes (Katz et al., 2023). This approach could yield highly granular codes, while also potentially facilitating the production of broader themes by directing the LLM to group similar-topic codes. The question of which NLP technique to use and which LLMs are most effective is still an open area of investigation. In Manuscript 1 we used MPNet and GPT-3.5 with an acceptable agreement to the traditional manual method. However, both of these methods have their strengths and weaknesses for different applications. For example, MPNet has the advantage of being open-source, but requiring coding knowledge in Python whereas GPT-3.5 could be used through the ChatGPT web application but has the downside of data privacy issues. The extent of the application of these models and other models in educational settings is still up for investigation. Large language models and NLP continue to advance rapidly and testing different embedding models for clustering and different generative models such as Anthropic's Claude, Google's Bard, open-source models, or the latest version of the GPT models, GPT-4 could be used for qualitative code generation.

Manuscript 2 resulted in acceptable accuracies of using the ZSL technique for classifying student exam wrapper responses into different categories based on a codebook generated in Manuscript 1. This application of the method was done with a specific theoretical framework within a specific context. This success opens the door for exploring the NLP ZSL technique in different contexts, using different theoretical frameworks. This study used three of the exam wrapper responses as proof of concept, leaving space to apply this method to the other exam wrapper questions. Another promising avenue for this research would be to develop a user-friendly API that enables researchers and instructors to input their codebooks and responses for classification or use a combination of the results from Manuscripts 1 and 2 to develop an API that generates the codebook from the responses in an inductive manner. Alternatively, we could explore ways of using ChatGPT as a tool for this analysis which could provide researchers and instructors with limited coding knowledge with a method to gain insights for research and, or practice. For example, instructors could input the data into ChatGPT to identify recurring themes, concepts, and key points in student data.

Manuscript 3 shed light on a few aspects of students' strategies, impediments, and SRL constructs discussed while preparing for an exam and in preparation for subsequent exams. While these results offer interesting insights, there is room to use other data sources to corroborate or enhance the findings of this study. For example, observational data or interviews could complement the data we used in this study. These additional sources could provide more depth and context, allowing for an enhanced understanding of SRL. These additional data sources could also be explored using the NLP techniques we used in this dissertation, but other tools and techniques could be tested in future studies.

5.4 Concluding Remarks

This research study arose from my dual passions in the realm of education. Firstly, I have always been intrigued by the process of learning - how individuals acquire knowledge and skills that they carry beyond the academic setting. This curiosity has its roots in my upbringing, where I witnessed the transformative effects of education firsthand. My siblings and I were raised by our mother who did not finish school and was of relatively poor socio-economic status. Through education, we were able to end some of the financial burdens we had through education and the careers our education afforded us. Therefore, I was naturally drawn toward understanding SRL and investigating ways to foster these skills within the classroom environment because this could help other students succeed in school and beyond. My second area of interest emerged from my experience as an educator. I have always believed that teachers often spend an excessive amount of time on grading and other tasks that could potentially be automated. The recent advancements in NLP and LLMs have introduced the capability to analyze qualitative data with a depth and nuance previously unachievable before the advent of transformer-based NLP models such as BERT and GPT. The time saved could allow instructors more time to devise engaging lessons, interact more with students, and ultimately, maintain a healthier work-life balance.

This work has shown promising results and the increased interest in NLP and LLMs in various fields excites me to continue the journey of using NLP and LLMs to improve outcomes for researchers and practitioners through the lens of SRL and other related theories that focus on the internal processes of how the individual learns, what motivates them, and how they feel as they are learning. I have a passion for teaching and learning and I hope that this work, and my future work, can develop students into critical thinkers, including cultivating their SRL skills which can lead to their success in the future.

References

AlYahmady, H. H., & Al Abri, S. S. (2013). Using Nvivo for Data Analysis in Qualitative Research. *International Interdisciplinary Journal of Education*, *2*(2), 181–186. https://doi.org/10.12816/0002914

Bayerlein, L. (2014). Students' feedback preferences: How do students react to timely and automatically generated assessment feedback? *Assessment & Evaluation in Higher Education*, *39*(8), 916–931. https://doi.org/10.1080/02602938.2013.870531

Bergin, M. (2011). NVivo 8 and consistency in data analysis: Reflecting on the use of a qualitative data analysis program. *Nurse Researcher*, *18*(3), 6–12. https://doi.org/10.7748/nr2011.04.18.3.6.c8457

Borrego, M., & Henderson, C. (2014). Increasing the Use of Evidence-Based Teaching in STEM Higher Education: A Comparison of Eight Change Strategies. *Journal of Engineering Education*, *103*(2), 220–252. https://doi.org/10.1002/jee.20040

Carpenter, T. S., Beall, L. C., & Hodges, L. C. (2020). *Using the LMS for Exam Wrapper Feedback to Prompt Metacognitive Awareness in Large Courses*. https://doi.org/10.14434/jotlt.v9i1.29156

Chew, K. J., Chen, H. L., Rieken, B., Turpin, A., & Sheppard, S. (2016). Improving students' learning in statics skills: Using homework and exam wrappers to strengthen self-regulated learning. *123rd ASEE Annual Conference and Exposition, June 26, 2016 - June 29, 2016*, *2016-June*.

Craig, M., Horton, D., Zingaro, D., & Heap, D. (2016). Introducing and evaluating exam wrappers in CS2. *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, 285–290.

Davis, K. C. (2021, July 26). *Examining the Efficacy of Exam Wrappers in a Computer Science Course*. 2021 ASEE Virtual Annual Conference Content Access. https://peer.asee.org/examining-the-efficacy-of-exam-wrappers-in-a-computer-science-course

Gezer-Templeton, P. G., Mayhew, E. J., Korte, D. S., & Schmidt, S. J. (2017). Use of Exam Wrappers to Enhance Students' Metacognitive Skills in a Large Introductory Food Science and Human Nutrition Course. *Journal of Food Science Education*, *16*(1), 28–36. https://doi.org/10.1111/1541-4329.12103

Grandoit, E., Bergdoll, R., Rosales, E., Turbeville, D., Mayer, S., & Horowitz, G. (2020). Exploring Organic Chemistry I Students' Responses to an Exam Wrappers Intervention. *Journal of the Scholarship of Teaching and Learning*, *20*(1), 105–129.

Greco, E. (2012). *Developing, deploying, and analyzing exam wrappers in a large introductory physics class*. *2012*, R15.001.

Grohs, J. R., Knight, D. B., Young, G. D., & Soledad, M. M. (2018). Exploring Academic Performance Paths and Student Learning Strategies in a Large Foundational Engineering Course. *International Journal of Education in Mathematics, Science and Technology*, 241–253. https://doi.org/10.18404/ijemst.428175

Hodges, L. C., Beall, L. C., Anderson, E. C., Carpenter, T. S., Cui, L., Feeser, E., Gierasch, T., Nanes, K. M., Perks, H. M., & Wagner, C. (2020). Effect of Exam Wrappers on Student

Achievement in Multiple, Large STEM Courses. *Journal of College Science Teaching*, *50*(1), 69–79.
Katz, A., Shakir, U., & Chambers, B. (2023). *The Utility of Large Language Models and Generative AI for Education Research* (arXiv:2305.18125). arXiv. https://doi.org/10.48550/arXiv.2305.18125
Liao, J., Kunberger, T., Papkov, G. I., Badir, A., O'Neill, R., & Nguyen, L. D. (2018, June 23). *Exam Wrappers, Reflection, and Student Performance in Engineering Mechanics*. 2018 ASEE Annual Conference & Exposition. https://peer.asee.org/exam-wrappers-reflection-and-student-performance-in-engineering-mechanics
Liu, M., Shum, S. B., Mantzourani, E., & Lucas, C. (2019). *Evaluating machine learning approaches to classify pharmacy students' reflective statements*. https://opus.lib.uts.edu.au/handle/10453/137785
Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, *55*(9), 195:1-195:35. https://doi.org/10.1145/3560815
Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook, 2nd ed* (pp. xiv, 338). Sage Publications, Inc.
Sallam, M., Salim, N., Barakat, M., & Al-Tammemi, A. (2023). ChatGPT applications in medical, dental, pharmacy, and public health education: A descriptive study highlighting the advantages and limitations. *Narra J*, *3*(1), Article 1. https://doi.org/10.52225/narra.v3i1.103
Sebesta, A. J., & Bray Speth, E. (2017). How Should I Study for the Exam? Self-Regulated Learning Strategies and Achievement in Introductory Biology. *CBE Life Sciences Education*, *16*(2), ar30. https://doi.org/10.1187/cbe.16-09-0269
Shaddock, A. (2014). *Using Data to Improve Learning: A practical guide for busy teachers*. ACER Press.
Stephen, B., Jiang, X., & McArthur, S. D. J. (2020). Extracting Distribution Network Fault Semantic Labels from Free Text Incident Tickets. *IEEE Transactions on Power Delivery*, *35*(3), 1610–1613. Scopus. https://doi.org/10.1109/TPWRD.2019.2947784
White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT* (arXiv:2302.11382). arXiv. http://arxiv.org/abs/2302.11382
Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P.-Y. (2023). Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces* (pp. 75–78). Association for Computing Machinery. https://doi.org/10.1145/3581754.3584136
Zimmerman, B. J., & Moylan, A. R. (2009). Self-Regulation: Where Metacognition and Motivation Intersect. In *Handbook of Metacognition in Education*. Routledge.

Appendices

Appendix A Full Exam Wrapper Questions

Table A1: Exam Wrapper Questions

Part	Question
Exam Reflection	Reflection - What did you do differently between Exam 1 and Exam 2? Did the changes that you made make an impact? Did you reach your goal from the last Exam Wrapper?
Exam Dissection	There are skills other than physics knowledge necessary to complete this exam. Can you identify any skills or fundamental knowledge (non-physics) that are weak that impeded your ability to show what you know about physics concepts? What evidence do you have to backup your answer?
Exam Preparation Reflection	Describe your process for learning/engagement during the regular week for this module. Can you identify any areas of improvement that could strengthen your learning during the regular week moving forward?
Preparation Process	Describe your process for preparing to take the module exam. Can you identify any areas of improvement that could strengthen your preparation activities?
	How confident were you when the exam was passed out that you were ready to show what you knew about this module? What is one thing YOU could do over the next three weeks to support building confidence? What is one thing your instructor could help with to support building your confidence?
Strategic Plan	Define a measurable goal you would like to achieve during our next class module. This goal should be measurable and attainable in the next three-week period.
	Identify one action you want to START doing that may better support your learning in this next module. Can you describe a specific action plan to support you in starting this action?
	Identify one action you want to STOP doing that is detrimental to your learning in this next module. Can you describe a specific action plan to support you in stopping this action?
	How will you plan to celebrate if your goal is achieved?

Appendix B Full Codebook for Manual Qualitative Analysis, NLPCA, and NLPGPT

Table B1. Full Comparison of Codebook Results for Manual Analysis, NLPCA, and NLPGPT for "Preparation Process"

Manual Codebook	**Manual Example**	**NLPCA Codes**	**NLPCA Examples**	**NLPGPT Codes**	**NLPGPT Examples**
Practiced old exams/past tests	In preparing for this exam I just made my equation sheet. In reviewing for the next exam I could also do practice exams.	practice past/old exams	Preparing for the exam I went over many practice exams. My process for preparing for this Exam included a lot of practice exams. My process for preparing for the exam involved looking over practice problems and reviewing previous exam examples. My process for preparing for this exam included printing out various practice exams and completing them first blank, and then going back and checking them with the completed exams.	Practice Exam Emphasis. Practice exams as preparation. Using practice exams for review. Practice exams. Practice exams as preparation. Practice Exam Emphasis. Practice old exams. Reviewing old exams and learning pages. Practice exams for preparation. Using Past Exams for Practice Practice with old exam problems.	Before this exam, I took a couple of practice exams and graded myself on the exams. Before this exam, I took a couple of practice exams and graded myself on the exams. Before the exam I take multiple practice exams without looking at the answers. The most I did was one previous exam and I've learned that they aren't the best representation of what the exam will look like. Taking the previous exams in a timed manner, then working out the questions I got wrong.
Study more and do more practice	I study alot the night before the exams. i could study more over the week to succeed more often.	study more	A way I could strengthen this would be studying the exam questions more intently and spending more time on the practice exams. I can definitely strengthen my preparation by reviewing notes and taking a practice exam each day the week of the exam. To strengthen my preparation activities I could probably do more practice exam questions. I think something I could improve on to strengthen my preparation activities would be to make sure that I have all of the practice problem "write-ups" completed and the learning page notes completed	Problem-solving through practice and video tutorials. Practice Problems. Practice and Review. Improving through Review and Practice. Need for more studying. Practice	to improve, I could've studied longer. I think if I was on time and had more study time, I could have done better. I could have studied more. I definitely could have studied more. I definitely could have studied more. As I would have gotten a better grades if I did do more work and preparation. But I would say If I was able to study I would've done better. I could have focused more directly on

			about a week before the exam, so then my focus is more heavily on reviewing rather than learning. Taking more practice exams could strengthen my preparation activities.	question review. Practice Problem Review Lack of Practice Problems. Practice problem repetition. Practice problem review. Practice exams and problem review.	a few of the subjects on the exam that I knew I would be weak in.
		do more practice exams	I need to spend more time on practice exams. I need to utilize practice exams more thoroughly, and practice well in advance. I think I need to work through more practice exams. I absolutely need to do more practice problems before exams.		
		studied earlier	These exams didn't go over as much as the practice problems so I should have focused more on those. I believe if I actually had done some problems from the practice exams I would have done better on my exam. I took more practice exams which gave me the opportunity to gather the questions that I didn't fully understand for the exam. I should have taken more practice exams. I should have taken more practice exams. I could have taken more practice exams I also studied by doing practice exams, but in hindsight I could have blocked off more time to take practice exams to prepare. I could use practice questions to study and I could have spent more time looking through old exams in order to make sure that I am thoroughly prepared.		
Equation Sheet	I prepared for the last module	make equation sheet	In addition to this, I also look over the practice problems to ensure that I understand how to use the information on the equation sheet.	Equation sheet and practice exams. Last-	I prepared for the exam by making my equation sheet and looking over a few old exams.

	exam by spending time studying alone in the library. I reorganized all of my notes throughout the week and complied an equation sheet of important concepts.		I also fill out my equation sheet during this time by going through the learning modules and the practice questions and adding whatever could possibly be on the exam. I look over my work for practice questions too, and see if there any equations there that were useful. I look over my equation sheet and update it if I need to, and look over the practice questions and my work for them. Then whichever topics that I feel that I need help memorizing , I put examples of problems we have done on my equation sheet. Then whichever topics that I feel that I need help memorizing , I put examples of problems we have done on my equation sheet.	minute studying and equation sheets. Equation sheet preparation. Creating an equation sheet. Using equation sheet for studying. Equation sheet strategy. Practice problems on equation sheet. Improving Equation Sheet Strategy. Improving equation sheet. Creating equation sheet for exams. Practice Exam and Equation Sheet Preparation.	Before taking the module exam, I prepared an equation sheet and worked through a few practice exams. On this module exam, I started by writing my equation sheet, then worked through a full practice exam. I prepared for the exam by filling out my equation sheet and going through past exams. My preparation for the exam began with reading back through the learning pages and pulling important equations for my equation sheet. While studying for the exam, I completed my equation sheet making sure I included every formula from the module, I reworked every homework problem and made sure that I understood the process of solving them, and worked through 5-6 old exams. While preparing for the module exam, I do practice exams with a friend, we write down important equations as needed and transfer them to our equation sheet, then I go over the concept questions.
		put examples problems on equation sheet	When preparing for the exam, I went back through the learning pages and took hand-written notes and rewatched the videos included. I then rewatched videos from the learning pages and watched the video solutions for some past exams and listened to a professor talk through it. To improvement my preparation, I believe I should also look back over the learning pages. I spent time going back throughout the modules but in the future, I may need t put more focus on the videos and past assignments so I can review the concepts in practice and not just in theory.		

Make equation sheet earlier	My preferred method of exam prep is to make my equation sheet and then to test it. Once I make my sheet I begin taking practice exams until I feel that I am comfortable with the material. One area I can improve is making my equation sheet earlier than the day I start studying so that it doesn't cut into my study time.	make equation sheet earlier	I could also make my equation sheet before the day of the test because I think i missed an equation that could have helped me. I didn't start preparing soon enough and I should have made my equation sheet sooner. On this exam I made my equation sheet the night before and only began to study then, which is a horrible prep process. I should've done my equation sheet a lot earlier, though. I also should have started writing my equation sheet earlier. I ended up creating my equation sheet really late and barley studying for this module exam. I probably could make my equation sheet earlier than I did. I also feel like rewriting my equation sheet the night before would help me on the exam.	Creating Equation Sheets Early. Creating Equation Sheets. Creating Equation Sheets in Advance. Last-minute equation sheet. Creating equation sheet throughout module.	To better prepare, I could make my equation sheet earlier and start studying earlier. One thing I could improve upon is writing my equation sheet earlier in advance, so I can practice using it when we take practice exams during lab. I suppose that I could create my equations sheet earlier to study sooner. I could improve by making my equation sheet earlier thus leading to me being able to spend more time studying and less time having to make a sheet last minute. I made the equation sheet earlier and spent more time taking practice exams so that worked well. I need to start studying earlier as well as making my equation sheet sooner.
		make equation sheet and do practice exams	I write my equation sheet and then go back and review and then take practice tests. Then I take a practice test and after I have it completed I add practice problems to my equation sheet. I take practice tests, and use writing examples and filling out the equation sheet as a study source.		
		reviewing notes/learning pages	I go through all learning pages and practice questions the night before the exam. To prepare for the exam, I went over all of the		

			learning pages and studied them for multiple hours. I needed to make sure to study all the time after class. I made sure to study many days in advance so that I would have time to get through everything. In preparing for exams, I work go through every learning page.		
Review Notes	I review notes and learning pages to make my equation sheet, and once I make my equation sheet, I do 2-3 practice exams.	review notes/watch videos	Then I watch videos over the previous exams. and then I watch all of the videos that are associated with previous exams. After that, I usually review old exams by trying some of the problems and watching videos for the ones that have them.	Reviewing and Checking Study Material Reviewing and Practicing Material. Reviewing Course Material. Reviewing skipped material. Reviewing and Relearning Concepts. Comprehensive review and equation sheet creation. Reviewing practice problems and questions. Reviewing notes and practice problems. Review and Practice. Reviewing and Revising. Reviewing and summarizing.	I usually start by looking through the prep material for each section to refresh my mind. Other than that, I believe I have come up with a good process for doing my preparations. My process for preparing involves me skimming over the module the day before trying to refresh on material and hopefully catch anything that I may have forgotten. I go through and check and correct my work. I usually begin by reviewing the preparation activities and learning modules. I check my work afterward, and look over a problem more closely if I am confused or unable to reach a solution.
		review notes/practice questions	I usually take at least two practice exams the night before. I finished all the prep and practice the week before the exam, so that I could start reviewing and going through notes. I finished all the prep and practice the week before the exam, so that I could start reviewing and going through notes. I should do all of the things I have listed more often (multiple times daily) and far in advance from the exam to feel the most confident when taking the exam. I plan to finish all the prep and practice for all the sections a week before the exam.		
Weak strategies	My exam preparation	weak study strategies	However, I had a lot of outside factors prevent me from getting time to do so such as family matters	Procrastination and Time	I feel that if I had studied a little more rather than just right before the exam, I

	process is pretty weak, mostly just consisting of loosely looking over the content, but one major area of improvement that could benefit me immensely would be doing a lot of preparatory practice questions before the exam, particularly to simulate the real thing.		and an incredibly busy week involving my other classes. I didn't have time because of my other classes. I let some things keep me distracted, and honestly, I really did not study much at all. I made the mistake of procrastinating on one of my CS labs and had to stay up late on Tuesday in order to finish it. I did not study very much. I also think that I slacked on meeting with my instructor and I think that harmed me. However, my efforts here were greatly stunted by a large assignment for another class that took up more of my time than expected. however I don't recall being able to go to a study session since it was interfering with my schedule to study for something else. I didn't study super intensely.	Management Last-minute practice exams. Last Minute Studying. Lack of Time Management. Overconfidence and lack of preparation. Cramming for exams. Last-minute cramming. Last-minute practice exams. Limited Practice Testing. Poor Time Management. Inconsistent study habits.	would have been more confident in the material, helping me work out the problems more carefully and faster. I could have studied up sooner to be better prepared, but other than starting my studying sooner I could have started studying earlier and done more. Looking back, I should have realized I had all this work to do and worked better the days before so that I would have more time to work on preparing for the exam. I could have taken more time to prepare for the exam and could have started preparing earlier.
Study earlier more consistently	In order to prepare for the module exam, I just looked at the exams from previous years and went through the questions. I could definitely improve on this studying	study earlier	The areas of improvement that could strengthen my preparation would be to study more and longer, like a week in advance. The areas of improvement that could strengthen my preparation would be to study more and longer, like a week in advance. I think i could strengthen my preparations by focusing more on concept questions and studying earlier I could strengthen my preparation by starting earlier and studying the weekend before. I could strengthen my preparation by starting earlier and studying the weekend before. This way, I get beneficial, focused study time and will be a lot more prepared on the exam.	Starting studying earlier. Practice exams and starting days before exam. Starting early for exams. Starting Early for Exams Starting studying earlier. Starting Early Strategy. Improving Time Management for Practice Exams. Early Preparation	I think the only thing I could improve on is studying more early in advance than right before. An area of improvement could definitely be to start studying earlier, as I did not start soon enough. To improve, I could start studying earlier along with focusing on problems that I struggle with. The one think I can improve upon is to start studying earlier. I could strengthen my preparation by starting my studying further in advance.

	habit by starting to study earlier, looking at the learning pages during the week to keep concepts fresh, and making sure I can grasp the concepts of every possible example problem.		Studying more would improve my preparation activities. I could start a week early in order to look over more practice exams and really strengthen my knowledge of the concepts.	Lack of Consistency in Exam Preparation. Need for Earlier Preparation Early Study Start	
Make study notes	Again, I think I need to go over the videos that I never watched in this section. I think that would be very helpful and is something I plan to do with my upcoming weekend. Other than that, my preparation is the same. I'm taking extensive notes on the	Areas of improvement	To improve I could do more practice or try to find what I have trouble with early and start correcting it now instead of later. An area I could improve is to review through a test to see where my knowledge stands then go back and apply whatever is needed to my weaker areas. This does not take much time and is the area I want to improve in the most. The area that I can improve on would have to be not getting enough practice for the concepts that I don't fully understand. Just doing this has benefitted me, but I definitely want to improve. I think this is where I can improve.	Note-taking and summarizing. Note-taking and Reviewing.	I went back and read through all of the learning pages and took notes on my equation sheet. I reread through all of the learning pages and wrote my equation sheet. I reread through the learning pages and then made my equation sheet all the night before. I went through all the learning pages and took notes and added it to my equation sheet.

	learning pages and the lectures, as well as paying attention to the questions asked during the learning questions and clicker questions. I'm also learning the section through a couple of videos on youtube, which just solidifies the concepts that we go over in class.				
Watch conceptual videos	For exams I make my equation sheet and do a few practice exams and watch the videos for the problems I don't understand. Studying with a friend could help			Note-taking and Video Learning. Learning through Videos and Group Work. Understanding concepts thoroughly. Emphasizing Conceptual Understanding.	I also need to not just look through the learning pages but take more notes. I also think that taking notes when watching the videos on the learning pages would be beneficial. Looking over the learning pages and making sure I understand the topic at hand before moving one. I'm also learning the section through a couple of videos on youtube, which just solidifies the concepts that we go over in class. Next I go and watch videos on some of the content to relearn the basics of it. but I would like to watch some of the

	strengthen that preparation.				conceptual videos for this next exam that are in each learning page. I think I should look through the learning pages more.
Study alone	I tried studying in a group this time but found it to be very distracting specifically because it was a group of people that I didn't know well. In the future, I will know to study alone or with closer friends that I feel more comfortable with.				
Study in focussed environment	I studied with usually one or two other people, and consulted friends when necessary. I also usually studied at the ROTC building, which is	Go to EF study room	one improvement i can make is to also look over practice and learning module problems that i have already solved to remind myself. When I prepare I could also try and work through the homework problems I had trouble with throughout the module and go get help in the EF study room. When I prepare I could also try and work through the homework problems I had trouble with throughout the module and go get help in the EF study room. When I prepare I could also try and work through the	Minimizing Distractions and Time Management.	Just going to study alone without distractions should help this. so I do not have to cram while trying to study. I need to make time in my day to study and not do it whenever possible. I need to work on using myself to study more and take breaks ONLY if necessary.

	often distracting. So far in the new module, I've gone to the EF study room almost daily, and switched my study area to either a quiet floor in Hodges or elsewhere by myself. I believe this will help me focus more.		homework problems I had trouble with throughout the module and go get help in the EF study room. Also, I need to spend more time recap the learning pages in order to perform well on all the concept questions. Also, I need to spend more time recap the learning pages in order to perform well on all the concept questions.		
Group study	The night before the exam i studied in a group with people in my dorm. I should start studying earlier to understand the material better.	Study group	I also feel I need to study with a group to gain more knowledge. I usually study with a group, and it's been helping a lot. I could study with a group, but I feel like my method works for me. I will also join a study group so that we can share our skills and knowledge. I can maybe try a group study which could help.	Collaborative Study. Group Study. Collaborative Study. Group Study Benefits. Studying with a friend. Longer and Group Studying. Study group collaboration.	For the next exam, I would start this process earlier, and maybe form a study group to bounce problems off of. Areas for improvement would be to start looking at in more in depth sooner before the exam I will try working on more problems before this exam. One way I could prepare better is doing these exams with someone else and I can explain to them the problems wherever they are confused.
Review learning pages	I focused mostly on the learning pages and rereading these. I need to spend more time on			Review and Reflect. Reviewing and Practicing Material. Reviewing learning pages. Reviewing class	This allowed me to find my common mistakes and understand the process behind each problem. One thing I could do better is review the notes I take throughout the module. This allowed me to identify which problems I did and did not understand how to do.

	practice exams.			materials and learning pages. Review and Self-Evaluation. Improving Learning Pages Study Strategy. Reviewing Concepts Before Exam. Practice and Concept Review.	Review the homework's and identify which ones gave me the most trouble. If there is a question I think I would have trouble with, I look back over the notes. After each one, I would look over every problem, especially the ones I missed or could not figure out. I also write down any multiple choice questions that caused me trouble.
Confidence	I review all the modules and complete several practice exams. I think once I develop confidence in one module, I should then start to focus the bulk of my attention on modules that I lack confidence in, rather than doing more practice exams and giving each module	confidence	I should've done more but I was overconfident and that turned out to be a bad thing. I completely lost sense of what I was doing prior and that reflected in me losing my common sense. This helped to boost my confidence but my lack of deeper understanding was still prevalent. I tended to check my answers as I went and didn't really time myself. However, it did become too much, and it hindered me. Overall, I believe that this was my downfall. I only took 2 this time around, and I think that little bit I left off was what really killed me this time around. I probably should have studied for a little longer, but I am happy with how it worked out.	Overconfidence in preparation.	Although this was valuable, it also gave me a false sense of security, as shown with problem 13 and problem 14. Although this was valuable, it also gave me a false sense of security, as shown with problem 15 and problem 10. Although this was valuable, it also gave me a false sense of security, as shown with problem 15 and problem 10. I think this was a big mistake on my part because it is a very risky approach. My process was one of the questionable choices. This was part poor planning on my part and part underestimating the testing of my lab.

	equal attention.				
Clicker questions	I start a week before. I practice past exams and study clicker questions. I also rework prep and practice questions. An area where I can improve would be to take more practice exams and test myself on conceptual questions.				
Do past exams and check answers	My preparations were the exact same as last time. I did a couple practice exams then checked my answers. After seeing what I did wrong, I reviewed those	practice exam and check answers	I did a couple practice exams then checked my answers. I did a couple practice exams then checked my answers. After that, I looked at a few practice tests and after completing them I would check my answers. I took a practice exam, and I went through and checked my answers.	Focused practice and self-assessment. Self-assessment of study effectiveness.	I think I need to focus more on what I struggle with after taking one practice exam, instead of taking a ton of practice exams. I do a pretty good job spacing out the practice exams and only focusing on the problems I don't know how to do. I usually practice on previous exams to see what I'm weak on. I don't use the practice tests and I think going through at least one would help me.

	sections and then practiced more. One thing I have neglected to do is study homework problems. That's one more piece that I'm going to add to my preparations.				
Set aside time to study	Honestly, I just need to set aside more time to take more practice exams.	time management	I should have gotten more sleep the night before, and prioritized it over the extra studying I did. This process works well when I give myself enough time to study, but this time I did not. I could not do that and study before hand, that way I am able to get a good night's sleep before the exam. I did not have much time to study so study time is the biggest factor.	Early preparation and longer study sessions. Varying study start times. Exam study timeline. More time studying.	The areas of improvement that could strengthen my preparation would be to study more and longer, like a week in advance. The areas of improvement that could strengthen my preparation would be to study more and longer, like a week in advance. I think i could strengthen my preparations by focusing more on concept questions and studying earlier I could strengthen my preparation by starting earlier and studying the weekend before. I could strengthen my preparation by starting earlier and studying the weekend before. This way, I get beneficial, focused study time and will be a lot more prepared on the exam. Studying more would improve my preparation activities. I could start a week early in order to look over more practice exams and

					really strengthen my knowledge of the concepts.
		improve time management	I think organizing my time could be improved. I also can improve on organizing my time in general. Giving myself more time to work on it would be the best improvement I can see. I struggled to correctly divide my time correctly, and I will focus more on time management between problems. My main area of improvement would have to be the time that I allocate to doing so, as this mostly took place during lecture, as I usually opt-out.		
Focus on harder concepts/questions	Before the exam, I prepare my exam sheet first by adding equations, conversions, and units that may be useful. After this I try to work through 2 exams from previous semesters to gauge my knowledge on each concept. I could try to do more exams, or at least work more	improve conceptual understanding	An improvement that I can think of is maybe looking over the past homework or even looking at the prep so I can understand the topic more if needed. As such, I might be able to see improvement in re-watching lecture videos and taking better notes. I could improve studying by going through the learning pages a bit more carefully to review theoretical concetps. An area of improvement would be to study by myself more to get a better understanding of the material. An improvement could be made on when I watched the previous exam videos.	Practice test focus.	Allowing myself to totally focus on the material has helped me study for the upcoming exam. I could strengthen my preparation by doing a practicetest start to finish without checking the key. I could strengthen my preparation by doing a practicetest start to finish without checking the key. I start the practice exams alone, and if I need help on a question I look at the key to give me some insight on what to do. I need to study more ahead. I need to study more for the final.

	problems on harder concepts so that I understand them better during the exam.				
		doing problems that are hard			
Problem understanding	I did not prepare that much for this exam. I waited till the last minute to look over some notes and to make my equation sheet the night before or two days before. For the next exam, I will start studying more in advance and I will work more problems and try to understand them.			Problem-solving and time management. Practice and Comprehensive Understanding. Breaking down material for better retention. Emphasizing importance of comprehension.	I still need to work on solving the problems faster. I also need to work problems on my own to make sure I remember how to set up the problems because it is so easy to say to yourself "Oh, this is easy, I know how to do this" but completely forget a step or two on the exam when it matters the most. If it looks like the problem will take a longer time, I try to set up the problem or write out the equations that I will be using and then skip to the next problems to see if any of them are easier.

Timed practice exams	1) Re-read the learning pages 2) Add equations to EQ sheet as I get to each check mark 3) Do timed practice exams: lowest class avg to highest class avg	Time practice exams	I will also begin working on the previous exams in a more timely manner to get a better feel of how the exam may be like. I also want to do more practice exams prior to the day before the exam so that I can ask questions on how to do them if I am confused. I can also try and emulate the practice exams as a real exam in order to get a better grasp of time management during the tests. I could strengthen my preparation activities by doing old exam problems earlier during the module, as opposed to just the few days before the exam.	Test-taking strategies.	The only thing that I could've prepared myself more for was multiple-choice, which I failed to apply my knowledge on during this test, costing me to lose a lot of points. I briefly looked at problems from previous exams, but I didn't ensure that I was practicing one from each section. However, I should've taken the time to go through and balance my focus on all topics. I blanked on two problems because I was weaker at that concept than I was at others, where I did well. I skipped through the exams sheets to try to find the easiest questions and left the harder ones for last.
Focus on multiple choice questions	First, I go over each of the summary pages of the learning modules and write down all the major equations on formula sheet. Then I look at the personal notes that I have taken and see if I added any personal equations that would assist me in the exam.	improve on multiple choice	I also go through older practice exams to do the multiple choice questions and go through the prep and lecture concept questions. I also go through older practice exams to do the multiple choice questions and go through the prep and lecture concept questions. I go through older practice exams to do the multiple choice questions and go through the prep and lecture concept questions.	Improving Multiple Choice Skills. Multiple choice question strategy.	Areas I could improve on is going back through the concepts so I could improve on the multiple choice questions. One thing that I could improve upon is taking more multiple choice questions because this is where I typically struggle. I wish I would have refocused more on the multiple-choice questions as an area of improvement because some of them I had no clue on. Any areas of improvement would probably doing more questions with concepts I wasn't as good at.

	Finally, I complete the exam review provided before every exam and review concepts that I personally think might be difficult on the exam. I could improve in these preparation activities by going over more exams from previous years and look at the multiple choice questions closely since that's where I struggle the most.				
Do lab questions	I usually take 3 practice exams, and I cover over half of the concept questions on the other	Do prep/practice questions	I still need to work on solving the problems faster. I also need to work problems on my own to make sure I remember how to set up the problems because it is so easy to say to yourself "Oh, this is easy, I know how to do this" but completely forget a step or two on the exam when it matters the most. If it looks like the problem will take a longer time, I try to set up the problem or write out the		

	practice exams. In addition, I take a look at the lab concept questions as well as they have proved extremely helpful.		equations that I will be using and then skip to the next problems to see if any of them are easier.		
Other commitments	I tried to study a couple days before, but like I said I had 3 other exams around this time and was very flustered. More time.				
Better study habits	In order to prepare for the module exam, I just looked at the exams from previous years and went through the questions. I could definitely improve on this studying habit by	effective study strategies	honestly this approach helps me a lot and is very effective. I think that this works well for me, I just needed to spend more time doing this for the previous test. I think it worked relatively well and do not think I need to improve my method. This I feel worked pretty well and made me confident. I found this to be extremely helpful and boosted my confidence. My exam grade was pretty good so I will continue this strategy. This method works very well for me. These preparation activities worked really well for me and I will be applying them again for the next text.	Pre-Test Breaks Comprehensive study approach. Effective study strategies. Study Strategy Improvement. Exam preparation strategies. Time Management Time Management. Time Management Strategies. Exam Review	I try to make sure all my practice is finished beforehand. Study a little on the day but not a whole lot, maybe like 15-30 minutes so I can give my brain a break for the upcoming test. Study a little on the day but not a whole lot, maybe like 15-30 minutes so I can give my brain a break for the upcoming test. Study a little on the day but not a whole lot, maybe like 15-30 minutes so I can give my brain a break for the upcoming test. I think that this is the best way for me to prepare.

	starting to study earlier, looking at the learning pages during the week to keep concepts fresh, and making sure I can grasp the concepts of every possible example problem.		This has worked for me the past exams so I do not know if there is any place I necessarily need to improve.	Strategies. Time Management and Practice Tests. Improving through targeted practice.	
		improve study strategies	I need to start going to the EF study room to figure out why I got a problem wrong so that I can ask questions about my mistake. I need to start going to the EF study room to figure out why I got a problem wrong so that I can ask questions about my mistake. I need to start going to the EF study room to figure out why I got a problem wrong so that I can ask questions about my mistake. However, I realize there is a large way questions can be presented, so I should make sure I know how to rearrange equations and unknowns given the problem.		
Nothing to improve on	My process for taking the module exam is to make sure I am prepared for it. I do this by going over at least two of the				

	previous exams in the past semesters. I feel no need for improvement as this works for me.				
Get enough sleep	i studied the night before and did not get the greatest sleep. i definitely should study before the night right before the test.			Importance of Sleep for Studying.	I should have gotten more sleep the night before, and prioritized it over the extra studying I did. This process works well when I give myself enough time to study, but this time I did not. I could not do that and study before hand, that way I am able to get a good night's sleep before the exam. I did not have much time to study so study time is the biggest factor.
		Few practice exams	I only did a few practice exams this time, which led to my absolute destruction on exam 3.		
Ask for help	I start studying a week in advance and take all the practice exams from 2016 on. When I get a question wrong I go and watch the video on it. I need to start going to the EF study			Seeking clarification and feedback.	I need to start going to the EF study room to figure out why I got a problem wrong so that I can ask questions about my mistake. I need to start going to the EF study room to figure out why I got a problem wrong so that I can ask questions about my mistake. I need to start going to the EF study room to figure out why I got a problem wrong so that I can ask questions about my mistake. However, I realize there is a large way questions can be presented, so I should make sure I know how to rearrange

	room to figure out why I got a problem wrong so that I can ask questions about my mistake.				equations and unknowns given the problem.

Table B2. Full Comparison of Codebook Results for Manual Analysis, NLPCA, and NLPGPT for "Exam Reflection"

Manual codebook	**Manual codebook Examples**	**NLPCA Codes**	**NLPCA Examples**	**NLPGPT Codes**	**NLPGPT Examples**
Study more	What I did differently between Exam 1 and Exam 2 is study more multiple choice questions/prep questions. This definitely made a difference because I only missed one multiple choice. My goal was to do 5 points better and I did 15 points better than Exam 1.	did more practice	I did more practice exams. I also did more of the practice exams. I also did more practice exams this time.	Consistent and focused studying. Increased study effort. Exam improvement strategies. Practice-based study strategy. Consistent studying over time. Increased Study Time. Consistent study habits. Intensive Pre-Exam Preparation. Improved study consistency. Exam preparation. Consistent study approach.	I practiced more problems and went over old exams, and I also set aside a consistent time to study. I spent more time studying, took more notes on my note sheet, and worked on practice exam questions. Between exam 2 and 3 I studied more consistently. Between exams I spent more time outside of class working on the learning and practice problems.
		spend more time on past exams	I did two practice exams and timed them. I did an extra practice test.		
		studied more consistently	started studying earlier than I did last time. I studied much more and I went about studying differently. All I did differently was studying sooner and more often.		
Better conceptual understanding	I solved all the previous exams to get an idea of what's going to be asked. Additionally, I also went through the concept questions and	increased awareness of conceptual understanding	I believe it helped show me topics I wasn't confident in. This helped me	Utilizing Resources and Understanding Concepts Improved Conceptual Understanding. Understanding concepts through	I worked on understanding the concepts instead of focusing on how to do

	videos which helped me understand concepts properly. I achieved 80% of my goal. And it helped me a lot.		understand what concepts I was confident in and the ones I was less confident. I was able to slow down and think about my answers and the processes I was going through.	preparation. Conceptual Understanding and Application. Improving Material Comprehension. Lack of Understanding and Effort.	the problem. I also went to the ef study room to get help when I was confused on problems instead of just looking at the discussion board. Outside of this, I also took more advantage of my resources like the EF study room tutors, or simply asking those around me for help if I needed it.
Increased grade	I watched the example videos on the learning pages much more carefully. This helped a little bit because my exam grade improved a slightly from exam 2. I still did not reach my goal from the last wrapper.	Increased grade	I increased my grade by 6 points. My grade went up one point. So my grade went up by about 2 points.	Positive changes in study habits. Effective study strategy. Exam score improvement. Changes improved exam scores. Practice Exams for Improvement. Exam success.	This change did improve my exam score because I got a three point increase. This change made an impact on my overall performance on Exam 2 by scoring 17 points higher than Exam 1. Overall this did make an impact as I scored five points higher on exam 2.
Spent more time reviewing prep and practice questions	Between Exam 1 and Exam 2 I spent more time reviewing prep and practice problems throughout the module. This did benifit my score but I was 3 points shy of my goal of a 20 point improvement	focussed on long question practice	I think it helped me get a better grasp on the concepts. It helped me get a better understanding of the topics. I worked more non-multiple choice questions which did greatly strengthen my ability to solve those questions.	Organized Practice and Preparation. Practice and Exam Review. Practice Problems Review.	I think it helped me get a better grasp on the concepts. It helped me get a better understanding of the topics. I worked more non-multiple choice questions which did greatly strengthen my ability to solve those questions.

Studying earlier	I started studying earlier for this exam. I think this definitely helped a lot because I had more time to do more questions on the practice exams. I did reach my goal from the last exam wrapper because I made higher than an 85!	started exam prep earlier	I started studying about a week before the exam rather than waiting just a few days to begin studying. I started studying for the Exam 2 weeks prior to ensure that I was actually understanding the material on a deeper level. I started studying at least a week before the exam so I was not rushing to understand the material.	Early and thorough practice. Early and Intensive Exam Preparation. Early and Consistent Studying. Preparing in advance for exams. Early Exam Preparation. Early and targeted studying. Starting early for exams. Starting studying earlier. Early study habits. Early studying.	I started studying further ahead of the exam. I studied a lot more for this exam than I did for the last one. I started studying a little bit earlier than the previous exam. I also took the time to study more for this exam than I previously had.
		Had less time to prepare for exam	did not do much different as I did not have a whole ton of time to prepare for this exam with family stuff and my midterms being pretty clumped together.		
Asked for help	I did reach my exam wrapper. I asked for help from my TA. I asked to go over concepts that I didn't get or that were confusing. We went over an old exam just to make sure. I think that help me reach my goal. I also went over practice problems which helped too.	asked for help	I also went to the ef study room to get help when I was confused on problems instead of just looking at the discussion board.		
Took more practice tests	I studied a lot more, looked at past clicker questions as possible multiple choice questions, and did more practice exams. I thought this would help me a lot more but it only helped a little.	Did multiple past exams	In addition I completed multiple practice exams in order to study before our exam.	Practice exams. Practice problem emphasis. Time Management. Practice testing. Practice old exams. Practice exam focus. Practice Exams for Exam Prep. Focusing on Past Exams. Practice exams for studying.	Before exam 2 I completed more practice exams than I had before. On this exam, I took a couple more practice

				Practice exam repetition. Using Past Exams.	exams before this exam. On this exam, I took a couple more practice exams before this exam.
Made equation sheet earlier	I stared my equation sheet earlier and it did impact my grade. I got about 12% higher grade. I almost reached my goal.	More effort into equation sheet	I organized my equation sheet better so that I could find equations/ examples easier.	Creating equation sheet gradually. Early equation sheet preparation. Early preparation and equation sheet.	Between Exam 2 and Exam 3, I decided to do the equation sheet as I was completing the modules which I think made a difference. For exam 2, I started my equation sheet and overall studying earlier than I did for exam 1. For exam 2, I began working on my equation sheet earlier than I did for exam 1.
		did not make equation sheet during module	I didn't do just about any practice work and wrote out my equation sheet all the night before the exam.		
More effort into equation sheet	I spent more time on my equation sheet and spent more time looking through my notes and practice problems.	made equation sheet during module	Between exam 2 and 3 I did the same things except I made my equation sheet as I went rather than at the end.	Creating equation sheet. Not using equation sheet. Improved equation sheet organization. Using equation sheet. Focus on Equation Sheet. Equation sheet preparation.	I worked on my equation sheet at the start of the test week and made sure to look at practice exams as the exam got closer. For this exam, I made sure to do more than one practice test and made sure to check every learning page for my equation sheet. I started going through all the notes and making my equation sheet on the Saturday before the exam.
Studied with others	I continue getting higher and higher scores on my exams, but I did not	Group study	I talked with the members of my group	Exam preparation with peers. Group Study and Collaboration.	

	reach my goal. I studied with another friend in this class for this exam which I did not do last time.		and got their help on the homework and eventually studying for the exam.		
Closer to reaching goal	I took more practice tests. I didn't; reach my goal but I got far closer.	Changes had positive impact	The changes did indeed have an impact as my grade went up from the last exam.	Study strategy effectiveness. Moderate changes for exam improvement. Exam wrapper impact. Impactful study strategies. Effective study changes. Goal Progression. Goal-oriented studying. Goal-oriented study strategy. Goal Improvement. Exam Wrapper Improvement Exam performance comparison.	I studied multiple different previous exams. In between Exam 2 and Exam 3 I ended up studying with my friends lab group for several days. I had two other exams and a computer science lab due around the time of exam 3
Other impediments or commitments	I tried to do my homework the night of the lectures, but could not do that because of other classes. Other than that, I'm content with the exam score I got and I studied about the same for Exam 2 as I did Exam 1.	other commitments	I got bogged down with other work and did not start study as early as I intended.	Lack of focus and preparation. Exam absence due to illness.	I was not as strenuous on my studying and was less focused on this exam as I had a lot of things going on at that time. I did however feel more prepared for the exam, so I think I still would have received a worse grade if I did not put in extra study time.
		bad study environment	What I did differently was study in an area where it was hard to focus.	Challenging study environment.	What I did differently was study in an area where it was hard to focus.
Did poorly on exam	Between Exam 2 and 3, I stayed relatively consistent with my strategies. I did, however, begin to work on prep and practice material a day in advance to the respective lecture. Ultimately, this seemed to make a negative impact on my	did poorly on multiple choice	However, I think it had more to do with the fact that I did not do well on the multiple choice.	Exam performance disappointment. Exam disappointment. Disappointed with exam results. Frustration with exam results. Declining exam performance. Disappointed with grades. Mixed emotions about	I still got a pretty terrible grade, even though it might've been slightly better than my previous scores. In result I ended up getting a bad grade.

	score, as I scored signinificantly poorer on Exam 3.			grades. Mixed feelings about performance.	
Did not reach goal	I prepared less for this exam, and I didn't understand the material as well as Module 1. I did not reach my goal for this exam wrapper, as I hoped to get close to what my last exam grade was, and while I was only 7 points away, I knew I could have done better.	did not reach goal	Because of this, I did not reach my goal from the last Exam Wrapper.	Ineffective practice testing. Unsuccessful goal setting. Unsuccessful goal attainment. Goal not achieved due to poor exam performance. Not achieving study goals. Not meeting exam goals. Lack of Goal Achievement. Exam goal not met. Falling short of goals. Goal not achieved. Frustration with unmet goals.	I did not reach my goal from the last Exam Wrapper.
Reached goal	I took more practice exams for exam 3 than for exam 2. These changes made a positive impact. I reached my goal of taking three practice exams before the exam.	reached goal	I did meet my goal of missing no more than one multiple choice question.	Goal achievement through exam improvement. Goal achievement through exam wrapper. Goal achieved. Exam wrapper goal attainment. Goal achieved but not exceeded. Goal achievement. Goal achievement through exam wrappers. Goal attainment through study strategies. Achieving study goals. Goal achievement through study strategies. Successful Exam Wrapper Implementation.	I did reach my goal. I did reach my goal. I did reach my goal.
Changes did not make an impact	I studied a little bit more throughout the week, rather than cramming it all in the night before. The changes did not make an impact, but I think that is because I had a better understanding of module 2 than I did on module 3. Even though that I did better on exam 2, I still reached my goal of getting above an 80 on exam 3.	Changes did not make an impact	The changes did not make much of an impact, the exam seemed harder, and I got 5 points lower.	Mixed results in studying. Ineffective study strategy. Impact of Changes. Minimal impact of strategy. Differences in study effort.	I do not think it made that much of an impact since I got the same grade as I did on this exam as the last exam. It doesnt seem to have impacted my grade too much
		Got the same grade in exam	I do not think it made that much of an impact since I got the same grade as I did on this exam as the last exam.		
Feel more confident	I did more practice exams for module 3's exam. I felt more confident going in, but somehow I	felt confident	These changes did not have much of an impact of my grade for	Repetitive affirmation. Confidence from past success. Increased Confidence in Exam Performance. Building Confidence.	I did really well on exam 1 I largely did the same between Exam 1 and

	ended up doing worse on the test. I did not reach my goal.		the test, but it made me more confident.		Exam 2 considering I was quite happy with my grade on Exam 1. Doing this made me feel very confident on the exam.
Did not do anything differently	I did not do anything differently. Before both exams I watched the videos on how to do old exams and that helped my understanding.	did nothing differently	I don't think that I did much differently between Exam 2 and Exam 3.	Exam Wrapper Neglect. No change in study strategy. Repeating old strategies. Lack of Strategy. Maintaining current grades. Lack of Strategy Adjustment.	I did not do anything too differently. I did not really do anything differently. I did not do anything differently. I did not do anything differently.
Reviewed notes	Between exam 1 and exam 2 I started going through my notes and making my equation sheet earlier before the second exam. This made me feel more prepared because I had time to include more practice problems on my equation sheet that helped with the exam. I did not get to as many practice exams as I had wanted but I was still able to raise my grade enough to pass the exam the second time around.	Took better notes	These changes helped me to understand the material before lectures, which also saved me a lot of time because I could opt out of class if I was confident enough in the material.	Impact of summarizing. Improved note-taking strategy. Note-taking for exam success. Self-assessment through summarization. Effective text summarization. Study strategy - Text summarization.	The a large change I made between Exam 2 and 3 was how I was taking notes. These changes helped me to understand the material before lectures, which also saved me a lot of time because I could opt out of class if I was confident enough in the material. I also felt like I spent more time doing the homework problems by myself as compared to the first module. I took more detailed notes on each module before each lecture through this module.
Focused on multiple choice more	For Exam 2 I did more practice tests and I focused on multiple choice a little bit more and it did make an	focussed on multiple choice practice	I asked more questions about concepts which allowed me to improve	Multiple-choice test-taking strategy.	No, my goal was to miss zero of the multiple-choice

	impact because on the first exam I got only a couple right and on the last exam I only got two wrong.		on the multiple choice questions.		questions and I missed three of them. I made mistakes and didn't get 3 of the multiple choice questions right I did meet my goal of missing no more than one multiple choice question.
Made a study plan	I did reach my goal from the last exam wrapper which I am very pleased with. I implemented the study plan from my last exam wrapper. I increased my study time but I did so in a way that was far more focused than the last exam. I also increased the number of old exams that I've studied which was also extremely helpful.	Changed study habits	I didn't meet my goal of doing as well or better on the exam, but i did learn some better studying skills.	Changing study habits. Effective self-made techniques. Effective study techniques. Effective study technique. Practice Exam Goals. Setting Exam Goals. Exam 2 study habits. Effective study strategies. Exam preparation strategies. Minimal effort strategy.	I also changed up my study habits by spending more time before the exam to review. Between exam 1 and 2, I changed my study habits.
		Did not implement change plan	Between Exam 1 and Exam 2, I didn't really implement my planned changes that I stated in the last exam wrapper.		
Focused on harder questions or concepts	For this test I focused harder on the topics that I knew I was uncomfortable with, and less time on the ones I was comfortable with. My goal was an 85, I made an 83 which isn't my goal, but still a solid score.	focussed on concepts struggled with	These changes helps a lot because I was able to have more practice on the things I don't know rather than repeating things that I do know.	Focusing on question types. Prioritizing difficult topics.	Between exam 2 and exam 3 there isn't much that I did differently except focusing more on the long answer questions than the multiple-choice. What I did differently between Exam 1 and Exam 2 is study more multiple choice questions/prep questions. For Exam 2 I did more practice tests and I focused on multiple choice a little bit more and it did make an impact because on the
		Focussed on questions struggled with	I focused on the specific types of questions I struggled with in order to better make use of my time studying.		

					first exam I got only a couple right
Did past exams without looking at solutions	I looked through more tests and actually took one without the solutions instead of just looking over it and saying yeah I can do that. The changes made a huge impact because I made mainly small errors on this test instead of missing an entire problem. I did reach my goal from the last exam wrapper.			Active Learning. Practice exam utilization.	For this exam, I tried to use the old practice exams and put more time into my equation sheet. I focused more on doing the practice exams this time around and used making my equation sheet as a way to study.
Spent more time on the learning pages	I did not reach the goal that I had set for myself. While I did make changes, the material was not as familiar to me and I was less confident going into the exam. I did put additional effort into the learning pages before class which helped my understanding a lot.	spend more time understanding learning pages and prep	I spent more time working on understanding the learning pages and prep questions as well as studying for the exam.	Focus on Learning Pages. Reviewing learning pages thoroughly. Allotting more study time.	I spent more time reading the learning pages and writing down every equation. I spent more time going back over the videos in learning pages and working examples. I spent more time on the learning pages and not as much time doing the practice pages.
Asked more questions	I was more on top of my work throughout the learning pages. I stayed after lecture to ask questions about practice problems. I did more practice exams, but did not reach my goal of the amount of practice exams I wanted to complete.				
Less stressed	I managed time very well on this exam. I was able to complete much more of the exam, and my grade was 20 points higher. I reached my goal from last exam wrapper, and it helped me so much stress-wise.				

Put example problems on formula sheet	I looked at more homework problems and put examples of them on the formula sheet. This made a big difference from exam 1 to exam 2 as I scored ten points higher on exam 2 than exam 1. Although this was a great improvement, I was still four points off from my goal I wanted to reach for exam 2.			Equation Sheet and Practice Problems. Emphasizing Equations and Examples. Adding personal examples.	I studied more and did a better Equation Sheet for Exam 3. I made a more elaborate equation sheet for Exam 2, and I also did a lot more practice problems.
Did work without help	I worked on practice problems while making my equation sheet and finished all homework without outside help. Yes, I finished the test, however, I made the lowest grade yet on a test.	Worked alone more	I also felt like I spent more time doing the homework problems by myself as compared to the first module.	Preparation and Self-Reliance.	I completed my entire equation sheet a few days before the exam and used it along the way as I filled it out. I worked on practice problems while making my equation sheet and finished all homework without outside help. From exam 2 to exam 3 I attended all classes and devoted more time to the pre-class assignments and notes.
Changes had negative impact	I spent less time studying to try and get more sleep the night before the test. The changes did make an impact but it was a negative one. I did not reach my goal unfortunately and I hope that I will be able to find the middle ground of 90% or higher whilst getting at least 6.5 hours of sleep the night before.	Grade decreased	In fact it impacted me by making me score around 10 points lower than the previous exam.	Ineffective study changes. Negative impact of exam changes. Negative impact of changes. Negative correlation between studying and grades. Inconsistent study habits. Less studying for exams.	The changes did not make an impact, my grade was worse and I did not reach my goal. My changes did not make a positive impact on my grade and I did not end up reaching my goal from the exam 1 wrapper.
Watched videos	I spent more time going back over the videos in learning pages and working examples. It helped and I got close, but I did not quite make my goal.				

Timed practice tests	I took more practice tests and alternated taking them timed and untimed so I would have the experience of a timed exam but could also take all the time I needed to work full problems slowly to find the simple mistakes I would make. The changes I made did have an impact I did much better on this exam than I did on my previous one and I felt more confident going into it.			Practice Exam Timing.	I did two practice exams and timed them. I did an extra practice test.
Improved time management	I started studying a-lot earlier than I did before and I timed my self during a couple practice tests which helped me with my time management. I did not reach my goal however, but i did do a-lot better.			Time management strategies. Procrastination and Time Management. Improved Time Management. Consistent but last-minute studying.	Spreading out my studying allowed me to better focus time on my weak points with plenty of time before the exam. I worked the problems as quick as I could, but I still didn't have enough time to finish the test.
Looked at past clicker questions	I studied a lot more, looked at past clicker questions as possible multiple choice questions, and did more practice exams. I thought this would help me a lot more but it only helped a little.				
Improved problem solving	Between exam 1 and exam 2 I worked on practice exams. This allowed me to understand how to solve more problems than I would if I did not study old exams. I did reach my goal of studying more, however I did not do as well as I did on the first exam.				
Paid more attention	I did not really do much differently. Though I did take more practice			Focused and thorough learning.	I spent a lot more time working on the

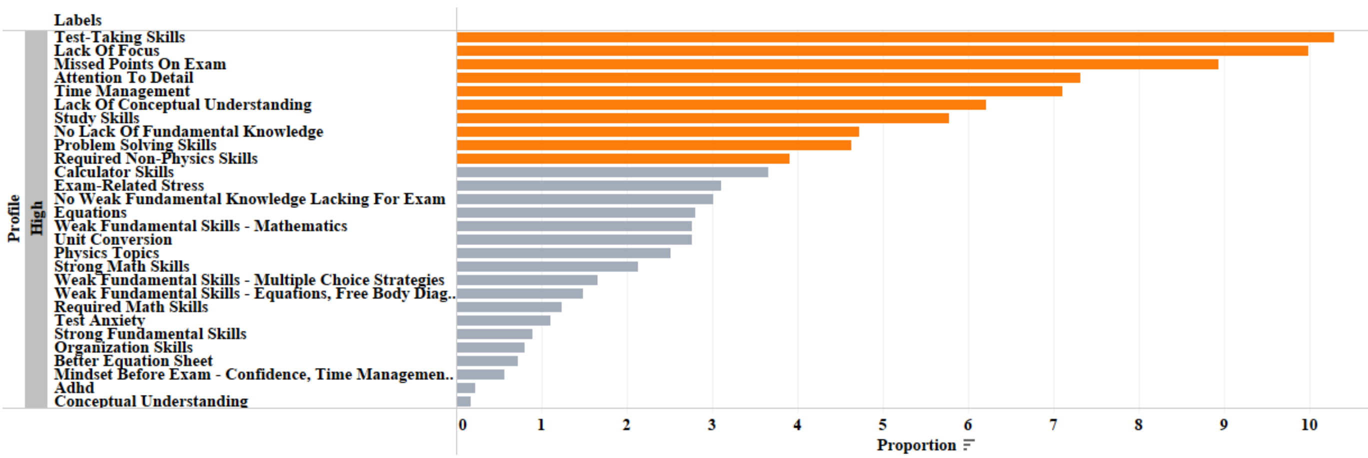
	exams and pay more attention, it did not really improve my score that much. I still did not get the grade I desired or achieve my goal of getting a 100.				homework and made sure I really understood the content. I also made sure I understood the material a lot more as I was doing the learning pages.
Stengthen math skills	I focused on incorporating a "recap" after every learning page to make sure I fully comprehend the math and concepts behind the processes. This helped me towards the end of the module to put together what I needed to study harder and what I already had a good understanding of.				
		didn't study as much	I did not spend enough time running through the material I needed to in order to really understand the concepts.	Lack of Preparation.	I did not spend adequate time for exam 2 to score well on the exam. I did not prepare enough for this exam. Between Exam 1 and Exam 2, I didn't really implement my planned changes that I stated in the last exam wrapper.
		exam harder than expected	This change probably helped but the exam was just harder than I expected and I got confused on a few of the problems.		
		missed points on exam	It was because I made simple mistakes.	Mistake prevention.	It was because I made simple mistakes. but I ended up making stupid mistakes.
		Not enough time in test	I worked the problems as quick as I could, but I still didn't have enough time to finish the test.	Lack of Time Management.	I did not spend enough time running through the material I needed to in order to really understand the

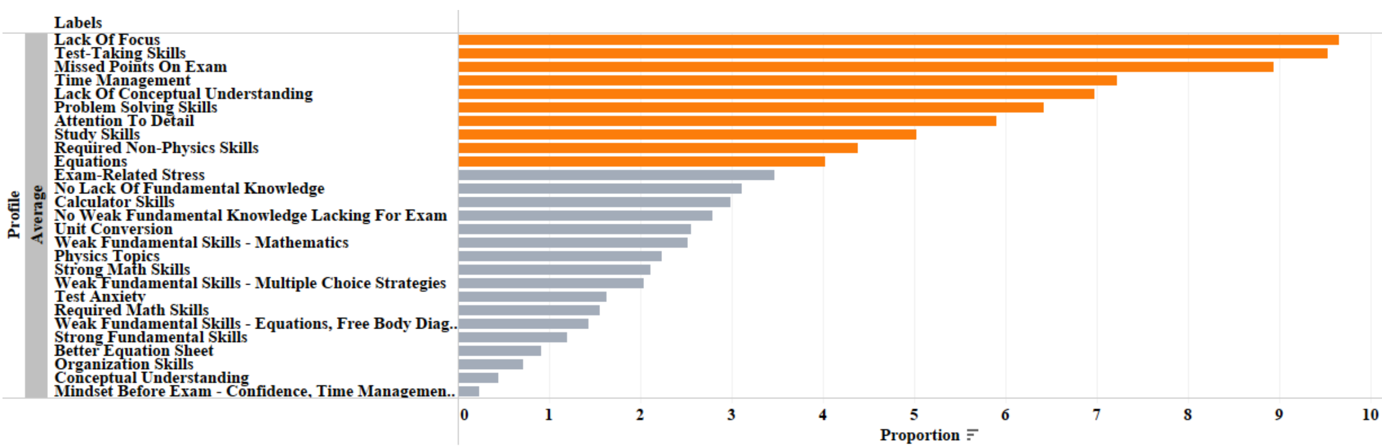
					concepts. Additionally, I read through all of my notes and worked problems I didn't understand. I also struggled to complete the learning modules early. While it may not be entirely associated with my study habits–as I seemed to grasp the content more easily than the last module–I can tell that they still made a difference regardless.
		did not complete exam wrapper	I missed the previous exam wrappers so I did not set any goals for them.		

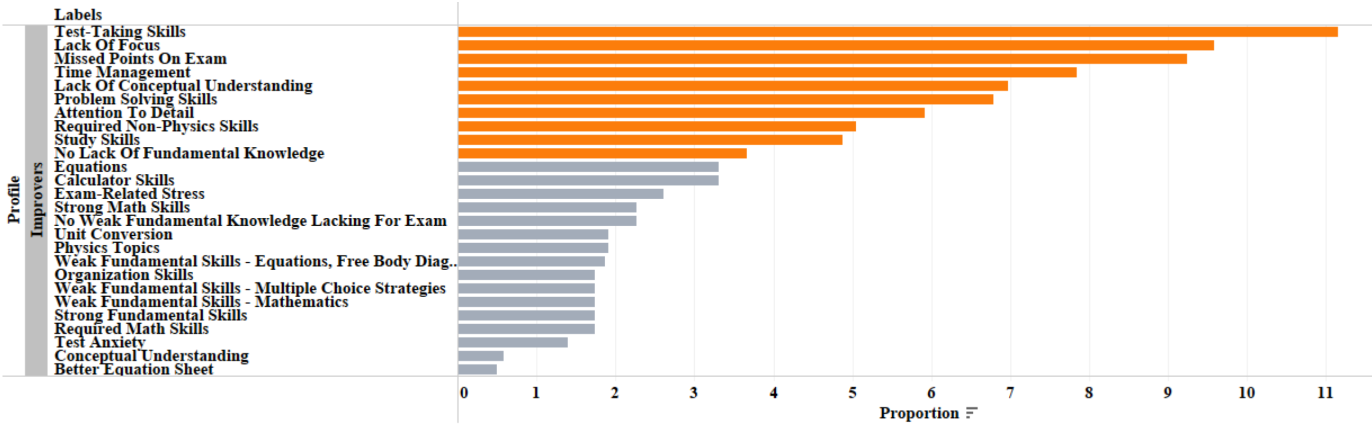
Appendix C Full Distribution of topics

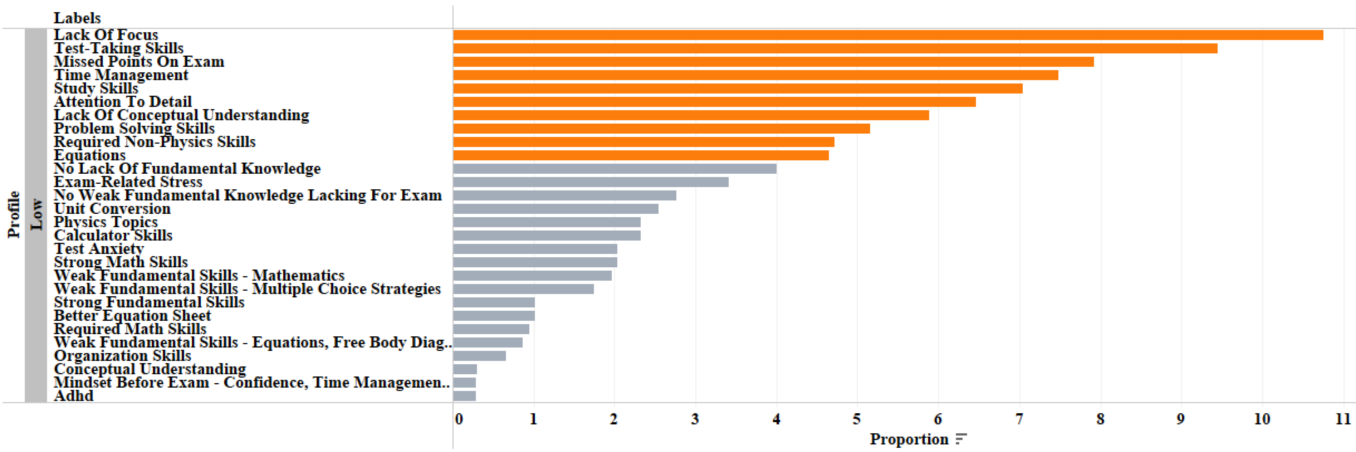
Table C1: Parameter settings for UMAP and HDBSCAN in clustering student exam grades

UMAP	
Number of neighbors	5
Minimum distance	1
Number of components	2
Distance metric	Euclidean
HDBSCAN	
Minimum cluster size	50

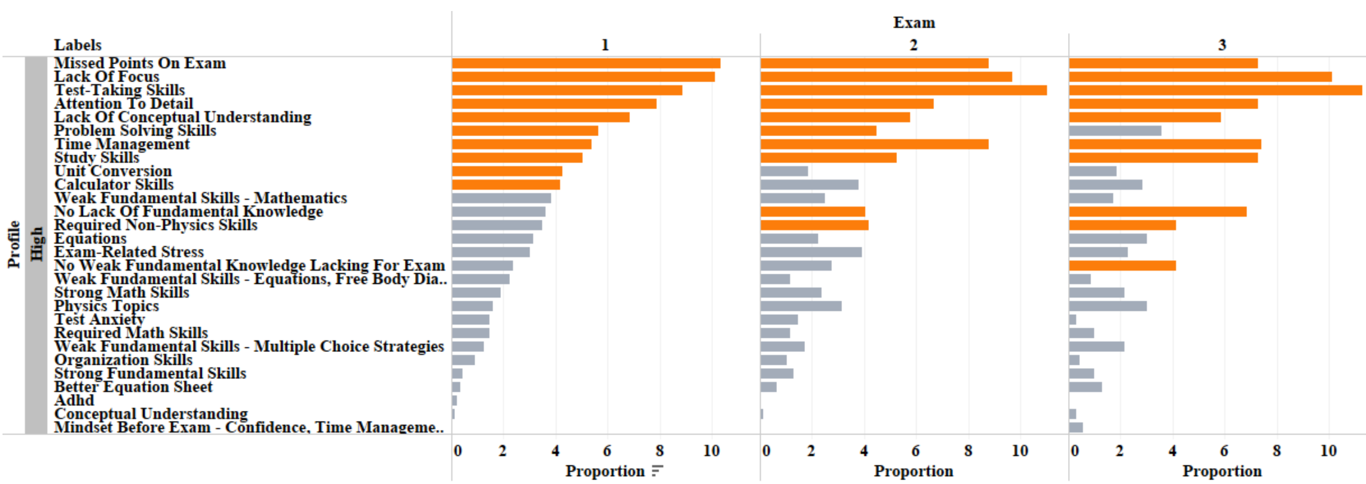


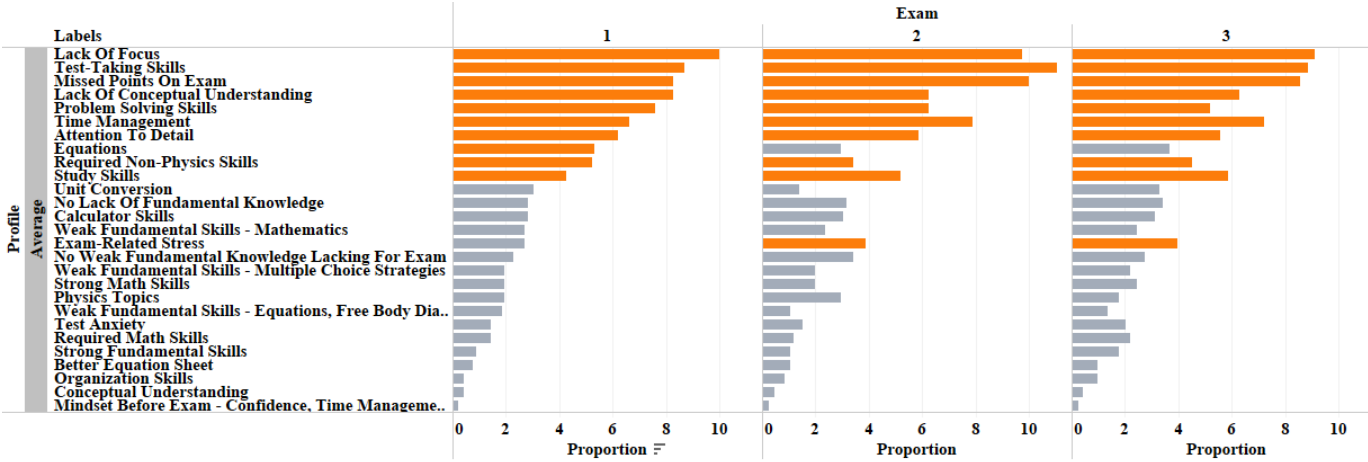


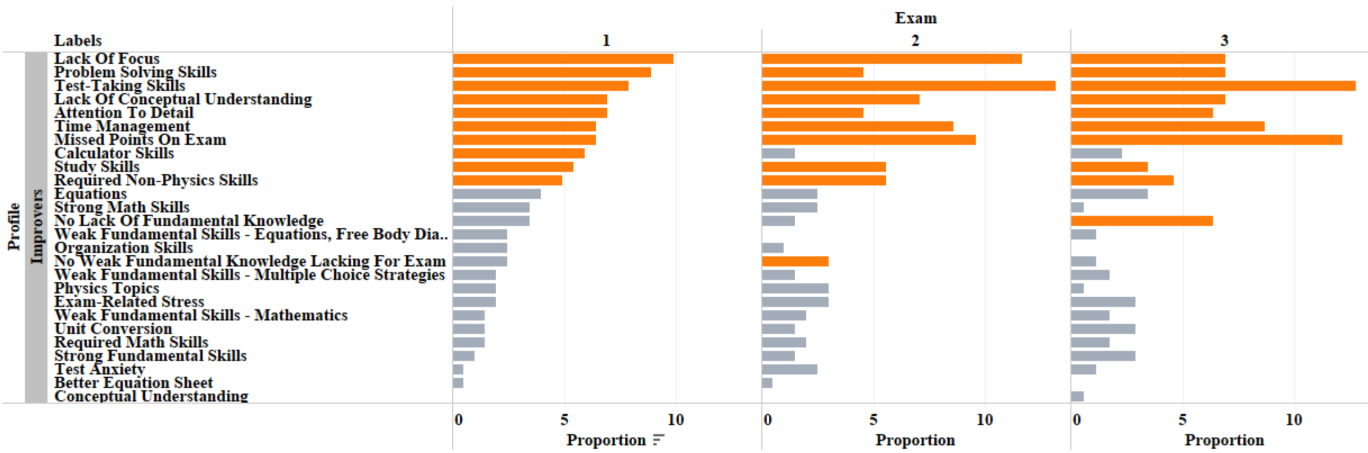


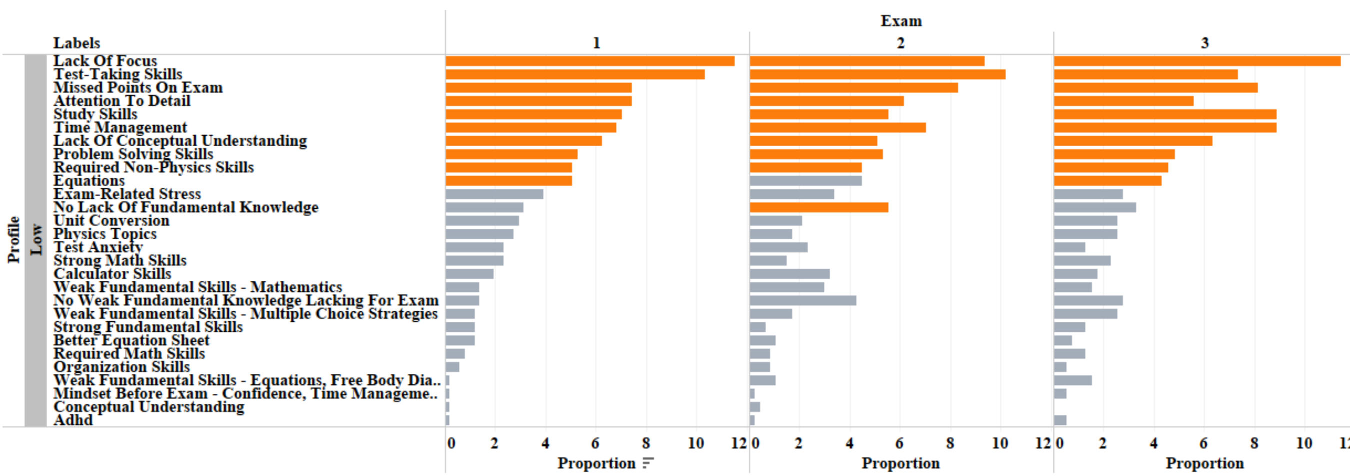


Distribution of topics across exams



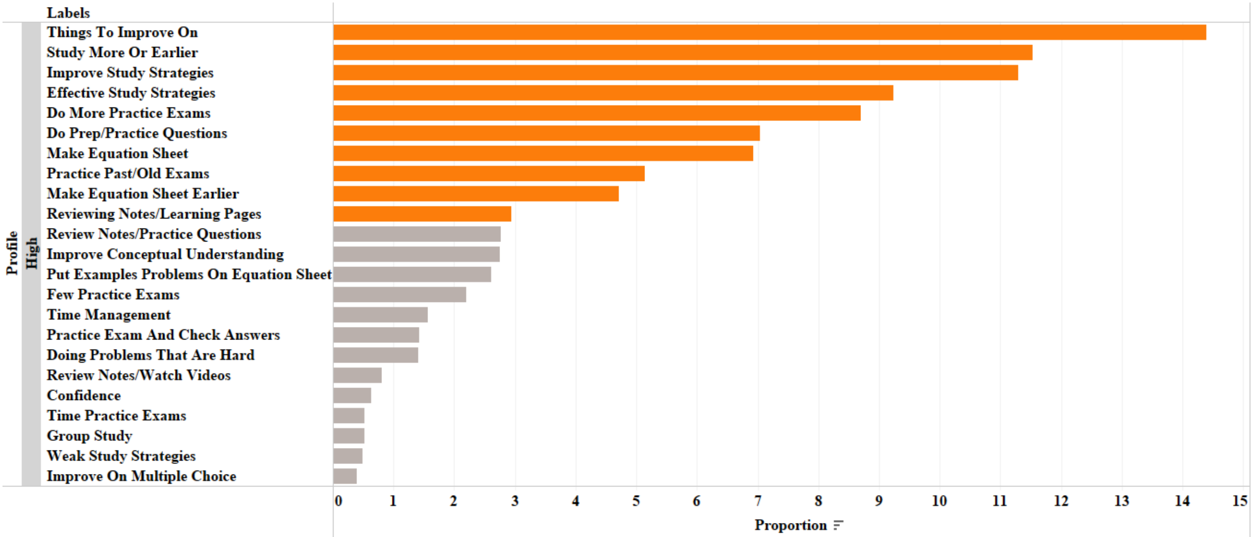


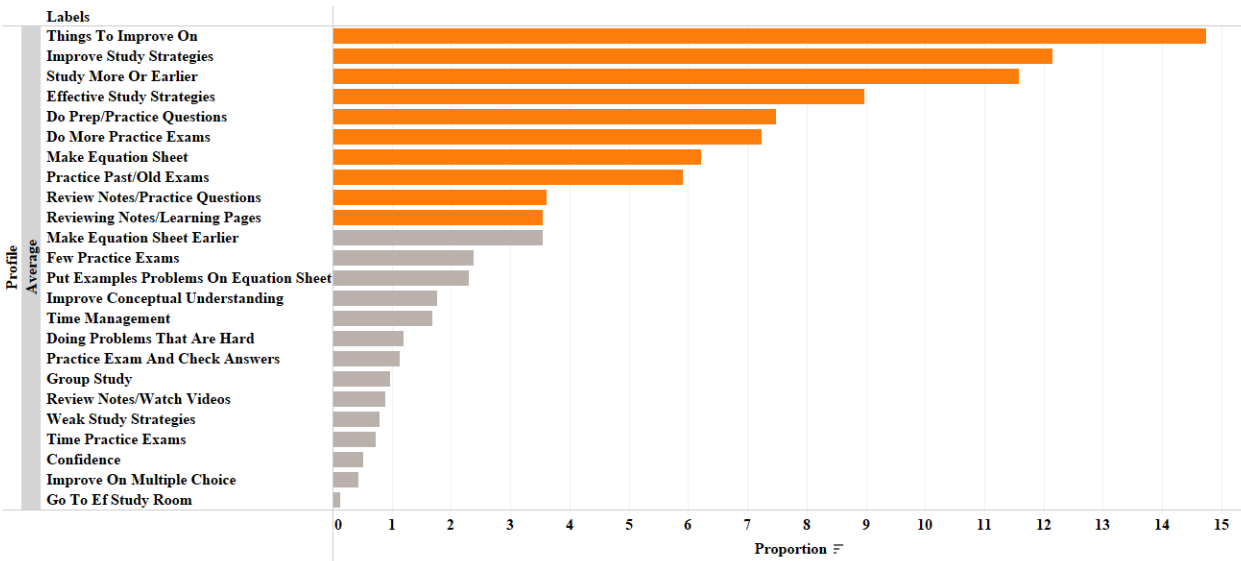


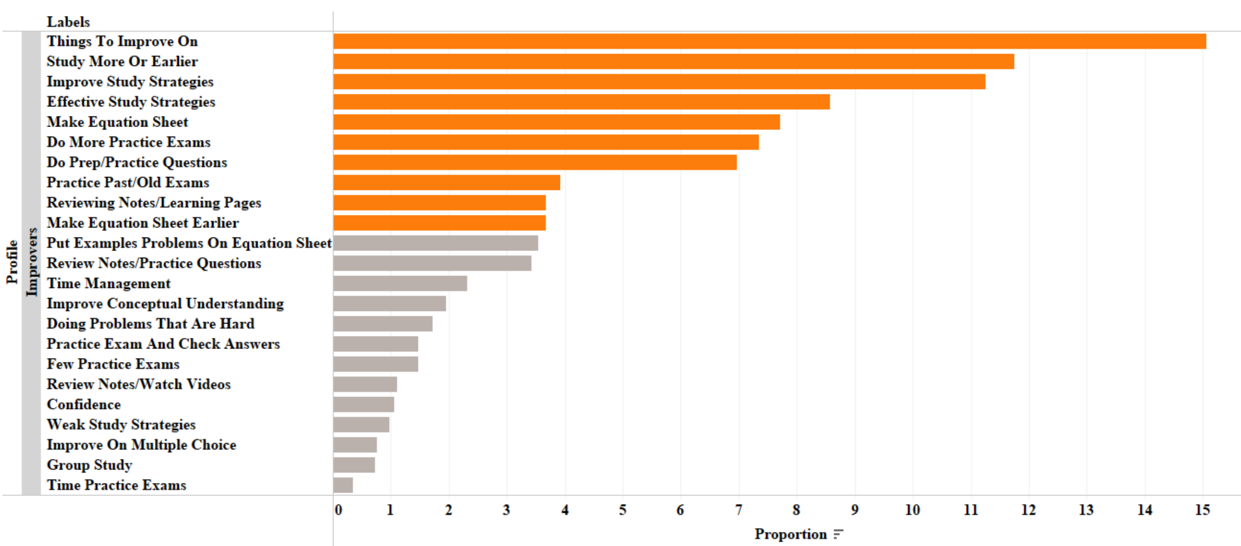


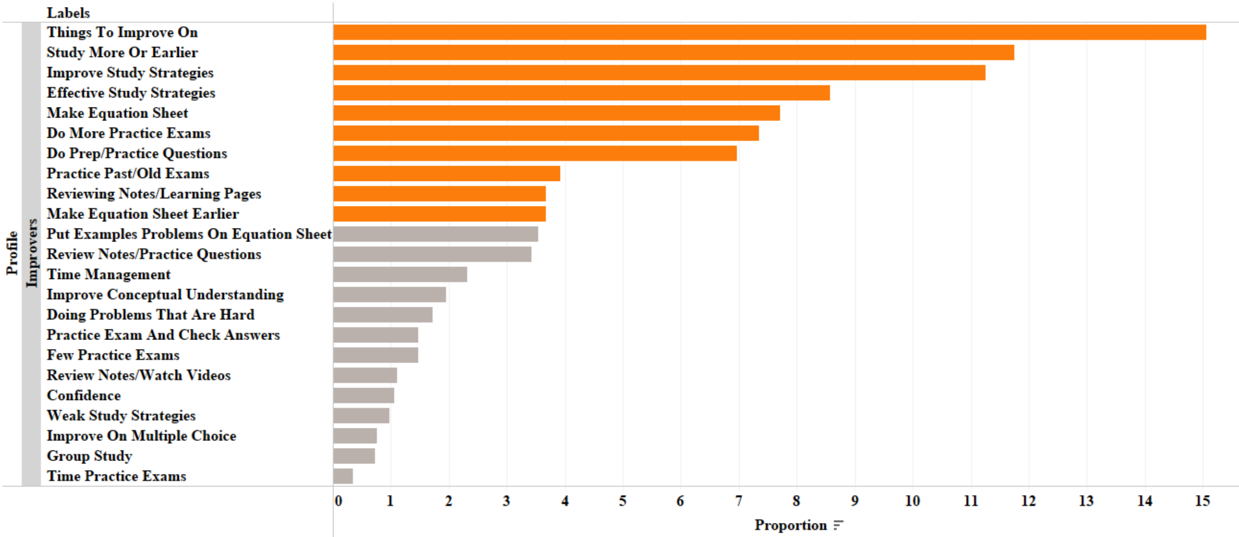
Preparations Process

Distribution of topics for different profiles

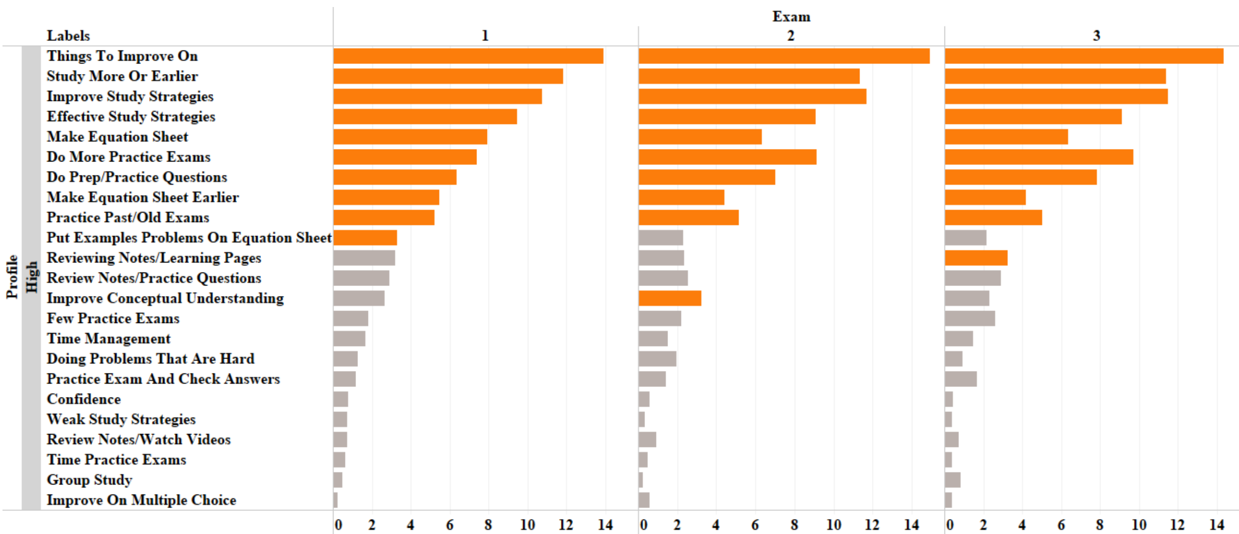

Labels
Profile
High
Things To Improve On
Study More Or Earlier
Improve Study Strategies
Effective Study Strategies
Do More Practice Exams
Do Prep/Practice Questions
Make Equation Sheet
Practice Past/Old Exams
Make Equation Sheet Earlier
Reviewing Notes/Learning Pages
Review Notes/Practice Questions
Improve Conceptual Understanding
Put Examples Problems On Equation Sheet
Few Practice Exams
Time Management
Practice Exam And Check Answers
Doing Problems That Are Hard
Review Notes/Watch Videos
Confidence
Time Practice Exams
Group Study
Weak Study Strategies
Improve On Multiple Choice
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Proportion

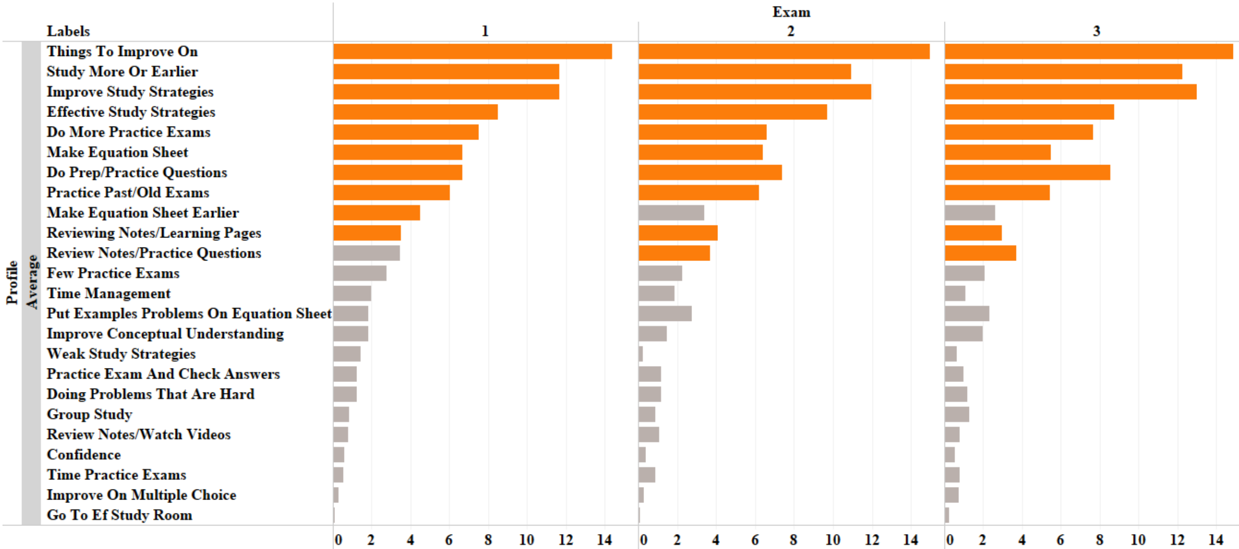

Labels
Profile
Average
Things To Improve On
Improve Study Strategies
Study More Or Earlier
Effective Study Strategies
Do Prep/Practice Questions
Do More Practice Exams
Make Equation Sheet
Practice Past/Old Exams
Review Notes/Practice Questions
Reviewing Notes/Learning Pages
Make Equation Sheet Earlier
Few Practice Exams
Put Examples Problems On Equation Sheet
Improve Conceptual Understanding
Time Management
Doing Problems That Are Hard
Practice Exam And Check Answers
Group Study
Review Notes/Watch Videos
Weak Study Strategies
Time Practice Exams
Confidence
Improve On Multiple Choice
Go To Ef Study Room
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Proportion

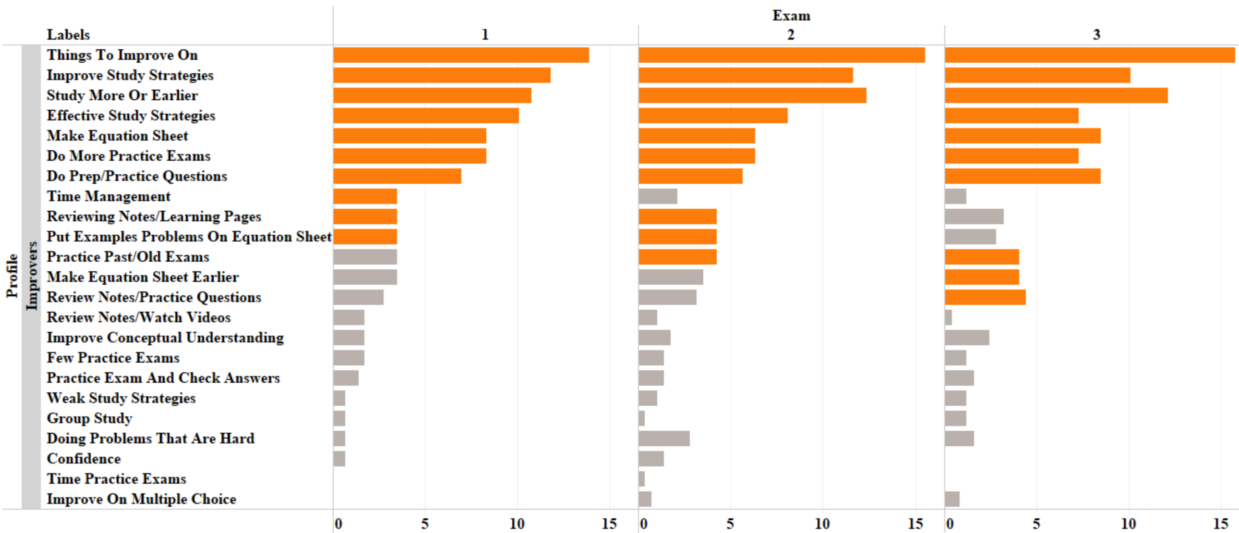

Labels
Profile
Improvers
Things To Improve On
Study More Or Earlier
Improve Study Strategies
Effective Study Strategies
Make Equation Sheet
Do More Practice Exams
Do Prep/Practice Questions
Practice Past/Old Exams
Reviewing Notes/Learning Pages
Make Equation Sheet Earlier
Put Examples Problems On Equation Sheet
Review Notes/Practice Questions
Time Management
Improve Conceptual Understanding
Doing Problems That Are Hard
Practice Exam And Check Answers
Few Practice Exams
Review Notes/Watch Videos
Confidence
Weak Study Strategies
Improve On Multiple Choice
Group Study
Time Practice Exams
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Proportion

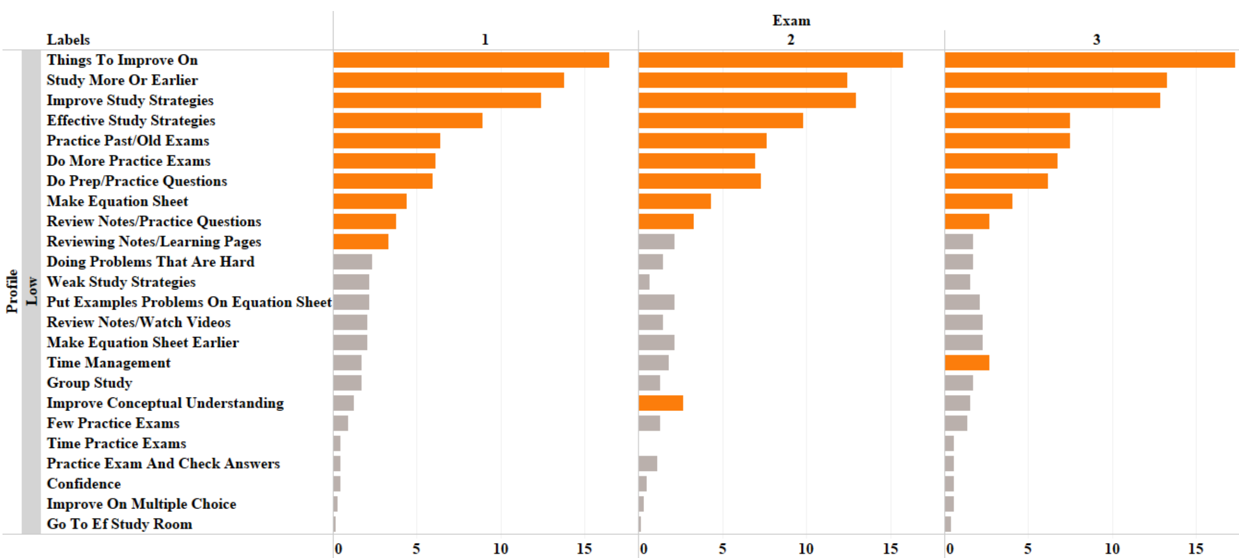


Distribution of topics across exams




Exam
Labels
1
2
3
Profile
Average
Things To Improve On
Study More Or Earlier
Improve Study Strategies
Effective Study Strategies
Do More Practice Exams
Make Equation Sheet
Do Prep/Practice Questions
Practice Past/Old Exams
Make Equation Sheet Earlier
Reviewing Notes/Learning Pages
Review Notes/Practice Questions
Few Practice Exams
Time Management
Put Examples Problems On Equation Sheet
Improve Conceptual Understanding
Weak Study Strategies
Practice Exam And Check Answers
Doing Problems That Are Hard
Group Study
Review Notes/Watch Videos
Confidence
Time Practice Exams
Improve On Multiple Choice
Go To Ef Study Room
0 2 4 6 8 10 12 14


Exam
Labels
1
2
3
Profile
Improvers
Things To Improve On
Improve Study Strategies
Study More Or Earlier
Effective Study Strategies
Make Equation Sheet
Do More Practice Exams
Do Prep/Practice Questions
Time Management
Reviewing Notes/Learning Pages
Put Examples Problems On Equation Sheet
Practice Past/Old Exams
Make Equation Sheet Earlier
Review Notes/Practice Questions
Review Notes/Watch Videos
Improve Conceptual Understanding
Few Practice Exams
Practice Exam And Check Answers
Weak Study Strategies
Group Study
Doing Problems That Are Hard
Confidence
Time Practice Exams
Improve On Multiple Choice
0 5 10 15


Exam
Labels
1
2
3
Profile
Low
Things To Improve On
Study More Or Earlier
Improve Study Strategies
Effective Study Strategies
Practice Past/Old Exams
Do More Practice Exams
Do Prep/Practice Questions
Make Equation Sheet
Review Notes/Practice Questions
Reviewing Notes/Learning Pages
Doing Problems That Are Hard
Weak Study Strategies
Put Examples Problems On Equation Sheet
Review Notes/Watch Videos
Make Equation Sheet Earlier
Time Management
Group Study
Improve Conceptual Understanding
Few Practice Exams
Time Practice Exams
Practice Exam And Check Answers
Confidence
Improve On Multiple Choice
Go To Ef Study Room
0 5 10 15