Investigating Student Perceptions of Equitable Grading Practices

Anmol Shukla

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Computer Science and Applications

Stephen H. Edwards, Chair

Donald S. McCrickard, Co-chair

Margaret O. Ellis

April 29, 2024

Blacksburg, Virginia

Keywords: grading, experience-report, computer-science-education

Investigating Student Perceptions of Equitable Grading Practices

Anmol Shukla

(ABSTRACT)

Grading is one of the key components of modern pedagogy. Grades are primarily meant to be a method of communication, which is initiated by the owners of the pedagogy (instructors, school, etc.) and are received by various parties such as students: as feedback on their learning outcomes, parents, potential employers, other institutions, etc. However, traditional grading practices in wide use across institutions globally suffer from many problems and can prove to be a hindrance to students achieving their learning outcomes. Many of these practices do not have a backing in education and social research and suffer from various problems such as inherent bias, rewarding of behavior over skill or knowledge, increasing student anxiety, etc. In contrast, EGPs have been backed in research, follow a more open assessment method and have been shown to induce increased learning. Thus, to mitigate some of these problems, we employ Equitable Grading Practices (EGPs) in an introductory CS-1 course at Virginia Tech for the Fall 2023 and Spring 2024 semesters. In this thesis, we introduce these practices and evaluate student perceptions of them thereafter to answer research questions so that we may be able to refine these practices. We gather student experiences primarily through two instruments: a survey and one-on-one interviews. We describe these in detail and evaluate them through the use of statistics as well grounded theory analysis to be able to extract student perceptions of these grading practices.

Investigating Student Perceptions of Equitable Grading Practices

Anmol Shukla

(GENERAL AUDIENCE ABSTRACT)

Grading is a fundamental aspect of education, shaping not only student learning but also their future opportunities. Traditionally, grades are intended to communicate a student's performance to various stakeholders, including parents and potential employers. However, standard grading methods suffer for biases and promote anxiety among students, rather than truly reflecting their knowledge and skills. In an effort to address these issues, our research at Virginia Tech investigates the implementation of Equitable Grading Practices (EGPs) in an introductory computer science course. These practices, supported by educational research, emphasize a more transparent and inclusive approach to assessment, aiming to enhance learning rather than merely evaluate it. This thesis examines how students perceive these new grading practices through surveys and detailed interviews. Our goal is to understand their experiences and refine these methods to better support student achievement and well-being. This study not only offers insights into the impact of EGPs but also contributes to broader educational reforms aimed at fairer and more effective grading systems.

Dedication

To the Lord Almighty, my parents and my brother.

Acknowledgments

I would like to express my deepest gratitude to Dr. Stephen H. Edwards for his invaluable guidance and continued support throughout this work. His insights and expertise have been crucial in shaping both the direction and success of my research.
I am also thankful to my thesis committee members, Dr. D.S. McCrickard and Ms. Margaret Ellis, for their thoughtful feedback and encouragement. Their perspectives have been instrumental in improving my work and covering my blind spots.

I also thank Dr. Adrienne Decker and all the members of our research group. Working alongside such talented and passionate individuals has been both inspiring and immensely rewarding. I am grateful for the collaborative environment we shared, which has significantly contributed to my personal and professional growth.

I deeply appreciate my parents and my brother, whose love and support have been my anchor and motivation throughout this journey. Thank you for believing in me, encouraging me, and providing the emotional foundation necessary for my success.

To my roommates and friends, thank you for being there through countless late nights and for all the moments of levity in between the stressful times. Your friendship and support have made all the difference.

Lastly, I would like to thank the Virginia Tech Computer Science Department. The resources, opportunities, and academic environment provided by the department have been fundamental to my research and education. Thank you all for your support and for contributing to this significant phase of my academic journey.

This material is based upon work supported by the U.S. National Science Foundation un-

der Grant Numbers 2235337, 2235643, and 2235644. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

ChatGPT was used in generation of the general audience abstract and the Microsoft Copilot tool was used as a proofreading assistant in this thesis.

Contents

List of Figures

List of Tables

List of Abbreviations

CS-1 course CS-1000 level undergraduate course

EGP Equitable Grading Praticices

MSLQ Motivated Strategies for Learning Questionnaire

TA Teaching Assistant

Chapter 1

Introduction

1.1 Background

Assessing student performance is a crucial aspect of most pedagogical approaches, with letter grading being the most prevalent method, utilized by 92% of all institutions affiliated with the American Association of College Registrars and Admissions Officers (AACRAO) [10]. By design, grades serve as a feedback mechanism and a communication medium between educational institutions and various stakeholders such as students, parents/guardians, and employers. Typically, a student's performance is represented through letter grades or percentages, reflecting their efforts on exams or throughout the course. While instructors employ a variety of grading policies, the impact of these policies on learning remains a significant topic of academic discourse. Although individual grading practices can have varying degrees of correlation with professional success (discussed further in Section 2.1), the achievement of learning outcomes has been shown to have a significant relationship with early career success [44].

Thus, grading practices play a pivotal role in shaping student learning experiences and outcomes, making them a crucial pedagogical tool. In most computer science classrooms, traditional grading is applied to assessments such as quizzes, exams, labs, assignments, etc., which are graded on a 100-point scale. Mistakes made by students result in point deductions, and a final grade is computed by dividing the 100-point scale into letter grades

and calculating a weighted average of all scores. However, traditional grading systems have been widely criticized for their subjectivity, lack of transparency, and potential to perpetuate inequalities. Over time, these flaws in traditional grading have been highlighted, and as corrective measures, alternative grading methods such as specification grading have emerged, aiming to provide more equitable, transparent, and less subjective assessment frameworks. This thesis delves into the qualitative perceptions of students regarding these two contrasting grading approaches.

Traditional grading systems typically rely on a combination of factors, including subjective teacher judgment, point accumulation, and a broad spectrum of assessment criteria. While these methods have been utilized in educational practices for decades, they often fall short in providing clear and meaningful feedback to students, fostering a growth mindset, and mitigating biases inherent in subjective evaluation. Moreover, many of these traditional methods do not have any backing in educational research despite their historical application [24, 42]. Some traditional grading systems can prioritize compliance with predefined standards rather than promoting deep understanding and mastery of the subject matter. Additionally, traditional grading practices can lead to various higher-order effects such as an increase in student stress and anxiety, students externalizing the locus of control, educators and instructors using grades for purposes other than to provide feedback to students (e.g., for gaining student compliance), grade grubbing (discussed further in Section 2.1), unethical behavior by students, etc.

To overcome these problems associated with traditional grading practices, we assess the use of Equitable Grading Practices (EGP), which generally follow more open assessment methods that have been proven to lead to greater student engagement and learning [37]. Equitable grading practices encompass a broad set of methods aimed at increasing the accuracy of student performance assessment, bias-resistance, and intrinsic motivation [7].

Some of the EGP practices include: separation of behavior from assessment (thus, reducing instructor bias), specification grading (e.g. a 0-4 grading scale), focus on the achievement of learning objectives instead of grades, provision of additional support, and finally, provision of multiple pathways, where a student can demonstrate their learning in multiple ways.

Specification grading offers a structured and transparent approach to assessment. Rooted in clear and detailed grading criteria, specification grading precisely delineates what students must accomplish to achieve certain grades. This method emphasizes mastery of specific learning outcomes and provides students with clear expectations coupled with detailed and often, criterion-referenced feedback. This allows students to focus on areas requiring improvement and take control of their learning journey, thus promoting self-reflection, which is one of the phases described by Zimmerman in his cyclical model of self-regulation [48]. Moreover, specification grading has been praised for its potential to reduce bias in grading, as it prioritizes objective evaluation based on predetermined criteria rather than subjective impressions.

1.2 Assessment of Equitable Grading Practices (EGP)

While various studies have evaluated the efficacy of EGPs in numerous settings, including high-school and undergraduate levels, our aim is to assess their deployment in the context of undergraduate CS education, specifically in a large, CS-1 Introductory course at Virginia Tech. Decker et al. [6] introduce and develop many of the EGPs assessed in this study. Another study by Najjar et al. reports on a similar line of questioning as does this study, however, to a lesser degree of depth and detail, stating that their study is a "preliminary exploration laying the groundwork for a comprehensive large-scale evaluation of

equitable pedagogical practices" [26]. A study by Vahid et al. evaluates the effects of employing a specific late policy in a large CS-1 course [43].

Many EGP practices have been used in previous settings in their entirety or in part. In this study, we aim to assess a subset of these Equitable Grading Practices, namely: specification grading, employment of a late policy, and replacing high-stakes midterm exams with a larger number of shorter quizzes. These specific EGP choices are described in greater detail in Chapter 3.

This thesis seeks to explore the qualitative perceptions of students regarding these two grading paradigms, aiming to uncover insights into their experiences, challenges, and preferences. By comparing and contrasting traditional grading with specification grading, this study intends to find out what undergraduate students in a freshman year, CS-1 course think about specification grading through instruments such as surveys and interviews and shed light on the potential benefits of adopting more equitable and transparent grading practices in educational settings. Thus, through these in-depth interviews and thematic analysis, this research aims to contribute to the ongoing discourse on assessment reform and student-centered pedagogy. By examining these contrasting approaches through the lens of student experiences, this study aims to provide valuable insights for educators and policymakers seeking to promote fairness, transparency, and student empowerment in the assessment process.

This master's thesis is a part of a broader effort of performing collaborative research to transform pedagogical tools, specifically grading practices and techniques in use within computer science education, in the National Science Foundation's grant award #2235337 titled "Transforming Grading Practices in the Computing Education Community" [9]. The broader work is a joint effort between Virginia Tech, University at Buffalo, and North Carolina State University.

The experimental observation and analysis are divided into two parts: reporting on objective student performances and capturing student perceptions. The first part is discussed in the research paper titled "Early Experiences with Specification Grading in Intro CS Courses"[5]. On the other hand, this thesis aims to report on the latter.

1.3 Research Questions

While instructors employ EGPs to achieve specific aims, this research focuses on how students perceive those interventions. However, it is beyond the scope of this effort to investigate the full range of EGPs. Instead, we investigate student perceptions of three specific EGPs deployed in the target course used for the study: one implementation of specification grading (referred to as the *EMRN grading scheme*), one specific *late policy*, and the use of smaller quizzes (rather than bigger exams/tests) in an attempt to reduce test anxiety.

We aim to answer the following research questions about these specific implementations of Equitable Grading Practices:

1. **RQ1**: What are student perceptions of the *EMRN grading scheme*?

2. **RQ2**: What are student perceptions of the employed *late policy*?

3. **RQ3**: What are student perceptions of being assessed using smaller quizzes in place of larger midterm exams?

4. **RQ4**: How do student self-reported information about their own self-regulated learning relate to their perceptions of these EGPs?

Elaborating on the research questions, with respect to RQ1, we aim to understand the relationship between EMRN grading scheme and the students and whether in their opinion, it is conducive to learning when compared to traditional grading schemes. With respect to RQ2, we aim to understand the students' perception of the relationship between late policy and their executive function, time management skills and general anxiety. Thirdly, with respect to RQ3, as discussed further in Section 3.1.3 there is a provision for students to omit one of the quizzes from their grade and we additionally analyze student perceptions on the same apart from the smaller quizzes. Lastly, we aim to contextualize student responses about the EGPs using the Motivated Strategies for Learning Questionnaire (MSLQ) [30] and assessing the student population's task-specific cognitions, motivations and learning strategies (**RQ4**).

The impact of the research questions is stated as follows: Assessment of student perception of the EMRN grading scheme lets us understand the relationship that the students have with the grading scale. This is important because the entire idea behind the employment of these *Equitable* Grading Practices is to do away with the feeling of oppression or lack of choice that many students experience with traditional grading practices. Thus, understanding student opinion of this system helps us understand the impact of this system on the relationship between students and their perceived notion of grading. A positive student-grading relationship should hopefully lead to greater student engagement, which could have other second-order positive effects on student learning.

Assessing this perception serves an additional purpose. As we observe quite often, grading becomes an end goal in itself for many students, which should instead be attaining learning and mastery of the subject. Understanding student perception should help us understand the degree of student motivation to attain learning (or achievement of learning outcomes) of the course. We speculate that if a student has learned how to use this system

and engage with it to their advantage, it should lead to a greater degree of achievement of learning outcomes of the course.

We speculate that the employment of the late policy in this course should help students get better at their executive function and time management. We also surmise that this should help in reducing the general feeling of anxiety around submission deadlines. Conversely, it could increase student tardiness if students do not realize that the late policy has an opportunity cost (in that it takes available time away from the next submission) or use it to procrastinate chronically. These issues are discussed further in Chapter 3. Assessing The employment of the late policy helps us understand the impact of the late policy on the students' executive function and time management and helps us understand their psychology as they engage with the late policy. This helps us design more effective late policies in the future and thus make grading more equitable.

The design choice of using smaller quizzes to assess knowledge as opposed to bigger exams has several advantages for students. It enables them to focus on the depth of a topic (Since a smaller subset of the syllabus is assessed as compared to a much larger one), it fits better in their schedules, offers insurance in the case of unexpected life situations, etc. Understanding this design choice's impact on test anxiety can help us design better assessments in the future. It can help ground future research into how the structure of test-like assessments can influence student test anxiety and help facilitate more equitable pedagogies.

Lastly, it is important for us to understand the context in which these student perceptions about the aforementioned EGPs exist. It is important for us to ground student experience in their task-specific cognitions, motivations and learning strategies. This helps us make a distinction between the efficacy of the grading practices themselves and the quality of the fit between these practices and the student population.

1.4 Overview

Following this chapter, in Chapter 2, we provide some background and ground our research efforts through pre-existing research work and literature and answer where our work fits among the library of work done on equitable grading practices. In Chapter 3, we dive into the details about the employed EGPs in the CS-1 course. We also describe the various instruments used to gather student perceptions and their deployment. In Chapter 4, we discuss the results obtained from the deployment of the aforementioned instruments and perform analaysis of the data obtained. Following this, in Chapter 5, we speak to the results obtained in the preceding chapter and the inherent relationships. Lastly, in Chapter 6, where we conclude the document with our insights and their meaning towards the answering of the research questions. We also talk about some of the future work that can be an offshoot to this work.

Chapter 2

Review of Literature

2.1 Problems with Traditional Grading

Samson et al. examined 35 different studies to understand how college grades, such as GPA, correlate with job success. Success was measured by various criteria, including job position, job satisfaction, income, performance ratings, and other factors. It is commonly believed that a student's performance in college predicts their success in the professional world post-graduation. However, the researchers found that the mean correlation between GPA and job success across all these studies was only 0.16 [34]. Grades and test scores accounted for only 2.4% of the variance in career success. In simpler terms, the grading methods employed by colleges do not seem to align strongly with the skills, knowledge, and traits valued and rewarded in the U.S. job market [27].

In the partial-credit system, grades and all other college assessments become akin to a game. The goal is to accumulate as many points as possible for your grade while minimizing the amount of time and effort expended. This is taken seriously by students because grades significantly impact a student's future endeavors [27]. This can lead to several issues:

- **Anxiety**: Traditional grading can induce anxiety among students. A study found that over half of the students reported feeling overwhelming anxiety, and around

13% sought professional help for anxiety within a year. Additionally, more than 31% reported feeling so depressed that it became difficult to function, with 11% seeking professional help for depression within the same time frame [1, 27]. Given that about 61% of teens already experience stress related to getting good grades [25], poorly designed grading practices can exacerbate this stress and increase the number of students who experience it. These practices can also be excessively punitive, potentially increasing the failure rate of students. Fear of failure was found to be a major contributor to student stress in high school [19, 46]. In general, classes utilizing "conventional" grading systems exhibit these symptoms because they often prioritize the end result rather than the learning process, with instructors exerting more control over the assessment process compared to students [13].

- **Grade-Grubbing**: The pressure from this stress can lead to "grade-grubbing," where students frequently complain about their grades and pressure their teachers to give them higher scores, even if they do not deserve them.

- **Unethical Behavior**: Another common way students try to "win the game" is by cheating and plagiarizing. Recent surveys have found that many students admit to copying text and answers from unauthorized sources for their homework, tests, and other major assessments [17, 27].

In traditional grading systems, grades are often falsely inflated due to rewarding behavior as opposed to learning. This practice rewards students unequally. It rewards students who are compliant and put in high effort (high quantity) but punishes students who are in the transition phase of understanding how to achieve learning outcomes or "do college." These practices reward quantity over quality and reward students based on the amount of work they complete, rather than on what they know [12, 29].

Grades and student reports are external factors that can affect student motivation [2]. Low grades can hold little or no value to students, and psychologically, students only count or pay attention to high grades [35] (Note that this study was performed only on students with learning disabilities). A possible consequence of this could be that poor grades can cause students to withdraw from courses rather than make an effort to improve their grade. Thus, grading systems that are extremely punitive can have an add-on effect on students in that they can inhibit their growth mindset as students disregard feedback in the form of low grades.

2.2 Equitable Grading Practices

Student motivation can be enhanced by various factors, such as clear understanding of the learning outcomes required to achieve a certain grade, visibility and comprehension of their achievements, and a positive and supportive environment rather than a negative and punitive one [28].

Promoting a risk-free and penalty-free environment where students can make mistakes without fear of grade loss can foster trust and relationships between students and instructors (teachers and teaching assistants) [11]. Traditional grading practices, on the other hand, offer limited opportunities for students to engage in lectures, labs, and exams. Allowing retakes on assessments, such as program assignments and labs, can promote a risk-free atmosphere. This approach enables students to review their work, learn from mistakes, and work with instructors to improve the quality of their work. According to Feldman [11], this approach can not only reduce failure rates and improve student performance, but also alleviate stress. Speaking on one his results, he states: ”Students were less stressed and were grateful ... to have flexibility to turn in assignments after a deadline,

and to be allowed to retake exams."

The Triage Theory of Grading, proposed by William J. Rapaport [32], offers an alternative approach to evaluating student work. Instead of the traditional range of intermediate grades, this theory simplifies the grading process by assigning only three possible grades:

1. Full Credit ("A" grade): An item receives full credit if it is clearly or substantially correct.
2. Minimal Credit ("F" grade): An item gets minimal credit if it is clearly or substantially incorrect.
3. Partial Credit ("C" grade): If an item falls somewhere in between i.e., it is neither clearly correct nor clearly incorrect. It receives partial credit.

This system does not assign any other intermediate grades. The triage theory aims to make grading more objective, fair, and transparent for students, while also reducing the burden on instructors.

Traditional grading suffers from the problem of students asking for "just a few more marks" to clear a borderline grade. Rapaport believes that intellectual work should be assessed via criticism. However, he notes that criticism can be subjective and that grades are necessary as they are required by most educational institutions and the industry as a measure of quality. Triage grading solves problems with traditional grading by emphasizing clear correctness and minimizing subjectivity. It aligns with the push for more socially just practices in education, ensuring that what is modeled for teachers reflects research in the field.

Rapaport suggests a refined grading scale, with four grades instead of three. He creates a fourth grade and reserves it for work that was worse than wrong, such as when the sub-

mission was blank or when it is clear that the answer is completely wrong. Thus, he proposes the following modified scale [32]:

1. Assignment done, and clearly adequate (full credit) = 3
2. Assignment done, but not clearly adequate or inadequate (partial credit) = 2
3. Assignment done, but clearly inadequate = 1
4. Assignment not done = 0

When employed incorrectly, traditional grading practices focus on behavior rather than a measure of a student's skill or knowledge [12]. Equitable grading practices can be applied to assess skill or knowledge as opposed to behaviors. These practices emphasize the belief that all students can learn and meet learning targets [41]. Traditional grading using a 100-point scale, which typically distributes students over a normal distribution, can prove to be inherently inequitable. Using a 100-point grading scale, the initial 40 percentage points are evenly allocated. However, this scale does not represent a student receiving a 0 correctly. In the scale typically, the grade "A" spans 90–100 percent, B spans 89–80 percent, and so forth, with a D falling between 69 and 60 percent. However, when students receive a zero, the reduction is not proportionate; instead, it entails a substantial 60-percent deduction. This particular grading approach places students who receive a zero in a challenging position, as the considerable impact makes it mathematically arduous for them to recover their overall grade. To correct this, Feldman [11] advocates that rather than using a 0 points for a student submission (e.g. no submission), we use "minimum grading", meaning that we instead award the student 50 points for such a submission. Additionally, the use of a 4-point scale rather than a 100-point one can reduce the ambiguity and increase the standardization of grading [32].

2.2.1 EMRN Grading Scheme

The EMRN rubric was created by Rodney Stutzman and Kim Race [40]. It originated as the "EMRF" rubric in a 2004 article in Mathematics Teacher magazine. In our scheme, the only modification made was replacing "F" (which students often associate with "failing") with "N" (for "Not Assessable"). As with Rapaport's scale, the EMRN scale also uses four levels, described this way by Stutzman and Race [40]:

- Excellent (E): Meets or exceeds expections; Complete, clear communication; Clear understanding; Any error is trivial.
- Meets expectations (M): Understanding is evident; Needs some revision or expansion, but written comments are enough; No additional teaching is needed.
- Needs Revision (R): Partial understanding is evident but significant gap(s) remain; Needs more work/ teaching/ communication.
- Fragmentary (F): Clearly misunderstands; Insubstantial attempt made.

They emphasized using labels as markers of proficiency to percentages to get the students to focus more on learning outcomes and less on smaller differences in their result percentages, on writing a rubric and thus standardizing the grading process, thus eliminating bias, getting student buy-in by explaining the rubric to students and finally also eliciting student feedback on the rubric itself to allow for revisions.

Thus, the "EMRF" (in our case, "EMRN") grading scheme supports a standards-based specification grading criteria. It helps manage student expectations by defining the quality of work required from them and helps teachers create a more open and equitable process of grading [40].

2.2.2 Late Policies

Late policies are a component of equitable grading practices that aim to provide students with flexibility and support in managing their coursework. These policies allow for unexpected challenges or life situations that can impact a student's ability to submit assignments on time. By allowing for late submissions, these policies can reduce student anxiety and alleviate some of the pressures associated with strict deadlines. Thus, late policies can play a crucial role in creating a supportive and equitable learning environment that prioritizes student well-being and success.

Various versions of late policies have been employed across different studies. For instance, a study conducted by Kim et al. investigates student perspectives on assignment deadline policies [20]. They explore the balance between flexibility and structure in deadline policies specifically comparing different deadline policies and assess their benefits and drawbacks. These include: "strict deadlines with no extensions, deadlines with limited extensions, deadlines that could be extended with notifying instructor, soft deadlines that could be extended without notifying instructor, no deadlines, and deadlines with an early incentive". They found that "none of the students selected strict deadlines or no deadlines as their favorite policy. Instead, most participants preferred a policy that provided both structure and flexibility, including limited extension, extension with notification, and soft deadlines allowing extension without notification."

Another study by Hott et al. [16] delves into the impact of a flexible late policy on student performance and submission behavior in two semesters of a lower-level required course (LL) and an upper-level elective (UL). Initially, both courses had stringent late policies, with the LL course imposing a 10% penalty per day for two days and the UL course rarely allowing late submissions. Subsequently, for both courses the late policy was modified

to offer two no-penalty late days for each assignment. Findings indicate that in the LL course, grades declined with each day past the deadline under the strict policy, and the introduction of no-penalty late days led to more submissions after the deadline and lower overall homework scores. Conversely, in UL, students submitting two days late in the second semester performed comparably to those with stricter penalties, and average scores for assignments submitted on or around the deadline notably improved.

Complementary to the studies above that demonstrate the relationship between late policies and student performance, a study by Chen et al. [4] aims to address some practical challenges of large CS-1 courses. These challenges include long wait times and overworked staff during office hours in computer science courses, particularly near assignment deadlines. Their study implemented a "split deadlines" policy where students were randomly divided into two groups with staggered release and due dates for assignments. The results showed improved office hour efficiency with decreased wait times, and no significant difference in student performance between the two groups despite different due dates. Student feedback indicated that most students perceived the policy as fair and effective. Thus, the concluded that split deadlines policies could enhance office hour efficiency without compromising student performance, thus addressing the pressing issue of limited resources for grading, instructors and TAs.

2.2.3 Test Anxiety, Test Performance and High Stakes Assessments

Hembree et al. [15] conducted a meta-analysis on over 562 studies related to academic test anxiety. They found that test anxiety leads to poor performance in tests. They also discovered that test anxiety has an inverse relationship with students' self-esteem and a

direct relationship with their fears of negative evaluation. They further found that "Contrary to prior perceptions, improved test performance and grade point average (GPA) consistently accompany TA reduction."

Von Der Embse et al. [45] examined the levels of test anxiety on high-stakes tests and the differing effects based on the students' socio-economic status. They noted that high-stakes exams can significantly influence a student's future success and found that students who experience test anxiety have lower achievement scores on high-stakes tests, which can result in many negative consequences for a student. They also acknowledged the cyclical effects of test anxiety being caused due to low performance, which in turn causes further low performance in the future. They suggested various interventions such as those based on verbal instructor communication about exams being an opportunity to demonstrate mastery as opposed to exams being high stakes and having severely negative consequences if a student fails them, and relaxation techniques such as deep breathing and muscle relaxation. However, it is important to note that these interventions are student-centric and do not change the pedagogical structure of the courses.

Silaj et al. [38] aimed to study the effects of the relative weight and quantity of assessments on test anxiety. To achieve this, they experimented with two groups: a smaller seminar-style class with six low-stakes quizzes (10% of the final grade each) and a larger lecture-style class with two high-stakes exams (40% of the final grade each). Results showed that higher post-state anxiety was linked to poorer performance in both formats, but anxiety decreased over time in the seminar-style class while remaining constant in the lecture-style class.

2.3 Self-Regulated Learning

Zimmerman [48] provides valuable insights into how self-regulation impacts learning and achievement. He defines self-regulation as the systematic process by which individuals activate and sustain their cognition, behaviors, and emotions. Functional self-regulation can prove useful to achieve one's learning goals. This perspective emphasizes the role of metacognitive strategies, effort regulation, and cognitive engagement in academic performance.

Zimmerman describes a model of self-regulation, which is a cyclical system that has three phases [48] also shown in Figure 2.1:

1. Forethought: Forethought pertains to influential cognitive processes that precede action and lead to it

2. Performance or Volitional Control: Performance or volitional control "involves processes that occur during motoric efforts and affect attention and action".

3. Self-reflection: Self-reflection encompasses processes that occur after-the-fact. They shape an individual's response to the experience. These introspective assessments, influence forethought in-turn thus perpetuating the cycle

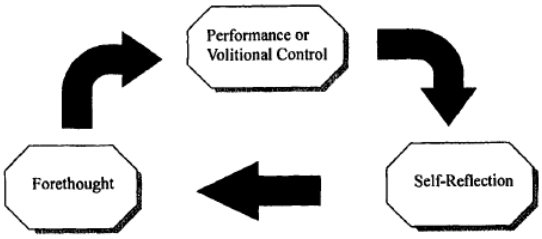


FIGURE 2 Cyclical phases of self-regulation. *Note*. From *Self-Regulated Learning: From Teaching to Self-Reflective Practice*. (p. 3), by D. H. Schunk and B. J. Zimmerman (Eds.), 1998, New York: Guilford. Copyright 1998 by Guilford Press. Reprinted with permission.

Figure 2.1: Zimmerman's cyclical self-regulation model

Additionally, Zimmerman suggests that proactive self-regulation results in better outcomes than reactive self-regulation. This he says, is because reactive self-regulators rely on after-the-fact task outcomes, which are often delayed, difficult to interpret, and socially stigmatizing, such as being a low-scoring student, as opposed to proactive self-regulators, who rely on goals, strategic planning, and have a sense of personal agency. Further, including pedagogical elements that promote proactive self-regulation should thus improve academic performance and outcomes. The EMRN grading system, as described further in Section 3.1.1, describes a clear mapping between learning outcomes and grades, which could promote planning and goal setting in students and thus promote the exhibition of proactive self-regulation.

Pintrich and De Groot [31] explored the relationship between motivational orientation, self-regulated learning, and classroom academic performance. Their study involved seventh graders from science and English classes. They found that self-regulation, self-efficacy, and test anxiety predict performance outcomes. Additionally, they developed and validated a questionnaire titled "Motivated Strategies for Learning Questionnaire" (MSLQ) [30], which is self-reported and contains questions that test self-regulation, intrinsic value, test anxiety, help-seeking, etc. The MSLQ scales of effort regulation and help-seeking map well onto self-regulated behaviors, and the MSLQ scales of critical thinking and metacognition map onto self-regulated cognition. Assessing these measures in students could be correlated with better academic outcomes.

2.4 Grading and Feedback

"Grades should provide feedback to students, document their progress, and help teachers make decisions about what instruction a student needs next." Timely feedback incorpo-

rates accountability into a student and improves their learning outcomes. "Assessment and feedback, particularly during the course of learning, are the most effective ways for students to learn accountability in their work and personal lives." [29, 47]

Providing "criterion-referenced feedback", which is feedback that informs students of their "standing relative to a target level of knowledge or skill, gives students more information about their learning than norm-referenced feedback, which tells them how their performance ranks relative to the performance of other students" [24, 33]. According to Marzano, rubrics can also serve as excellent tools for "providing students with criteria that describe specific levels of performance for content that is informational in nature as well as content that is process oriented" [24]. For example, having discrete levels of performance, where each level indicates a certain degree of a student's demonstration of knowledge and skill and providing students with these levels can help students better achieve learning outcomes.

Chapter 3

Methodology

This chapter discusses the undergraduate course in which the Equitable Grading Practices (EGPs) were implemented, the practices introduced in the two iterations of the course, and the experimental approach to gathering observations for this qualitative study. We first describe the syllabus and how EGPs were employed in the course. We supplement this with additional details and information about the employed EGPs, which the syllabus does not cover. We then describe the choice and structure of the course. We conclude the section by delving into details about the instruments that we employed to gain student perceptions about EGPs to answer the research questions.

In this thesis, we focus on gathering student perceptions about specific equitable grading practices employed within the context of two instances of CS-1114. While a wide variety of equitable grading practices exist, including different versions of specification grading, various approaches to late policies such as token-based systems, and different structures for quizzes and exams, our study narrows its scope to the practices implemented in our course. These include the EMRN grading scheme, a 7-day late policy extension with a maximum attainable grade, and the use of 8 smaller quizzes in place of midterms. By focusing on these specific practices, we aim to gain insight into how students perceive these particular strategies and how they may impact their learning experiences and outcomes.

3.1 The Course

The EGPs were implemented in the course "CS-1114: Introduction to Software Design" at Virginia Tech's Blacksburg campus, which is typically taken by undergraduates in their first year of college. The experiment and observations include those of the students who took this course in the Fall 2023 and Spring 2024 semesters.

This course is a CS-1 level standard undergraduate course taught during the regular dates of a semester. Both course versions were taught in-person and did not have any remote components. The instructors for these two versions were different. This course does not have any pre-requisites and is open to all majors. It covers data structures and their implementation in Java.

While the course version taught during the Fall 2023 semester was held for 16 weeks, the version taught during the Spring 2024 semester was also held for 16 weeks, excluding national holidays and mid-semester breaks.

Pertaining to the aforementioned research questions from Section 1.3, the following sections present text directly taken from the course syllabi. This does not include the fourth research question, which involves grounding the results from the survey and interviews using the MSLQ.

3.1.1 EMRN Grading scheme

As mentioned before, In CS-1114, one of the aspects of the Equitable Grading Practices (EGPs) is employed through the use of the "EMRN Grading Scale". Of course, when compared to a traditional 100-point grading scale, the EMRN grading scale is simpler and coarser. With respect to the EMRN grading scheme, the syllabus mentions:

"Grading for the course is based on the distribution of credit shown in the table above. Final letter grades for the course will be determined using the following grading scale, based on the percentage of possible points achieved:

A	**A-**	**B+**	**B**	**B-**	**C+**	**C**	**C-**	**D+**	**D**	**D-**	**F**
$\geq 93\%$	$\geq 90\%$	$\geq 87\%$	$\geq 83\%$	$\geq 80\%$	$\geq 77\%$	$\geq 73\%$	$\geq 70\%$	$\geq 67\%$	$\geq 63\%$	$\geq 60\%$	$\geq 0\%$

Table 3.1: Grading scale

Grading of lab and programming assignments in this course will be a bit different than you typically see in other courses. We will be using a form of specification grading, where the focus is on mastering the required skills rather than point deductions for individual mistakes. Lab and programming assignments are instead graded on a four-valued categorical scale that is similar in spirit to a pass/fail evaluation.However, you can repeat and resubmit lab and program assignments as much as necessary to ensure you pass the assignment.

Lab and programming assignments will be assessed on the following EMRN scale. Each assignment will receive a grade that is one of:

Grade	**Meaning**
(E) Excellent	Exceeds expectations with no nontrivial errors.
(M) Meets Expectations	Meets expectations by clearly demonstrating intended skills, with no significant gaps or errors, although some revision is needed.
(R) Revision Needed	Partial understanding/skill is shown, but significant gaps or errors are present requiring revision.
(N) Not Assessable	The work is too fragmentary or has too many gaps to assess.

Table 3.2: Grading criteria and meanings

Remember that you can revise and resubmit these assignments (within a defined period) to increase your grade to the EMRN value you wish. However, Excellent ratings are only available for work that is completed on time before the assignment deadline. Work turned

in after the posted deadline will only be eligible for a maximum M grade. For lab and program assignments, E grades will be counted as 100% credit when computing your grade. M grades will count as 80% credit, R grades will count as 50% credit, and N grades will not receive credit."

It is important to note that the scale is tilted towards the higher end of the range, i.e. there is a straight jump from 0% to 50%, which continues on to 80% and finally 100%. By design, this makes it easier for a student to score at least 50% credit in their labs and programming assignments, which corresponds to a "F" grade. In practice, as long as a student makes a submission by putting in bare minimum work, they are awarded a 50% grade. This means that the likelihood of a student scoring a 0% is quite low, as this grade is usually only meted out for non-submissions.

Merits of the EMRN Grading Scheme

1. **Nuanced Assessment:** The EMRN grading scale allows for a nuanced assessment of student performance, offering detailed feedback beyond a binary pass/fail system. This coarse granularity enables a more accurate reflection of a student's understanding and skills.

2. **Opportunity for Revision:** The "Revision Needed" category acknowledges that learning is an iterative process. Students have the opportunity to revise and resubmit assignments to improve their understanding and skills, which promotes a growth mindset.

3. **Focus on Mastery:** The emphasis on mastering required skills, rather than point deductions for individual mistakes, aligns with the principles of specification grading. This approach fosters a focus on understanding and achieving proficiency in key

concepts.

Challenges of the EMRN Grading Scheme

1. **Unfamiliarity and Complexity:** The EMRN grading scale is novel to the majority of students taking the course, who have so far been exposed to typical point-based grading. This coupled with its multiple categories and credit percentages, may be perceived as complex. Students might find it challenging to navigate and understand the implications of each grade category. Naturally, this unfamiliarity entails acclimation time for students. This is discussed further in Section 5.1.1.

2. **Subjectivity:** While every programming assignment and lab has clear instruction, which describe the mapping between learning outcomes and the EMRN scale, a student may still disagree with this. For example, we found that some students think that the "Excellent" grade should not assess the comment quality of the code. This personal disagreement may cause resentment towards the EMRN grading scheme scale and the student may want to be graded using traditional schemes.

3. **Potential Stress for Students:** While the opportunity for revision is beneficial, the pressure to achieve an "Excellent" rating on assignments completed before the deadline might create additional stress for students, potentially impacting their overall learning experience.

3.1.2 Late Policy

The EGPs employed by us also consist of the late policy. It applies to multiple assessments including weekly readings, assignments (programs), labs and quizzes. The following list describes the late policy for these assessments:

1. **Readings**: "If you miss the due date, Reading Activities can be completed up to 7 days late. Late submission does not affect your grade, but does take up time you should be spending on the next Reading Activity. Also, students are responsible for the material in the covered reading assignments on each in-class quiz, whether or not the reading has been completed on time."

2. **Programs**: "If you are unable to complete your program assignment by the deadline, you may continue to submit your program assignment up to 7 days late. Late submissions are limited to a maximum grade of Meets Expectations (the opportunity to earn a grade of Excellent is only available for submissions completed on time)."

3. **Labs**: "If you are unable to complete your lab by the end of your assigned lab period, you may continue to submit your lab assignment up to 7 days late from the end of your lab. Late submissions and submissions without lab attendance are limited to a maximum grade of Meets Expectations (the opportunity to earn a grade of Excellent is only available for submissions completed on time with lab attendance)."

4. **Quizzes**: late policy: "students must take in-class quizzes at the scheduled time. If you are forced to be absent from class on the day of a quiz due to an emergency or illness, you must inform your instructor immediately and obtain verification of absence from the Dean of Students Office. Makeups for missed quizzes are at the discretion of the instructor and require a clear explanation of the reason for the absence together with official absence verification."

In CS-1114, the late policy is designed to accommodate students' varying schedules and unforeseen circumstances, providing flexibility while maintaining a fair and structured approach to assignment submissions. The late policy extends the deadline by seven days in all cases. It is designed to ease the timeline, but at the same time not be too lax. It is

another tool given to the students designed to be something that is used in moderation. While the late policy gives students the ability to turn in their work a week later, in doing so they also inadvertently or otherwise eat up the time available for next week's assessments. The same is also mentioned in the official syllabus of the course.

Thus, the late policy in CS-1114 is designed to strike a balance between providing flexibility for students facing unexpected challenges and maintaining a structure that encourages timely completion of assignments and assessments. It aims to nurture a supportive learning environment while emphasizing the importance of meeting deadlines to optimize the learning experience. Student perception around the late policy is discussed in greater detail in Section 5.2.

Merits of the Employed Late Policy

1. Flexibility: The late policy allows for flexibility in accommodating unexpected circumstances that may arise for students. This can be particularly beneficial for students facing personal or health-related issues, ensuring they are not unduly penalized for factors beyond their control.

2. Popularity: Implementing a late policy that provides some leeway for late submissions may increase student satisfaction. Students may appreciate the understanding and flexibility offered by such a policy, which can contribute to a positive learning environment.

Challenges of the Employed Late Policy

1. No encouragement of good time management: One challenge of a lenient late policy is that it may not effectively incentivize students to manage their time effectively.

Knowing they have some flexibility with deadlines, students may be less motivated to submit work on time, potentially leading to procrastination and lower quality work.

2. Chronic lateness: Another concern is that a lenient late policy could contribute to a culture of habitual late submissions. If deadlines are consistently extended without significant consequences, students may develop a habit of submitting work late, which can be detrimental to their overall learning experience and the efficiency of the course.

3.1.3 Smaller Quizzes and Quiz Dropping

As previously mentioned, we use quizzes as opposed to bigger exams, midterms, or finals. There are 8 quizzes held in total, with one of the quizzes being able to be "dropped" by the students (not counted towards the grade). This approach aims to promote a continuous and manageable evaluation process, allowing students to consistently engage with course material and receive timely feedback. The quizzes, administered approximately every other week, play a crucial role in gauging students' understanding of the ongoing coursework. With regard to quizzes, the syllabus mentions:

- "Given in-class approximately every other week."
- "This course has no midterm exams or tests–these more frequent, shorter in-class quizzes are used instead."
- "Students in face-to-face lecture will see these as pop quizzes given during class. Students in online lecture sections will see these show up on Canvas during your scheduled synchronous meeting time."

- "Quizzes are "closed notes, closed web"–no outside information, tools, or resources can be used while taking quizzes."
- "At the end of the semester, your lowest quiz grade will be dropped when calculating final grades."

The deliberate choice of smaller quizzes over comprehensive exams aligns with a pedagogical consideration – the mitigation of test anxiety. Test anxiety is a common challenge faced by students, which can be alleviated through a more frequent and low-stakes assessment structure. The smaller quiz format reduces the pressure associated with high-stakes exams, providing students with regular opportunities to demonstrate their knowledge without the overwhelming weight of a single, decisive assessment event. The true capability of a student is thus more accurately determined by their skills and knowledge as opposed to traditional exams, where their ability to handle test anxiety is an often unnecessary inclusion.

By breaking down the assessment components into bite-sized quizzes, students are encouraged to approach learning as a continuous process rather than one-off performances centered around major exams. This approach not only promotes consistent engagement but also allows for incremental skill development and knowledge retention.

Moreover, the regular quizzes contribute to a dynamic learning environment where students can track their progress, identify areas for improvement, and adapt their study strategies accordingly. The ongoing nature of these assessments fosters a proactive and adaptive mindset, nurturing a deeper understanding of the course material.

The implementation of smaller quizzes in CS-1114 not only is meant to address issues around evaluation using bigger exams as well as test anxiety. The smaller quiz size and increased quiz frequency encourages a more sustained and less stressful learning experience,

contributing to a positive and supportive academic environment for students. Student perception around this design choice is discussed further in Section 5.3.

Merits of Smaller Quizzes

1. Assessment Frequency: Smaller quizzes offer the advantage of more frequent assessments, providing students with regular feedback on their understanding of the material. This can help students stay engaged and motivated in their learning process, as they receive continuous reinforcement and guidance.

2. In-class format: Smaller quizzes that can be completed in class save time compared to having two longer exams. In-class quizzes can be administered efficiently, allowing instructors to assess student understanding without the need for extensive preparation or grading time outside of class.

Challenges of Smaller Quizzes

1. Grading for absent students: With more frequent quizzes, grading can become more challenging, especially if there are absent students who need to make up missed quizzes. Clear and well-defined makeup policies are essential to ensure fairness and consistency in grading.

2. Dropping quiz grades: Allowing students to drop one or more quiz grades can lead to reduced effort or engagement with quizzes. Students may prioritize other tasks or courses if they know they have the option to drop a quiz grade, potentially impacting their overall learning experience.

3. Seriousness of quizzes: If students know they have the option to drop one quiz grade, they may not take all quizzes as seriously as they should. This could affect their

preparation and performance on quizzes, ultimately impacting their learning outcomes and understanding of the material.

3.2 Survey

To capture student perceptions around the grading practices implemented in CS-1114, a structured survey was employed. The survey was administered during two distinct semesters, namely Fall 2023 and Spring 2024, enabling a broad examination of student perspectives across different groups. The survey instance deployed in the Fall 2023 semester was a post-survey, as it was deployed in week number 13 of the semester. On the other hand, the survey instance deployed in the Spring 2024 semester was a pre-survey, as it was deployed in week number 2 of the semester. This temporal diversity allows for the identification of potential variations in perceptions based on evolving course dynamics, student optimism, instructional adaptations, and student demographics. The complete survey can be found in Appendix B.

3.2.1 Motivated Strategies for Learning Questionnaire (MSLQ)

The survey instrument was based on the Motivated Strategies for Learning Questionnaire (MSLQ)[30] developed in 1991. This questionnaire and the corresponding manual served as a foundational basis for constructing a survey tailored to the specific context of the course.

The MSLQ is renowned for its effectiveness in assessing students' motivations and learning strategies in a college-level course. The MSLQ is divided into two sections: student motivation and learning strategies, with the former containing 31 questions and the lat-

ter containing 50 questions. The MSLQ covers a wide variety of question categories aimed at motivation scales and learning strategies such as Intrinsic Goal Orientation, Test Anxiety, etc. However, we selected only a subset of these categories and questions, which were directly applicable to our course. These categories include:

1. Intrinsic Goal Orientation
2. Metacognitive Self-Regulation
3. Critical Thinking
4. Test Anxiety
5. Self-efficacy for Learning and Performance
6. Peer Learning
7. Help seeking

Each question asks the student to indicate their alignment with the question on a 7-point Likert scale, with "1" being "not at all true of me" and "7" being "very true of me". Some questions on the scale are reversed. Each question can contribute to the category positively or negatively. For example, a question judging the "Task Value" is "I am very interested in the content area of the course", which contributes positively to this category. However, a counterexample is under the category "Metacognitive Self-Regulation" the question "During the class, I often miss important things because I'm thinking of other things", which contributes negatively to this category.

Leveraging this instrument provided a solid theoretical foundation and allowed for the incorporation of nuanced questions related to equitable grading practices, the EMRN grading scale, late policy, and the frequency of smaller quizzes.

3.2.2 Refining the MSLQ

The MSLQ is general in nature and can be applied to a wide range of contexts, provided they are targeting tertiary education. However, its wording is a bit dated as it was developed about 33 years ago. To refine the survey instrument and ensure its relevance to the context of CS-1114, insights were gleaned from the study titled "Measuring medical students' reflection on their learning: modification and validation of the motivated strategies for learning questionnaire (MSLQ)"[39]. These modifications aimed to make the survey more modern and consider unique aspects of the CS-1114 course, providing a better understanding of how students perceive the implemented grading practices. The modifications made to the scale are described by [39] in their table titled "Table 1 Modifications of MSLQ selected items".

3.2.3 Structure of the Survey

The survey was structured into three parts: Likert scale questions, open-ended long-text questions, and questions for gathering email addresses, consent, and such. The full list of questions is given below:

1. Q1: Gathering the email address of the respondent
2. Q2-4: Multiple choice, Likert Scale sub-questions referenced from the MSLQ and refined as mentioned in Section 3.2.2. Also contains the EMRN grading scheme and late policy Likert scale questions which are discussed more in the next Section 3.2.4.
3. Q5-7: Open-ended long-text questions
   (a) Q5: What was the most beneficial aspect of the grading policies in this course?

(b) Q6: What was the most anxiety-provoking aspect of the grading policies in this course?

(c) Q7: How can we improve the grading policies in this course?

4. Q8: Gathering consent for student survey responses along with a check that the student is 18+

3.2.4 Additional EMRN Grading Scheme Subscale

To gather student perception about the EMRN grading scheme, we added a subscale to the beginning of the Likert scale section. It should be noted that while the questions from the MSLQ have been validated [18, 22, 30], the questions included in the EMRN subscale have not been validated. However, they are aimed at testing grading perception and efficacy of the EMRN grading scheme. Note that while the MSLQ questions use a 7-point Likert scale, the EMRN grading scheme subscale uses a 5-point Likert scale instead. This choice was made to simplify these questions and choices around them. Even more so, because it has not yet been validated, and having a smaller range reduces the potential misrepresentation of response distribution by restricting it to a smaller range. The questions are present in table 3.3. We should note here that the 5th question in this subscale, which is "EMRN grading increases my anxiety in this course" aims to make the student comment about their anxiety in the course, related specifically to the EMRN grading scheme. This is different from the Test Anxiety subscale questions as described further below.

Subscale	Question
EMRN grading scheme	Earning a good grade is a strong motivating factor for me
EMRN grading scheme	I am very familiar with the EMRN grading scale
EMRN grading scheme	EMRN grading is a reliable way of assessing my skills and knowledge
EMRN grading scheme	EMRN grading causes me to focus more on what I should learn in each assignment
EMRN grading scheme	EMRN grading increases my anxiety in this course
EMRN grading scheme	EMRN grading means I don't have to focus on earning each individual point on an assignment
EMRN grading scheme	I prefer EMRN grading to using regular points for grading
EMRN grading scheme	I expect my course grade to be higher because of EMRN grading

Table 3.3: Included subscale on the EMRN grading scheme

3.2.5 Additional Late Policy Subscale

To gather student perception about the late policy, we added a subscale to the beginning of the Likert scale section. Apart from the details are identical as the previous section, which talks about the additional EMRN grading scheme subscale (Section 3.2.4). The questions are given in table 3.4.

Subscale	Question
Late policy	The late policy reduces my stress about assignment deadlines
Late policy	The late policy gives me the flexibility I need to deal with unanticipated situations
Late policy	The late policy is unfair
Late policy	The late policy encourages me to start working on assignments earlier
Late policy	The late policy makes it easier for me to delay working on assignments until the last minute
Late policy	The late policy helps me succeed in this course

Table 3.4: Included subscales on the Late policy

3.2.6 Scoring the Likert Scales

The survey responses from questions 2 through 4 were scored using averaging and summing as described further in Section 4.2 under Results.

3.2.7 Structure and Analysis of the Free Responses Questions

Analyzing the open-ended responses is crucial for gaining a nuanced understanding of how students perceive the grading system within the course. Unlike closed-ended questions that offer predefined options, open-ended questions allow participants to express their thoughts freely, providing rich qualitative data. Through this analysis, we can identify patterns, common concerns, and suggestions for enhancement, which can inform future iterations of the grading policies to better cater to students' needs and optimize their learning experience.

The survey responses from questions 5 through 7 were analyzed using Grounded Theory Analysis[14]. Grounded Theory is inductive, meaning it starts with specific observations and data rather than preconceived hypotheses. The collection and analysis of data are followed by the identification of patterns, themes, and concepts that emerge from the data itself. This is achieved by first coding and categorizing the data, which breaks the data into separate components. Then, focused coding helps identify core categories and concepts. The constant comparison process helps refine and clarify these categories by removing repetitions, combining two very similar categories, and removing infrequent categories. Grounded theory aims for theoretical saturation, which means that data collection continues until new data no longer contribute to the development of new concepts or the refinement of existing ones. This helps ensure that the theory is well-developed and comprehensive. Lastly, the final output of grounded theory research is an emergent theoretical

framework or a set of concepts that explain the studied phenomenon. These theories are grounded in the data and can be used to guide further research or inform practical applications. Since the hypotheses are developed from the data rather than the other way around, grounded theory helps eliminate bias and explain the causality of ideas that form the structure of the theory.

The two survey instances were coded separately, and then emergent themes were identified from the two sets of codes. Omissions include empty responses, those marked as a non-response by a student such as ”N/A”, or an answer to just fill up the form field, and responses which do not have enough data or are incoherent or ambiguous. For example, when asked ”What was the most beneficial aspect of the grading policies in this course?”, a student responded ”The grade I got”, which could mean that either they got exactly the grade they were expecting or that the course was easy and they were pleased with their grade. The coding was done separately for different questions, and then the data was combined across questions to see if the results observed per question translate similarly across the three questions.

3.2.8 Survey Deployment

The first survey was a post-survey deployed towards the end of the semester, and the second survey was a pre-survey as it was deployed earlier in the semester. The survey was deployed as a course assignment having “0” credits within the CS-1114 course page on Canvas. Thus, not completing the survey, completing the survey partially, completing the survey fully and giving consent for data collection, or completing the survey fully but denying consent all led to the same result with respect to the number of points awarded to students, i.e., there were no direct academic consequences for the students introduced by the

survey.

3.3 One-on-One Interviews

As previously mentioned, while the Likert scale and open-ended long-form questions in the survey are beneficial, they are insufficient to capture a deeper understanding of student perceptions. A more extended personal interview can be used to delve deeper into the issues that students face. Reasons for this include the ability to capture emotion through facial expressions, body language, and voice, the ability to ask follow-up questions, the ability to cross-question, and allowing the student to deliberate over their responses over a greater amount of time.

Thus, personal interviews lasting about 50 minutes to an hour were conducted both in-person and over Zoom (with the video turned on) to supplement the survey responses.

The full script of the personal interviews, including the consent readout, can be found in Appendix E.

3.3.1 Timeline

For the Fall 2023 semester, we aimed to interview students in the latter half of the semester, i.e., in the last 4-6 weeks. On the other hand, for the Spring 2024 semester, we aimed to interview students in the first half of the semester from week 6 through 11.

A logistical concern around this was the analysis of results and writing of this thesis. This is the reason we decided not to conduct any more interviews beyond week 11 of the semester.

3.3.2 Recruitment

We aimed to recruit students from both semesters to participate in the interviews. We wished to divide students into various groups according to their performance. The reason for this was that we hypothesized that students belonging to different performance groups would respond differently to the Equitable Grading Practices. Specifically, students who are high performers and students who are low performers would not be influenced greatly because in the case of high performers, they already have skills which are suitable to adapt to these new practices and be successful, regardless. On the other hand, while the low performers could potentially see an improvement in their course performance, they lack the basic skills and motivation necessary to perform in the course and the new grading practices could not influence this very much.

While the perceptions from these aforementioned groups were valuable, we were more interested in the student perceptions of the students belonging to performance groups that lie between these groups. While these groups have the basic skills necessary, we hypothesized that the alternative, more equitable grading practices could end up being more favorable for these groups and their perception of these practices could help guide further development to a greater degree.

In order to classify students based on performance, we needed their final grades, which we did not have access to. As a result, we needed a way to predict final student performance. After some trial and error with classification techniques, we settled on using the current student score to classify them into four quartiles. Thus, we were more interested in interviewing students from quartiles 2 and 3 as opposed to those from quartiles 1 and 4.

A Google form containing a calendar link and other questions was created for the purpose of slot booking. The calendar contained 1-hour slots in the aforementioned time period

for both semesters. This form was sent to students via email. The email template and the Google form in their entirety can be found in Appendices C and D respectively.

The following recruitment steps were followed:

- Fall 2023:

  1. We sent emails out to students from quartiles 2 and 3 in a staggered manner
  2. If they did not respond, we sent them a reminder email
  3. We then sent emails to more students from the same quartiles minus the students from step-1
  4. We finally sent emails to the entire class minus students from steps 1 and 3
  5. Lastly, in order to recruit more students, we sent a reminder email to all students around January 2024 (after the semester had ended and during Winter break)

- Spring 2024:
  Since we did not have a high sign-up rate in the previous semester, we decided to not opt for the staggered approach this time around. The sign-up rate is discussed further in Section 4.1.2.

  1. We sent emails out to all students from quartiles 2 and 3
  2. If they did not respond, we sent them a reminder email
  3. We then sent emails to more students from the quartiles 1 and 4
  4. If they did not respond, we sent them a reminder email

3.3.3 Logistics

A student could opt to attend the interview in-person or over Zoom. The interviews were held in 2000 Torgersen Hall, 620 Drillfield Dr, Blacksburg, VA 24060. The personal meeting room of Anmol Shukla was used as a Zoom meeting option. If held over Zoom, the participants were asked to turn their video on during the interview. If attending in-person the student was offered some snacks as compensation.

3.3.4 Interview Structure

The interview was divided into three parts: Consent form logistics, questions for the interviewee, and questions for the interviewer. The first five minutes were used for both parties to settle in and partake of any snacks. The next 45-55 minutes were used by the interviewer to ask questions of the interviewee. The last 5 minutes were reserved for any questions the interviewee may have.

The questions were divided into the following categories:

1. Personal History, experiences and perception with respect to grading
2. EMRN grading system
3. Late policy
4. Smaller quizzes vs bigger assignments

Chapter 4

Results

This chapter delves into the outcomes derived from the analysis of data acquired through two instruments: surveys and one-on-one student interviews. We describe and analyze the results of two survey instances conducted with students enrolled in CS-1114 in the Fall 2023 and Spring 2024 semesters, and six in-depth student interviews from the same semesters. By eliciting responses and perspectives from a diverse range of participants, we aim to present a comprehensive examination of the key findings, shedding light on patterns, trends, and valuable qualitative insights that contribute to the overarching themes of our study.

It should be noted that these results represent a subset of the total number of responses from both the surveys and interviews. The size of this subset is primarily determined by the number of students who consented to have their data included in the study. This is described in greater detail in Section 4.1.

Another important point to mention, as previously stated, is that our focus is on gathering student perceptions about specific equitable grading practices employed within the context of two instances of CS-1114. Our study narrows its scope to the practices implemented in our course. These include the EMRN grading scheme, a 7-day late policy extension with a maximum attainable grade, and the use of 8 smaller quizzes in place of midterms.

This chapter is structured as follows. We first describe the high-level sign-up and consent

statistics for both the survey and the interviews in Section 4.1. We then analyze the Likert Scale questions on the survey in Section 4.2 and describe the results pertaining to various subscales. Following this, we present the results of the Grounded Theory analysis of the interviews, as well as the Grounded Theory results of the free-response questions on the survey in Section 4.3.

4.1 Statistical Analysis

4.1.1 Survey

The survey was conducted in the Fall 2023 and Spring 2024 semesters. The Fall 2023 survey, deployed after a significant portion of the course had concluded, was a post-survey. Conversely, the Spring 2024 survey, deployed towards the beginning of the semester, was a pre-survey. We compare here and in section 4.2 the results of the surveys as pre and post-surveys, even though these survey instances were deployed on different student populations. We discuss this limitation in section 5.5.

The survey instances were deployed to the entire batch of students via a course assignment that had zero course credits within the CS-1114 Canvas course page.

In the Fall 2023 semester, the survey was viewed by 461 students. Out of these, 345 students (74.84%) completed the survey either partially or wholly. As mentioned in Section 3.2.3, Q.8 of the survey collected student consent to include their responses in the research study, along with an 18+ age check. Out of the 345 students who partially or wholly completed the survey, 217 (63.45% of 345) students gave consent for their data to be included in this research study.

In the Spring 2024 semester, the survey was viewed by 509 students. Out of these, 404

students (79.37%) completed the survey partially or wholly. This was an increase as compared to the Fall 2023 semester. This difference can be possibly attributed to a lesser student workload or "busy-ness" towards the beginning of the semester as opposed to later; recall that the survey deployed in Spring 2024 was a pre-survey and the survey deployed in Fall 2023 was a post-survey. Out of the 404 students who partially or wholly completed the survey, 276 (69.52% of 404) students gave consent for their data to be included in this research study. The increase in the percentage of students who gave consent could be possibly attributed to general privacy concerns or there being fewer students who were under 18 years of age.

These results have been summarized in Table 4.1.

Survey	# students who viewed	# students who partially or wholly completed	# students who gave consent
Fall 2023	461	345 (75%)	217 (47%)
Spring 2024	509	404 (79%)	276 (54%)

Table 4.1: Summary of student participation in survey instances

Thus, from the Fall 2023 survey, 47.07% of the responses were usable, and from the Spring 2024 survey, 54.22% of the responses were usable.

4.1.2 Interviews

This section centers on the sign-up rates derived from in-depth, one-to-one interviews with six students who were part of the Fall 2023 and Spring 2024 semesters. These interviews were designed to offer a deeper understanding of individual experiences and perspectives. They mostly supplement, yet at a few places complement the broader qualitative insights obtained through the survey. While the open-ended questions of the survey capture some perspectives, the word limit coupled with the low amount of time spent by a student an-

swering the question highly limits the amount of information the student is able to furnish. Moreover, a survey is also incapable of asking follow-up questions to student answers. Consequently, we were able to find some interesting points through these interviews, which we would have certainly missed if we had only relied on the survey.

While our initial aspiration was to engage with approximately 20 students across both the Fall 2023 and Spring 2024 semesters, logistical constraints and unforeseen circumstances led to a more limited sample size. Nevertheless, the narratives provided by the interviewed students furnish valuable qualitative data, providing a more detailed perspective on the topics under scrutiny and enriching our overall comprehension of the study's focal points.

We conducted one-on-one interviews with 5 students from the Fall 2023 semester and 1 student from the Spring 2024 semester.

Aforementioned in Section 3.3.2, we sent out an email inviting students to participate in a one-on-one interview lasting from 45 minutes to an hour. This email contained a Google Form link, which asked the students if they wished to participate in the research study through the interviews. If they marked their answer as yes, they would be asked to book a time on a public calendar link (Google Calendar was used in the Fall 2023 semester and Microsoft Outlook was used in the Spring 2024 semester) and then asked to fill in the booked time in the form. This process ensured that the students who marked "yes" on the participation question ended up booking a time.

In the Fall 2023 semester, we received a total of 14 responses. Out of these, 11 students indicated "yes" for participation and 3 students indicated that they would rather not participate in the study. Despite the higher than expected sign-up rate, only 5 out of the 11 students ended up attending the interviews. The remaining 6 students were absent due to health issues, conflicting appointments, or they were unresponsive no-shows. We reached

out to these 6 students after the no-show, but they did not end up participating in the interviews.

In the Spring 2024 semester, we initially invited the 2nd and 3rd quartiles. However, we experienced a sign-up rate of 0%. After experiencing such a low sign-up rate, we decided to invite the remainder of the class via email, as opposed to staggering the invites within the quartiles. In the end, a total of 7 students responded. This was half of the 14 students who had signed up in the Fall 2023 semester. Out of the 7 students who responded, 5 students indicated "yes" for participation in the study and 2 students indicated that they would rather not participate. Out of the 5 students, only 1 student ended up attending the interview. This was 20% of the sign-up rate from the Fall 2023 semester.

These results are summarized in Table 4.2.

Semester	**# students responded**	**# students who accepted**	**# students who showed up**
Fall 2023	14	11 (79%)	5 (36%)
Spring 2024	7	5 (71%)	1 (14%)

Table 4.2: Summary of student participation in one-on-one interviews

In conclusion, 36% of students from the Fall 2023 semester and 14% of students from the Spring 2024 semester ended up participating in the research study through interviews.

Another statistic from both semesters is the distribution in the student demographic with respect to their academic standing at the time of the interview. As mentioned before, we sent out emails to all the quartiles of students (with respect to their grade at the time of invitation).

4.2 Survey - Likert Scales MCQs

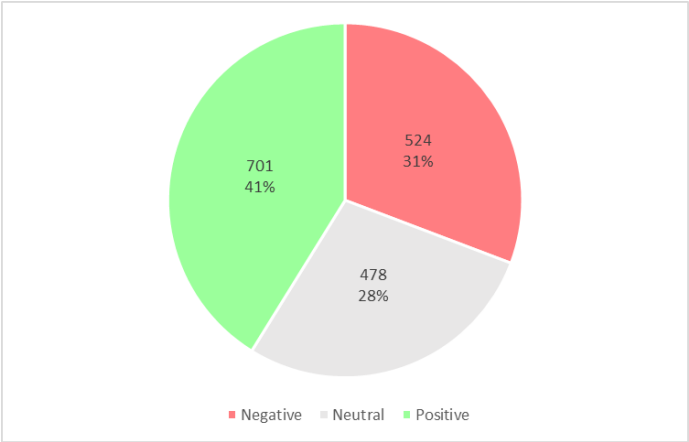
The survey contains Likert subscales for the EMRN grading scheme, late policy, and MSLQ subscales. The information gathered from these subscales helps us answer all research questions. As previously mentioned, as part of the analysis, we compare the results of the Fall 2023 semester, which was deployed in week 13 of the semester as a post-survey and the Spring 2024 semester, which was deployed in week 2 of the semester as a pre-survey. Although the contents of the survey are the same, the survey instances were deployed on different populations of students, who took the course in these two semesters. This limitation is described in section 5.5.

4.2.1 EMRN Grading Scheme Subscale

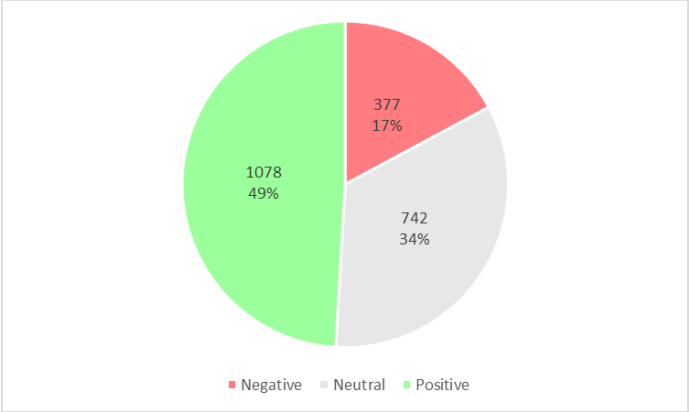
This subscale provides us information to answer RQ1: "Understand student perceptions of the employed specification grading scheme, called the EMRN grading scheme".

It is important to mention a few limitations in our results. Firstly, the grading of EMRN on a 100-point scale may have influenced student perceptions. This approach conflates perceptions about the grading scale with perceptions about the EMRN itself, making it difficult to separate the two. Secondly, the inconsistency in grading, with some aspects being graded on a 100-point scale and others using the EMRN grading scheme, further complicates the interpretation of student perceptions. A uniform grading system would have provided clearer insights into student opinions.

The method used for analyzing this subscale is different from the MSLQ subscales, which are discussed further in Section 4.2.3. Note that this subscale was a 5-point Likert scale with the response 1 being "strongly disagree" and 5 being "strongly agree".



(a) Fall 2023 (Semester Week 13)



(b) Spring 2024 (Semester Week 2)

Figure 4.1: EMRN grading scheme subscale - sentiment analysis

Firstly, we classified the Likert scale responses 1, 2 as negative, 3 as neutral, and 4, 5 as positive responses. Then we summed all the positive, negative and neutral responses. The results are shown in Figure 4.1.

For Fall 2023, as shown in Figure 4.1a, some students felt neutral about the system, with 27% (about a quarter) of the students responding 3 on the Likert scale. Students felt rather positively about the EMRN grading scale with about 41% marking either 4 or 5 and a lesser amount of students felt negatively about the scheme with about 28% marking either 1 or 2.

For Spring 2024, as shown in Figure 4.1b, again some students felt neutral about the system, with 34% (about a third) of the students responding 3 on the Likert scale. Again, the students felt rather positively about the EMRN grading scale with about 49% marking either 4 or 5 and a lesser amount of students felt negatively about the scheme with about 17% marking either 1 or 2.

It is noteworthy that students felt more neutrally and positively about the EMRN grading scheme in the Spring rather than the Fall, even though both groups of students felt it to be an overall positive influence on their learning.

Thereafter, we analyzed each question individually. For each question, we plotted the distribution of student responses (from 1 to 5) by summing them up. These questions are described in detail in Table 3.3. The result of this analysis is shown in Figure 4.2, and has 16 bar graphs; 8 graphs per semester; 2 graphs per question.

The questions that displayed a difference in distribution (and hence the shape in graphs) from figure 4.2 when comparing semester-wise were as follows. The questions that are not part of this list received almost consistent responses from students across both semesters.

Looking at the distributions across semesters in Figure 4.2, we can say that the first two questions, i.e., “Earning a good grade is a strong motivating factor for me” and “I am very familiar with the EMRN grading scale” are relatively weighted towards the positive end of the Likert scale, whereas the other questions are more neutrally distributed.

The results of the Mann-Whitney U-test [23] across both semesters are shown in Table 4.3. Here, we notice that there is no significant difference between the two semesters for any of the questions, as the p-value for all the tests is >0.05.

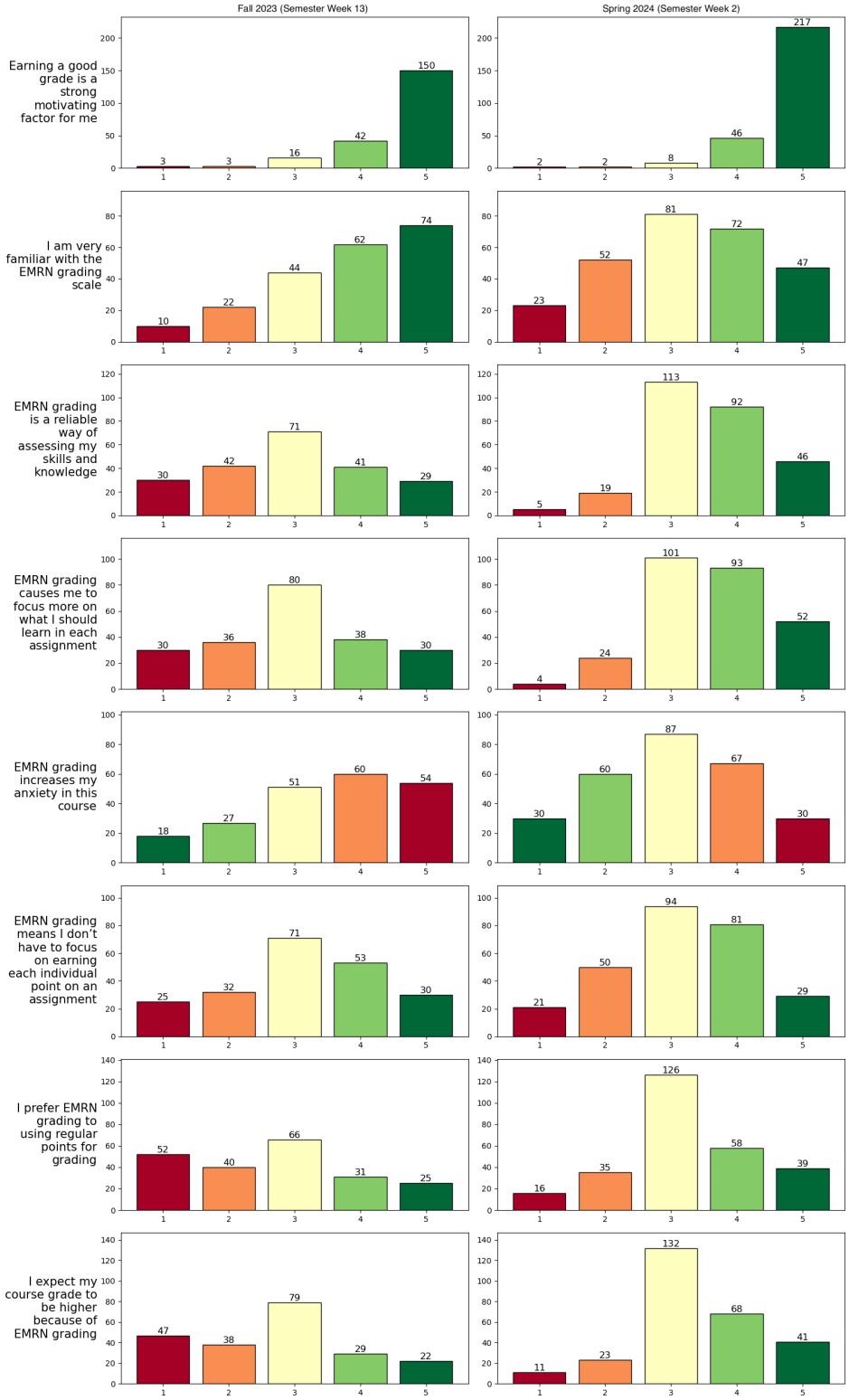


Figure 4.2: EMRN grading scheme subscale - per question summations

Question	U-value	p-value
Earning a good grade is a strong motivating factor for me	14.0	0.8335
I am very familiar with the EMRN grading scale	8.0	0.420635
EMRN grading is a reliable way of assessing my skills and knowledge	11.0	0.8413
EMRN grading causes me to focus more on what I should learn in each assignment	11.0	0.8340
EMRN grading increases my anxiety in this course	6.5	0.2477
EMRN grading means I don't have to focus on earning each individual point on an assignment	11.0	0.8413
I prefer EMRN grading to using regular points for grading	12.0	1.0

Table 4.3: Mann-Whitney U-test results for the EMRN grading scheme subscale

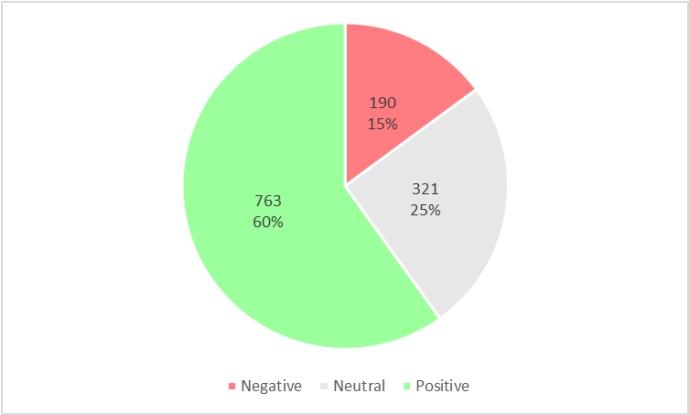
4.2.2 Late Policy Subscale

This subscale provides us with information to answer RQ2: "Gather information about student perceptions of the employed *late policy*" and its impact on their executive function and time management skills.

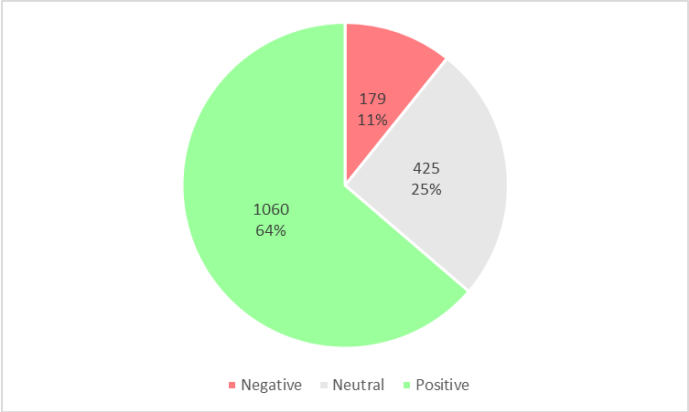
The analysis for this subscale is similar to that of the EMRN grading scheme described in Section 4.2.1. The method used for analyzing this subscale differs from the MSLQ subscales, which are discussed further in Section 4.2.3. Note that this subscale was a 5-point Likert scale with the response 1 being "strongly disagree" and 5 being "strongly agree".

Firstly, we classified the Likert scale responses 1, 2 as negative, 3 as neutral, and 4, 5 as positive responses. Then we summed all the positive, negative, and neutral responses. The results are shown in Figure 4.3.

For Fall 2023, as shown in Figure 4.3a, about a quarter of students felt neutral about the system, with 25% of the students responding 3 on the Likert scale. Students felt mostly and rather positively about the late policy with about 60% marking either 4 or 5, and a smaller percentage of students felt negatively about the policy with about 25% marking



(a) Fall 2023 (Semester Week 13)



(b) Spring 2024 (Semester Week 2)

Figure 4.3: Late Policy subscale - sentiment analysis

either 1 or 2.

For Spring 2024, as shown in Figure 4.3b, again about a quarter of students felt neutral about the system, with 25% of the students responding 3 on the Likert scale. Also, yet again the students felt mostly and rather positively about the late policy with about 64% (about two-thirds) marking either 4 or 5, and a lesser amount of students felt negatively about the policy with about 11% (less than one-eight) marking either 1 or 2.

The students from both semesters did not have a very different perspective on the Late policy. Also, the late policy is a positively rated aspect of the suite of EGPs. This too is discussed further in Section 5.2.

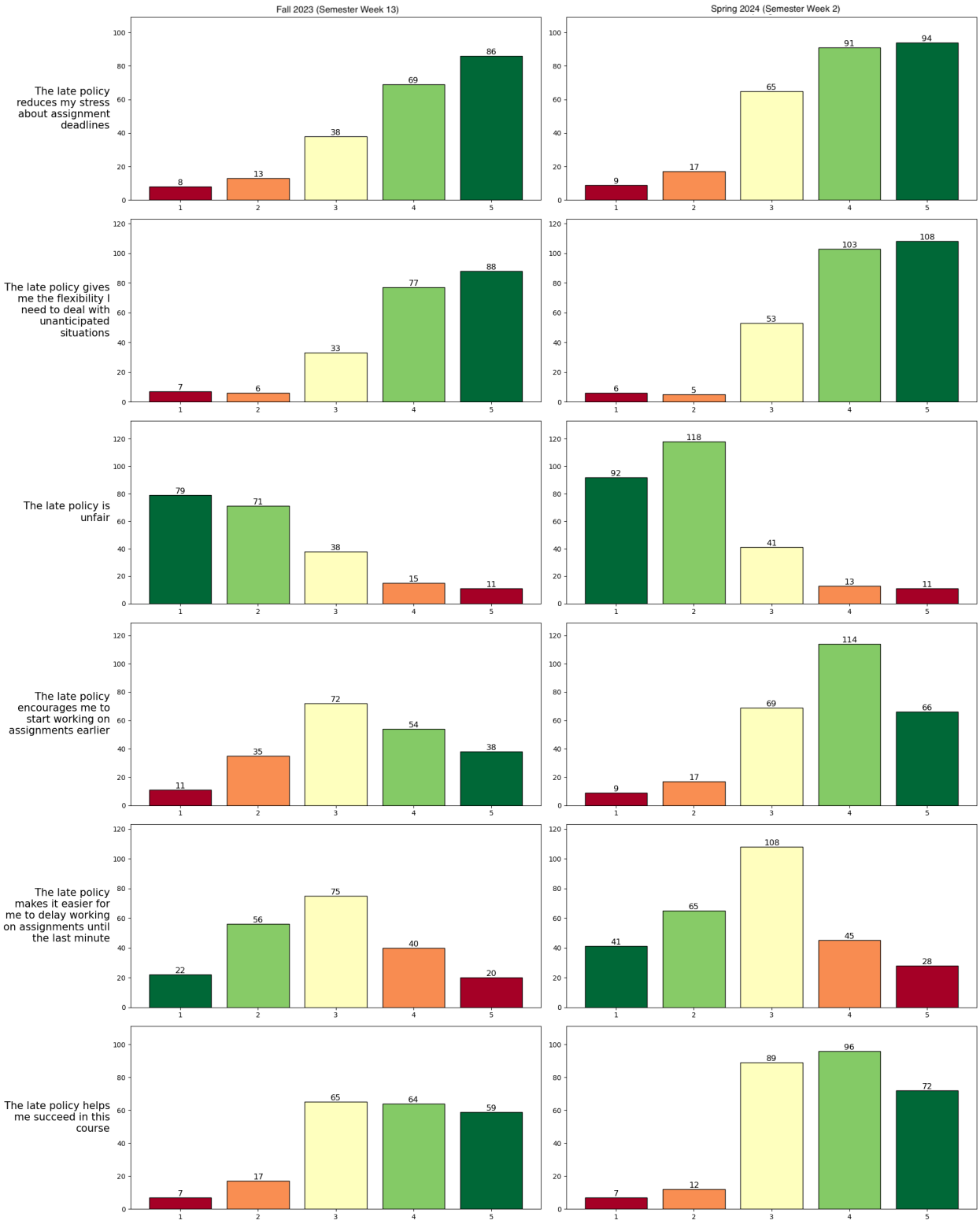


Figure 4.4: Late Policy subscale - per question summations

Thereafter, we analyzed each question individually. For each question, we plotted the distribution of student responses (from 1 to 5) by summing them up. These questions are described in detail in Table 3.4. The result of this analysis is shown in Figure 4.4, and has 16 bar graphs; 8 graphs per semester; 2 graphs per question.

Looking at Figure 4.4, we see that only the following questions displayed a neutral distribution on the Likert scale: "The late policy encourages me to start working on assignments earlier" (Fall 2023), "The late policy makes it easier for me to delay working on assignments until the last minute" (Fall 2023 and Spring 2024), "The late policy helps me succeed in this course" (Fall 2023). All the other questions display a more positively weighted distribution on the Likert scale.

The results of the Mann-Whitney U-test [23] across both semesters are shown in Table 4.4. Here too, as in the EMRN grading scheme subscale, we notice that there is no significant difference between the two semesters for any of the questions, as the p-value for all the tests is >0.05.

Question	**U-value**	**p-value**
The late policy gives me the flexibility I need to deal with unanticipated situations	11.5	0.9166
The late policy is unfair	10.5	0.7533
The late policy encourages me to start working on assignments earlier	11.0	0.8413
The late policy makes it easier for me to delay working on assignments until the last minute	8.0	0.4206
The late policy helps me succeed in this course	8.5	0.4633

Table 4.4: Mann-Whitney U-test results for the Late Policy subscale

4.2.3 MSLQ Subscales

These subscales provide us evidence to answer RQ4: "How does students' self-regulated learning beliefs relate to their perceptions of EGPs". In our analysis, we also include the EMRN grading scheme and late policy subscales, specifically in Table 4.5, where we discuss the measures of central tendencies of these two scales. An important caveat to mention here is the use of ANOVA to analyze the relationship between the MSLQ subscales and the subscales of EMRN grading scheme and late policy, which are not validated, raises concerns about the validity of the results.

We also compare the relative distributions of these subscales across semesters, which has the following limitations:

- Possible demographic changes between the Fall and Spring semesters could confound the results.
- Comparing pre and post-survey results from different populations introduces a potential threat to the validity of the study. Differences in these populations may limit the ability to draw meaningful conclusions.

As mentioned previously in Section 3.2.1, the MSLQ subscales include the following:

1. Critical Thinking
2. Help Seeking
3. Metacognitive Self-Regulation
4. Peer Learning
5. Self-efficacy for Learning and Performance

6. Test Anxiety

7. Intrinsic Goal Orientation

There is a scoring mechanism for each subscale. Students provide their response on a seven point Likert scale, which ranges between "not at all true of me" and "very true of me." Then, a subscale's score is calculated by normalizing the items that make up that subscale on a 100-point scale. For example, critical thinking has five questions within it. An individual's score for critical thinking would be computed by:

1. Summing the response (ranging from 1 to 7) from each question.

2. Ignoring blank responses.

3. Now, we have the sum of five responses, which ranges from 5 to 35 (assuming there were no blank responses).

4. Normalizing the 5-35 range over a 0-100 point scale.

The subscales in the MSLQ and their corresponding questions have been validated. This means that when all the questions from a subscale are added up, this sum represents a student's level on that subscale. For example, if after normalizing John Doe's responses on the Peer-Learning subscale, it turns out that the final sum is 90%, this can be taken to mean that they rate themselves as a highly collaborative student.

In Table 4.5 we compare the means of the various subscales across the two semesters along with comparing it to the baseline means obtained from the original MSLQ study.

As seen in Table 4.5, most subscales have both their mean and median between 60% and 70%. Notable exceptions to this are the Peer-Learning (Spring 2024), Self-efficacy (Fall

2023), Intrinsic Goal Motivation (Fall 2023) with lower mean and median than the others, and Intrinsic Goal Motivation (Spring 2024) with a higher mean and median than the others.

In this table, we compare the results across the two semesters. By doing so, we aim to find out whether agreements or disagreements exist between the results of the two semesters. If they do, this gives us some reassurance that both semesters point to the same results. If they disagree, however, then we can conduct further analysis to uncover any latent variables or causes. However, there are limitations around this. While conducting this study, limitations were noted regarding the comparison of pre-semester (Spring 2024) and post-semester (Fall 2023) differences. The data available did not allow for a direct comparison between semesters, which would have provided a more comprehensive analysis. Additionally, there are population differences to consider, including seasonal variations between student populations, some students repeating the course, and variations in the number of students taking the course each semester. Despite these limitations, the comparisons were conducted for completeness in the analysis. Overall, while some differences were observed, they were not substantial enough.

This table also lists the means obtained by the original authors of the MSLQ. We include this list here for comparison. The MSLQ was employed on a different demographic by the authors specifically, "The data presented in this document were gathered from a sample of 380 Midwestern college students. Most of these students (N=356) attended a public, four-year university; the remaining students (N=24) attended a community college. This version of the MSLQ was administered towards the end of the Winter 1990 (January to May) semester. [30]" In the study, we include baseline MSLQ means for various subscales in the table to provide a reference point for our results. However, it is important to note that a direct comparison between our means and those from the MSLQ study cannot be made.

The MSLQ study was conducted at a different time period, on a different course, with a different number of students, and at a different university. Additionally, there are many other differences between the populations in the two studies. Therefore, while the baseline MSLQ means serve as a useful reference, they should not be interpreted as directly comparable due to these substantial differences. We see that the means from our study of the subscales: Help-Seeking, Peer Learning, Self-efficacy for Learning and Performance, Test Anxiety, and Intrinsic Goal Motivation differ *significantly* from the means of the MSLQ study for at least one semester.

Table 4.5: MSLQ subscales: measures of central tendency and their differences

Subscale	Fall 2023		Spring 2024		MSLQ Baseline	Cohen's d	Kruskal-Wallis test	
	Mean	Median	Mean	Median	Means	Metric	F-value	p-value
Critical Thinking	61.4%	60.0%	61.1%	60.0%	59.4%	0.01	0.01	0.91
Help-seeking	49.9%	50.0%	44.2%	50.0%	54.9%	0.28	8.70	**<0.01***
Metacognitive Self-Regulation	62.6%	61.7%	67.2%	66.7%	64.9%	0.25	9.44	**<0.01***
Peer Learning	61.4%	66.7%	45.1%	50.0%	41.3%	0.65	55.12	**0.00***
Self-efficacy for Learning and Performance	60.2%	59.5%	68.5%	69.0%	78.0%	0.42	24.31	**<0.01***
Test Anxiety	60.7%	60.0%	60.6%	60.0%	51.9%	<0.01	0.02	0.88
Intrinsic Goal Motivation	62.0%	58.3%	71.7%	75.0%	71.9%	0.48	32.75	**0.00***
EMRN Grading Scheme	57.7%	56.3%	62.4%	62.5%	-	0.26	13.17	**<0.01***
Late Policy	59.0%	58.3%	60.0%	58.3%	-	0.06	0.29	0.59

After confirming that none of the subscales follow a normal distribution using the Shapiro-Wilk test [36], we applied the Kruskal-Wallis test [21] to the subscales across two semesters. This was done to determine if there were statistically significant differences between the two semesters for the subscales. As seen in Table 4.5, all subscales, except for Critical Thinking and Test Anxiety, show statistically significant differences between their distributions across the two semesters.

To understand these differences, we compared the means of these subscales across the two

semesters. This mean difference is contextualized in terms of the Cohen's d metric, as shown in the last column of Table 4.5. We observed that none of the subscales exhibit a substantial difference. However, the use of Cohen's D metric, despite the subscales not being normally distributed, raises concerns about the appropriateness of the statistical analysis. Non-parametric methods may be more suitable for comparing metrics in this context.

Next, we look at figure 4.5, which summarizes the normalized scores for all subscales across the two semesters. It is important to note that the student populations, as well as many other variables, differ between the two semesters. Despite these differences, we compare them side-by-side since we use the same instrument for both semesters. A key distinction is that the Fall 2023 semester is a post-survey, while the Spring 2024 semester is a pre-survey.

1. **Critical Thinking**: For both semesters, the maximal frequency for the Critical Thinking subscale lies slightly above the halfway point. However, for Spring 2024, we observe slightly more positive responses. Additionally, very few students rate themselves poorly in terms of Critical Thinking skills.

2. **Help-Seeking**: The mean for this subscale is lower than the halfway point. This is more pronounced for the Spring 2024 semester, suggesting that students in this semester rated themselves as less likely to seek help compared to the Fall 2023 semester.

3. **Metacognitive Self-Regulation**: This subscale to follows a similar pattern as with Critical Thinking.

4. **Peer-Learning**: There is a disparity with this subscale across the two semesters. We see that the students from Spring 2024 semester do not utilize their peers to aid in their learning as much as students from the Fall 2023 semester. This subscale is one of the more evenly distributed subscales, particularly with respect to negative

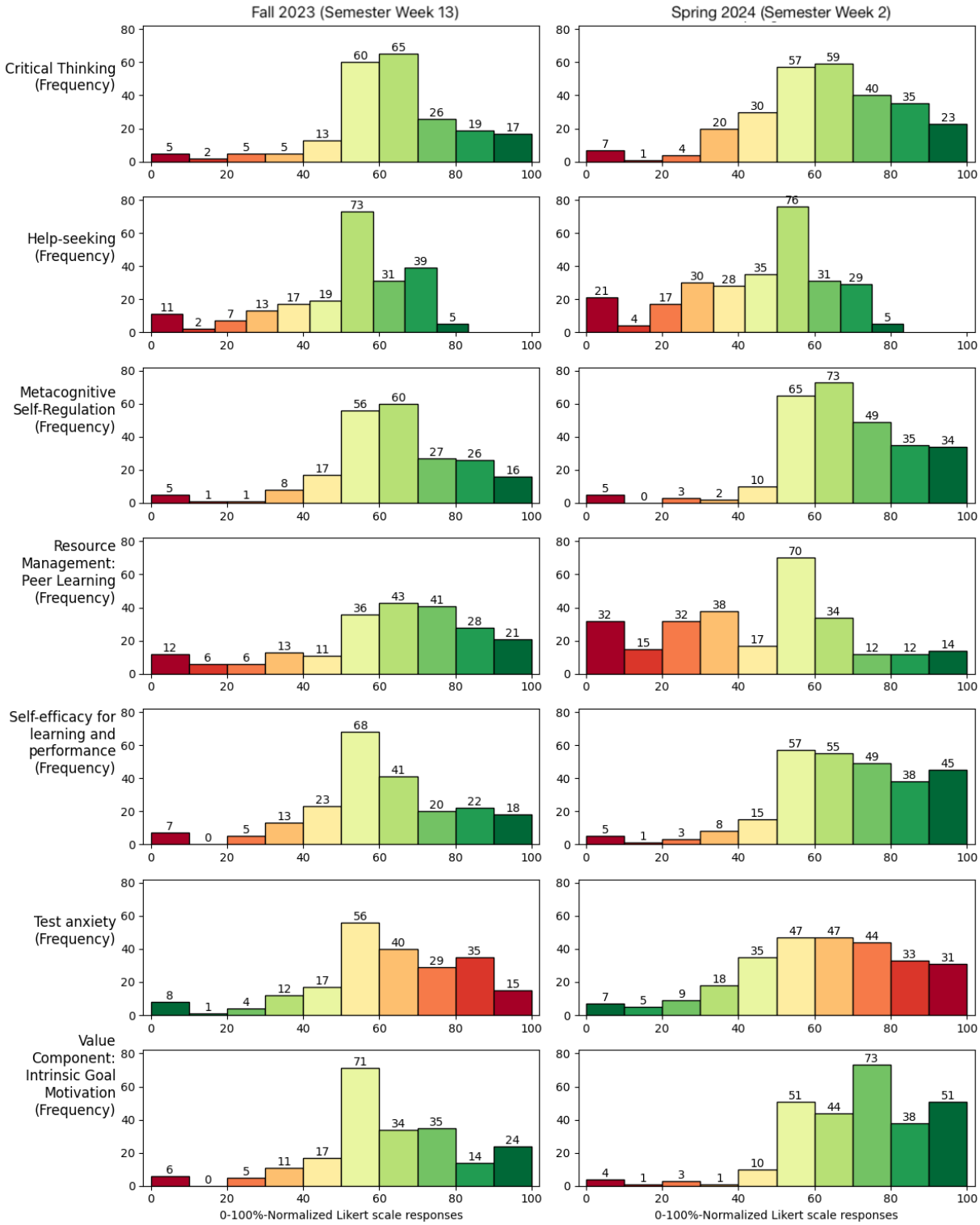


Figure 4.5: MSLQ subscales - 100-point normalized histograms (For e.g. A higher frequency towards 100 corresponds to a higher level of Critical Thinking)

responses. The highest frequency for the 10th percentile on this subscale, with 32 students. So too, is the case with Fall 2023 with 12 students being in the 10th percentile for this subscale, the highest among all others.

5. **Self-efficacy for learning and performance**: Here too, a disparity exists between the two semesters. We see that while the Fall 2023 semester has a more centrally peaking distribution, the Spring 2024 semester is skewed more towards the latter end of the 100-point scale.

6. **Test anxiety**: A similar distribution exists here as with the Self-efficacy subscale right above. We see here that students from the Spring 2024 semester indicate that they experience a higher degree of anxiety as opposed to the students from Fall 2023.

7. **Intrinsic Goal Motivation**: This subscale, like the two above, follows a similar distribution, with the Fall 2023 semester being more centrally distributed and the Spring 2024 being slightly skewed towards the latter end of the 100-point scale.

We next present the results of statistical tests to find relationships between the MSLQ subscales and the EMRN grading scheme, late policy, and Test Anxiety.

1. **EMRN Grading Scheme**

   With the EMRN grading scheme subscale, we find that the other MSLQ subscales have a statistically significant relation with it as seen in Table 4.6 that the analysis of variance p-value is < 0.01. Additionally, when fitting an estimation model by using the the least squares methods, the parameter values of Metacognitive Self-Regulation, Peer Learning, and Self-efficacy for learning and performance show statistically significant results (their p-value is < 0.05) with the EMRN grading scheme

subscale, as shown in Table 4.7. These subscales are positively related to the EMRN grading scheme subscale.

Source	DF	Sum of Squares	Mean Square
Model	7	2.67	0.38
Error	472	9.54	0.02
C. Total	479	12.21	-

F Ratio	Prob > F
18.90	< 0.01

Table 4.6: Analysis of variance - EMRN subscale vs MSLQ subscales

Term	Estimate	Std Error	t Ratio	Prob>|t|
Metacognitive Self-Regulation	0.29	0.07	3.90	< 0.01
Resource Management: Peer Learning	0.13	0.04	3.49	< 0.01
Self-efficacy for learning and performance	0.14	0.05	2.76	< 0.01
Help-seeking	-0.10	0.05	-1.90	0.06
Value Component: Intrinsic Goal Motivation	0.06	0.06	1.13	0.26
Critical Thinking	-0.04	0.06	-0.77	0.44

Table 4.7: EMRN grading scheme - sorted parameter estimates for fit least squares model

2. **Late Policy**

   With the late policy subscale, like the EMRN grading scheme subscale, we find that the other MSLQ subscales have a statistically significant relation with it as seen in table 4.8 that the analysis of variance p-value is < 0.01. Additionally, when fitting an estimation model by using the the least squares methods, the parameter values of Metacognitive Self-Regulation and Peer Learning show statistically significant results (their p-value is < 0.05) with the late policy subscale as shown in Table 4.9. These subscales are positively related to the late policy subscale.

Source	DF	Sum of Squares	Mean Square
Model	7	1.74	0.25
Error	472	8.87	0.02
C. Total	479	10.61	-

F Ratio	Prob > F
13.24	< 0.01

Table 4.8: Analysis of variance - Late Policy vs MSLQ subscales

Term	Estimate	Std Error	t Ratio	Prob>|t|
Metacognitive Self-Regulation	0.32	0.07	4.51	< .01
Resource Management: Peer Learning	0.14	0.04	3.68	< .01
Help-seeking	-0.09	0.05	-1.77	0.08
Critical Thinking	0.05	0.05	0.90	0.37
Self-efficacy for learning and performance	-0.03	0.05	-0.52	0.60
Intrinsic Goal Motivation	-0.03	0.05	-0.52	0.61

Table 4.9: Late Policy - sorted parameter estimates for fit least squares model

3. **Test Anxiety**

With the Test Anxiety subscale, like the two subscales discussed previously, we find that the other MSLQ subscales have a statistically significant relation with it as seen in table 4.10 that the analysis of variance p-value is < 0.01. Additionally, when trying to fit an estimation model by using the the least squares methods, the parameter values of Self-Efficacy, Self-regulation, Help-seeking and Intrinsic Goal Motivation show statistically significant results (their p-value is < 0.05) with the Test Anxiety subscale as shown in table 4.11. While Self-Efficacy is negatively related to Test Anxiety subscale (meaning that a student who rated themselves as being more self-efficient experienced lesser test anxiety), Self-Regulation, Help-Seeking and Intrinsic

Goal Motivation are positively related to the Test Anxiety subscale.

Source	DF	Sum of Squares	Mean Square
Model	7	3.91	0.56
Error	472	15.19	0.03
C. Total	479	19.10	-

F Ratio	Prob > F
17.3651	< 0.01

Table 4.10: Analysis of variance - test anxiety subscale vs MSLQ subscales

Term	Estimate	Std Error	t Ratio	Prob>|t|
Self-efficacy for learning and performance	-0.35	0.06	-5.35	< 0.01
Metacognitive Self-Regulation	0.39	0.09	4.19	< 0.01
Help-seeking	0.17	0.06	2.66	< 0.01
Value Component: Intrinsic Goal Motivation	0.17	0.07	2.43	0.02
Critical Thinking	0.09	0.07	1.24	0.22
Resource Management: Peer Learning	0.01	0.05	0.23	0.81

Table 4.11: Test Anxiety - sorted parameter estimates for fit least squares model

We next describe the multivariate statistics between various subscales. In Table 4.12, each cell contains multivariate non-parametric Spearman's ρ values. This table is represented in the form of a sparse matrix. The subscale names have been condensed here for brevity. We only consider values statistically significant if they are above 0.6. Additionally, all the values in the table barring the one between Intrinsic Goal Motivation and Peer Learning have a p-value < 0.05. We see that the following subscales have statistically significant correlation scores:

- Self-Regulation and Critical Thinking

- Peer Learning and Help-Seeking

- Self-efficacy and Self-Regulation

- Intrinsic Goal-Motivation and Self-Regulation

- Intrinsic Goal-Motivation and Self-efficacy

Table 4.12: Multivariate non-parametric Spearman's ρ values

Subscales	Critical Thinking	Help-seeking	Self Regula--tion	Peer Learning	Self-efficacy	Test anxiety	Intrinsic Goal Motivation	EMRN	Late policy
Critical Thinking	-	-	-	-	-	-	-	-	-
Help-seeking	0.44	-	-	-	-	-	-	-	-
Self-Regulation	**0.67**	0.53	-	-	-	-	-	-	-
Peer Learning	0.40	**0.65**	0.32	-	-	-	-	-	-
Self-efficacy	0.47	0.30	**0.64**	0.14	-	-	-	-	-
Test anxiety	0.33	0.36	0.39	0.25	0.16	-	-	-	-
Intrinsic Goal Motivation	0.56	0.30	**0.64**	0.08 (p-value > 0.05)	**0.67**	0.29	-	-	-
EMRN	0.26	0.17	0.39	0.16	0.35	0.19	0.34	-	-
Late policy	0.29	0.23	0.36	0.25	0.20	0.27	0.23	0.36	-

4.3 One-on-One Interviews and Free Response Questions

4.3.1 Grounded Theory Analysis Themes

As mentioned in Section 3.2.7, we used grounded theory analysis to code the interviews conducted with students. The interview was conducted around perspectives of students regarding topics such as the EMRN grading scheme, examining various dimensions such as personal history, experiences with grading systems, late policies, and the dichotomy between smaller quizzes and larger exams. Through comprehensive hour-long interviews, students shared their insights, providing rich data for analysis.

After conducting the interviews, each interview session was transcribed, and notes were compiled to capture the essence of student responses. Grounded theory principles guided the analysis process, beginning with open coding to identify individual points raised by students. These points were then systematically organized into codes, allowing for the extraction of overarching themes. Axial coding facilitated the exploration of relationships between codes, while selective coding honed in on central concepts. Theory saturation was achieved when no new insights emerged from the data, indicating a comprehensive understanding of the phenomena under investigation.

We must mention a few caveats to these results, which are also discussed in Section 5.5. Firstly, conducting interviews towards the beginning and towards the end of the course could introduce biases in the responses. The timing of the interviews may influence participants' recollection and perception of their experiences, potentially skewing the data. Secondly, the limited number of interviews conducted may restrict the generalizability of the findings.

The following themes emerged out of the grounded theory analysis:

1. EMRN grading scheme
2. Grading perception
3. Late policy
4. Learning and grading
5. Quiz dropping
6. Quiz-vs-exams
7. Teacher perception
8. Webcat

4.3.2 Coding Results

This section provides an in-depth analysis of the codes obtained. We only discuss codes here that were mentioned by at least two students. In this context, the frequency represents the number of times a statement was made that falls under a particular code. Table 4.13 presents an overview of the various codes, their observed frequencies, and the number of students who mentioned them.

1. **"emrn-not-granular"**: Frequency=18, Students=6

   A significant code that emerged from the grounded theory analysis is "emrn-not-granular." This code reflects students' observations regarding the EMRN grading system's lack of granularity, particularly its tendency to compress a wide range of

Table 4.13: Summary of grounded theory analysis codes and frequencies

Codes	Frequency	Students
emrn-not-granular	18	6
emrn-system-achieves-learning-outcomes, EMRN-makes-learning-outcomes-clear, EMRN-represents-skill-well, CS1114-assessments-induce-learning	12	4
EMRN-represents-skill-poorly, emrn-doesnt-reward-skill-equally	7	6
emrn-increases-effort-toward-perfection, emrn-pushed-me-to-get-better-grade, emrn-scoring-100-diff-and-time-consuming	10	2
highly-motivated-for-good-grades, start-out-course-with-goal-as-A-grade	7	4
late-policy-used-sparingly, late-policy-could-promote-tardiness, late-policy-has-opportunity-cost	20	4
late-policy-increased-quality-reduced-stress, late-policy-helped-with-other-courses, late-policy-helped-manage-time	19	4
quizzes-preferable-to-exams	38	6

skill levels into a single grade category. Students noted that this lack of granularity often results in the system pushing them toward a B- for an above-average level performance, regardless of their actual skill level. This compression of diverse skill levels into a single grade category can lead to a perception of harshness among students. Even minor mistakes or discrepancies may disproportionately impact their final grade, potentially resulting in a downward shift of an entire letter grade. Thus, while the EMRN grading system aims to provide clear standards for mastery-based assessment, its limited granularity can pose challenges for students in accurately reflecting their skill levels and perceptions of fairness in grading outcomes.

2. **"emrn-system-achieves-learning-outcomes", "EMRN-makes-learning-outcomes-clear ", "EMRN-represents-skill-well", "CS1114-assessments-induce-learning"**:

Frequency=13, Students=3

Firstly, "EMRN explains learning outcomes well" indicates that students perceive the EMRN grading system as effectively articulating and elucidating the expected learning outcomes for a given course or assignment. This suggests that the grading criteria within the EMRN framework are transparent and closely tied to the educational objectives. Additionally, the code "EMRN relates to learning outcomes" underscores the perceived alignment between the EMRN grading system and the intended learning outcomes of the course. Moreover, the code "EMRN maps to skill levels well" highlights students' belief that the EMRN grading system effectively assesses and reflects their skill levels and competencies. Lastly, the code "EMRN promotes learning" suggests that students perceive the EMRN grading system as conducive to their academic growth and development. This indicates that students view the grading criteria within the EMRN framework as supportive of their learning process, motivating them to engage actively with course material and strive for improvement.

3. **"emrn-was-confusing", "assessment-instructions-unclear-undetailed" and "emrn-doesnt-relate-to-learning-outcomes"**: Frequency=12, Students=4

   Firstly, "EMRN can be confusing to some students because of unfamiliarity" highlights the challenges some students face due to the novelty of the EMRN system, suggesting a need for clearer explanation and guidance. Additionally, "EMRN requires effective and clear communication because of low granularity" underscores the importance of transparent communication in navigating the EMRN system, especially given its limited granularity. This code emphasizes the necessity for educators to articulate grading criteria and expectations clearly to mitigate confusion among students. Moreover, "emrn-doesnt-relate-to-learning-outcomes" suggests a

disconnect between the EMRN grading system and the intended learning outcomes, indicating a potential mismatch between assessment methods and educational objectives. Meanwhile, codes such as "assessment-instructions-unclear-undetailed" and "assessment-instructions-not-clear-at-start" highlight concerns regarding the clarity and specificity of assessment instructions. This suggests a need for clearer, more detailed guidance to ensure that students understand the expectations and requirements of assessments from the outset. Lastly, the code "emrn-somewhat-achieves-learning-outcomes" suggests that while the EMRN system may partially align with learning objectives, there may be room for improvement in fully realizing these outcomes through the grading approach.

4. **"EMRN-represents-skill-poorly", "emrn-doesnt-reward-skill-equally"**: Frequency=7, Students=6
   One such code, "Some students thought that traditional grading related to learning outcomes better," indicates a perception among certain students that traditional grading methods may be more closely linked to the achievement of learning objectives. This suggests a potential preference for conventional grading approaches over alternative methods like EMRN. Moreover, the code "EMRN-represents-skill-poorly" highlights a perceived ambiguity or inconsistency in the interpretation of grades, particularly within the context of mastery-based grading systems. This suggests that students may not always equate a grade of "E" with actual mastery of the subject matter, indicating potential discrepancies between grading criteria and perceived levels of achievement. Furthermore, the code "Two students could put in hugely different amounts of effort but could receive the same grade" underscores a common concern among students regarding the fairness and equity of grading practices. This observation suggests that grading systems may not always accurately reflect differ-

ences in students' levels of effort or engagement, potentially leading to perceptions of injustice or inequality in academic assessment.

5. **"emrn-increases-effort-torward-perfection", "emrn-pushed-me-to-get-better-grade", "emrn-scoring-100-diff-and-time-consuming"**: Frequency=10, Students=2

   These codes highlight a perceived tendency of the EMRN grading system to escalate the effort curve towards perfection. This suggests that the grading criteria within the EMRN framework may incentivize students to strive for perfection in their academic endeavors to achieve higher grades. However, depending on individual student characteristics and coping mechanisms, this heightened emphasis on perfectionism may have contrasting relationships. Some students may feel motivated to pursue perfection, potentially leading to increased stress or general anxiety in the course (related to the grading scheme), while others may opt to settle for more moderate achievements, such as scoring an 80 or 50, to mitigate the pressure associated with pursuing perfection. Moreover, the code suggests that the difficulty in achieving perfection within the EMRN system may contribute to a perception of harshness among students. Even minor errors or deviations from perfection may result in a downgrade from a higher grade category to an "E" (Exceeds Expectations). This underscores the potential challenges and pressures associated with striving for perfection within the EMRN grading framework, as students may perceive the grading criteria as stringent and unforgiving.

6. **"highly-motivated-for-good-grades", "start-out-course-with-goal-as-A-grade"**: Frequency=7, Students=4

   Two noteworthy codes derived from the grounded theory analysis are closely linked to the Motivated Strategies for Learning Questionnaire's (MSLQ) intrinsic goal mo-

tivation and self-efficacy subscales. Firstly, the code "highly-motivated-for-good-grades" suggests that students exhibit a strong intrinsic motivation driven by the desire to achieve good grades. This intrinsic motivation aligns with the MSLQ's emphasis on the importance of personal goals and aspirations in driving academic performance. Students who are highly motivated by the prospect of earning good grades may be more likely to engage actively in learning activities and persist in the face of challenges. Furthermore, the code "start-out-course-with-goal-as-A-grade" indicates that students often begin a course with the goal of attaining an "A" grade. This goal-setting behavior reflects a high level of self-efficacy, as students believe in their ability to achieve academic success and set ambitious targets accordingly. This aligns with the self-efficacy subscale of the MSLQ, which emphasizes individuals' beliefs in their capability to perform specific tasks and achieve desired outcomes. Students who set their initial course goal as earning an "A" grade may be more likely to exert effort and employ effective learning strategies to attain their desired academic outcome.

7. **"late-policy-used-sparingly", "late-policy-could-promote-tardiness", "late-policy-has-opportunity-cost"**: Frequency=20, Students=4
   These are some negative codes that emerged from the grounded theory analysis regarding the late policy. Firstly, the code "late-policy-used-sparingly" suggests that the late policy was not abused, indicating that students exercised discretion in its utilization. On the other hand, the code "late-policy-could-promote-tardiness" raises a potential concern regarding the unintended consequences of the late policy. This code suggests that the availability of a late policy may inadvertently incentivize or enable students to submit assignments past the designated deadlines without facing significant repercussions. Additionally, students recognize that utilizing the late pol-

icy for an assignment means that a lesser time is available for the next assignment. Consequently, they recognize that this leniency in enforcing deadlines could foster a culture of tardiness, potentially undermining the punctuality and accountability expected in academic settings.

8. **"late-policy-increased-quality-reduced-stress", "late-policy-helped-with-other-courses", "late-policy-helped-manage-time"**: Frequency=19, Students=4 The code "late-policy-increased-quality-reduced-stress-helped-with-other-courses-helped-manage-time" encapsulates several positive outcomes associated with the utilization of the late policy. This code was created by amalgamating several positive codes about the late policy. Firstly, it suggests that the availability of a late policy allowed students to allocate additional time and effort towards completing assignments, consequently enhancing the overall quality of their work. By providing students with the opportunity to refine and revise their submissions, the late policy may have contributed to improved academic outcomes and a deeper understanding of course material. Moreover, the code indicates that the late policy played a role in reducing stress levels among students. The flexibility afforded by the late policy likely alleviated the pressure associated with strict deadlines, allowing students to approach assignments with greater composure and focus. This reduction in stress may have positively impacted students' overall well-being and academic performance throughout the course. Furthermore, the late policy appears to have facilitated better management of students' time and responsibilities. By accommodating unforeseen circumstances or competing demands, such as those from other courses or personal commitments, the late policy enabled students to effectively balance their academic workload and prioritize their tasks accordingly. This adaptive approach to time management may have contributed to students' overall success and satisfaction

with the course.

9. **"quizzes-preferable-to-exams"**: Frequency=38, Students=6

   The Grounded Theory analysis revealed several codes pertaining to the employment of short quizzes to assess knowledge instead of exams within the course's equitable grading policies.

   "Quiz-require-short-performance-time" suggests that quizzes demand less time for completion compared to exams, making them more efficient assessment tools. This characteristic allows for more frequent assessment opportunities and reduces the cognitive load on students during evaluation periods.

   "Quiz-test-more-topics" underscores the ability of quizzes to cover a wider range of topics compared to exams, providing a more comprehensive assessment of students' understanding and retention of course material.

   Conversely, "exams-don't-fit-in-schedule" highlights logistical challenges associated with exam scheduling, which may create conflicts for students with busy schedules or conflicting commitments.

   "Quiz-don't-require-great-memorization" emphasizes that quizzes typically assess understanding and application rather than rote memorization, aligning with pedagogical approaches that prioritize critical thinking and problem-solving skills.

   "Small-quizzes-increase-my-score" suggests that the inclusion of small quizzes positively impacts students' overall grades by providing opportunities for consistent performance improvement throughout the course.

   "Quiz-test-aptitude-better-than-exam" indicates that quizzes may be more effective in assessing students' innate aptitude or understanding of concepts, as they require rapid recall and application of knowledge.

"Quiz-test-recent-knowledge" underscores the ability of quizzes to evaluate students' understanding of recent course material, encouraging regular review and reinforcement of key concepts.

"Exams-induce-greater-test-anxiety" highlights the psychological stress associated with high-stakes exams, which may negatively impact students' performance and well-being.

"Quiz-promote-growth-mindset" suggests that quizzes foster a growth mindset by emphasizing learning and improvement over fixed performance outcomes.

"Quiz-dropping-insurance" indicates that the option to drop quiz scores provides students with a safety net, reducing anxiety and encouraging risk-taking in their learning endeavors.

"Exams-reduce-scoring-increase-failure" suggests that exams may be more stringent in scoring and lead to increased rates of failure compared to quizzes.

"Small-quiz-better-than-exam" indicates a preference among students for smaller quizzes over exams, likely due to their perceived fairness, reduced anxiety, and ability to assess a broader range of skills and knowledge.

4.3.3 Sentiment Analysis

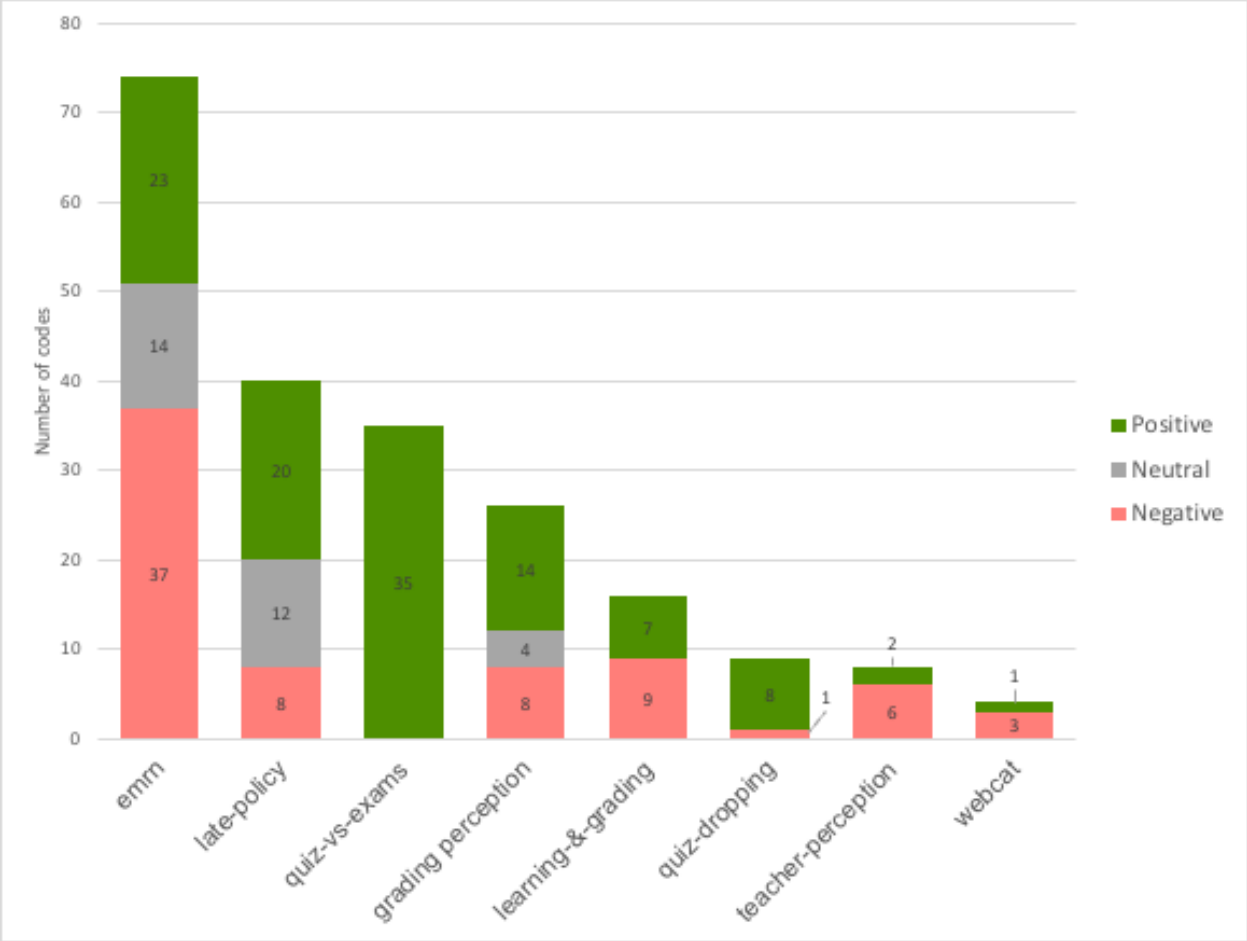


Figure 4.6: Sentiment analysis of the various themes observed in interviews' grounded theory analysis

As shown in figure 4.6, the grounded theory analysis of the interviews reveals varied sentiments across different thematic categories. The thematic sentiment analysis graph shows a distinct distribution of positive, neutral, and negative codes. Unlike Section 4.3.2, this section considers even low frequency codes while aggregating. In the category of "emrn", a predominantly negative sentiment is observed with 37 codes, compared to 14 neutral and 23 positive codes. This points to a challenging aspect in this area. Conversely, "grading perception" exhibits a balance between negative (14) and positive (20) sentiments, with

a smaller portion of neutral responses (12), indicating mixed feelings about grading practices.

The theme with the highest number of negative sentiments was "emrn", with 37 negative mentions, indicating significant concerns or challenges associated with it. In contrast, the quiz-vs-exams theme stood out with a remarkable positive sentiment, having 35 positive mentions and no negative or neutral sentiments, suggesting a strong preference or positive perception towards this aspect. Other themes such as late policy, and grading perception, also showed a balanced mix of sentiments, with late policy receiving 20 positive mentions against 8 negative ones, and grading perception having 14 positive mentions compared to 8 negative ones. Notably, the quiz dropping theme had an overwhelmingly positive response with 8 positive mentions and only 1 negative mention, indicating a favorable view towards this policy. The teacher perception theme, however, had more negative sentiments (6) compared to positive ones (2), pointing towards areas of potential improvement in teacher-student interactions or perceptions. Lastly, WebCAT, with 3 negative and 1 positive mentions, indicates a need for further investigation to understand the specific issues or positive aspects associated with it.

These findings offer insights into the participants' perceptions and sentiments regarding the equitable grading practices and general pedagogical topics. The positive sentiments observed in certain themes suggest areas of strength, while the negative sentiments highlight potential challenges or areas requiring attention and improvement.

4.3.4 Survey - Free Response Questions

In this section, we delve into the analysis of the open-ended responses gathered from the survey participants. The survey aimed to gather insights into the perceptions and expe-

riences regarding the grading policies implemented within the course. As mentioned in Section 3.2.3, participants were invited to provide their thoughts on the most beneficial aspect (Q5), the most anxiety-provoking aspect (Q6), and suggestions for improvement (Q7) concerning the grading policies.

The responses were analyzed using grounded theory. The analysis process began with categorizing responses into "Not Applicable" (N/A) and not N/A categories, both globally and per-question, the analysis proceeds with a thematic analysis utilizing grounded theory methodology. This approach allows for the identification of recurring patterns and themes within the dataset. Subsequently, sentiment analysis is conducted on the extracted themes to discern the overall sentiment expressed by participants towards each theme. This sentiment analysis provides valuable insights into the emotional tone associated with different aspects of the grading policies. Further, a per-theme analysis is undertaken, wherein subthemes are identified and analyzed within each overarching theme. This granular examination facilitates a deeper understanding of the nuances and variations within each thematic category, offering actionable insights for refining the grading policies to better align with the needs and preferences of the students.

Responses from students that were left empty or contained content irrelevant to the questions were categorized as Not Applicable. These are shown in table 4.14 and are discussed in the following couple of paragraphs.

In the Fall 2023 semester, Across all questions, the total number of responses is consistent at 345. However, there are variations in the number of responses categorized as not applicable, with Question 7 having the highest count (51, accounting for approximately 14.78% of total responses), followed by Question 6 (38, accounting for approximately 11.01% of total responses) and Question 5 (33, accounting for approximately 9.57% of total responses). This discrepancy in the prevalence of not applicable responses may indicate varying levels

Table 4.14: Free-response questions N/A analysis

Question	Fall 2023		Spring 2024	
	Not N/A	N/A	Not N/A	N/A
What was the most beneficial aspect of the grading policies in this course?	312	33	230	46
What was the most anxiety-provoking aspect of the grading policies in this course?	307	38	200	76
How can we improve the grading policies in this course?	294	51	162	114

of ambiguity or relevance perceived by respondents across the different questions. Further analysis could explore the reasons behind the higher occurrence of not applicable responses for Question 7 compared to the other questions, providing insights into areas where clarification or improvement may be needed in survey design or question formulation.

On the other hand, with the Spring 2024 semester, there are notable differences in the distribution of responses across the two categories N/A and not N/A. For instance, Question 7 exhibits the highest count of not applicable responses (114), indicating that approximately 41.38% of the total responses for this question were deemed not applicable by respondents. Conversely, Question 5 has the lowest count of not applicable responses (46), suggesting that approximately 16.67% of the total responses for this question fell into this category. These variations in the prevalence of not applicable responses could imply differences in the clarity or relevance of the questions within the survey. Further investigation into the reasons behind the higher occurrence of not applicable responses for certain questions may provide valuable insights for refining the survey instrument and optimizing data collection procedures.

The high number of N/A responses in questions 6 and 7 could potentially be attributed to the timing of the survey, likely conducted at the beginning of the semester. It is plausible that at this early stage, students may not have had sufficient exposure to the grading poli-

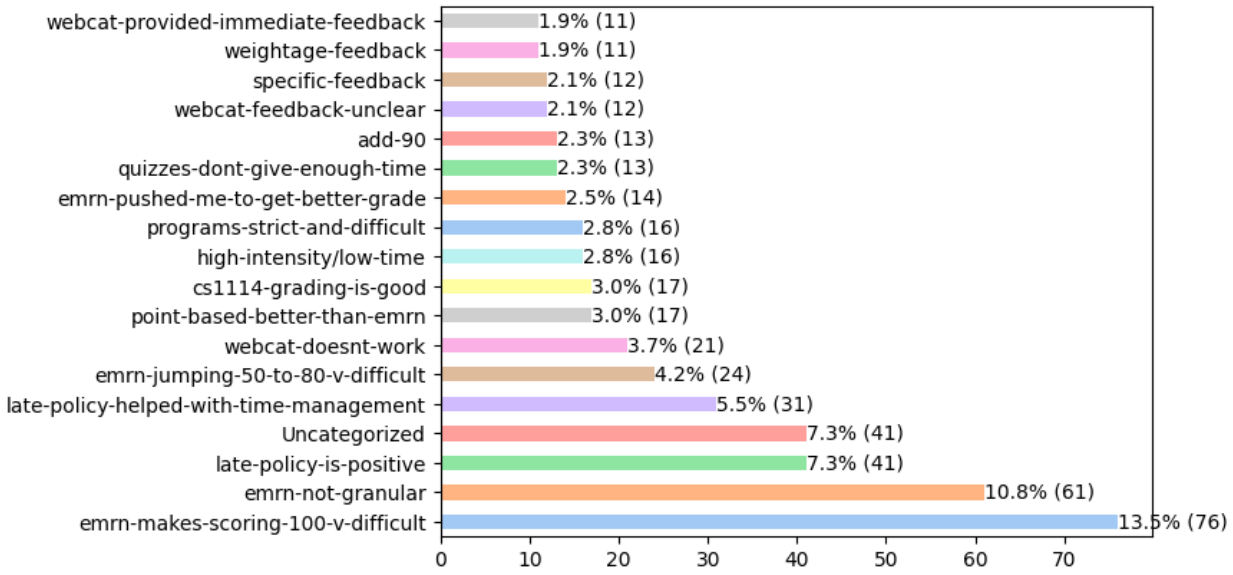


Figure 4.7: Free response questions - most frequent codes observed for Fall 2023 semester (Semester Week 13)

cies or experiences within the course to provide meaningful feedback on potential improvements. This hypothesis suggests that as the semester progresses, students may become more acquainted with the grading policies and their impact, leading to more informed responses.

4.3.5 Grounded Theory Code Analysis

In this section, we discuss the most frequently encountered codes resulting from the grounded theory analysis.

1. Fall 2023 semester

Figure 4.7 presents a bar chart analyzing responses categorized into various codes, with each code representing a distinct feedback point. The data suggests that the most common issue, with 13.5% (76 responses), is associated with the difficulty of scoring 100% in

the EMRN grading scheme. This is followed by the feedback that EMRN is not granular enough with 10.8% (61 responses). Next, a general "Uncategorized" code, which accounts for 7.3% (41 responses), indicating a diversity of other concerns not captured by specific codes. Additionally, a significant portion of respondents, 7.3% (41 responses), expressed that a late policy had positively impacted their time management. It is worth noting that the analysis only considered codes that received at least 10 responses, ensuring that only recurrent themes were highlighted. Responses that did not meet this threshold were discarded to maintain focus on the more prevalent issues.

2. Spring 2024 semester

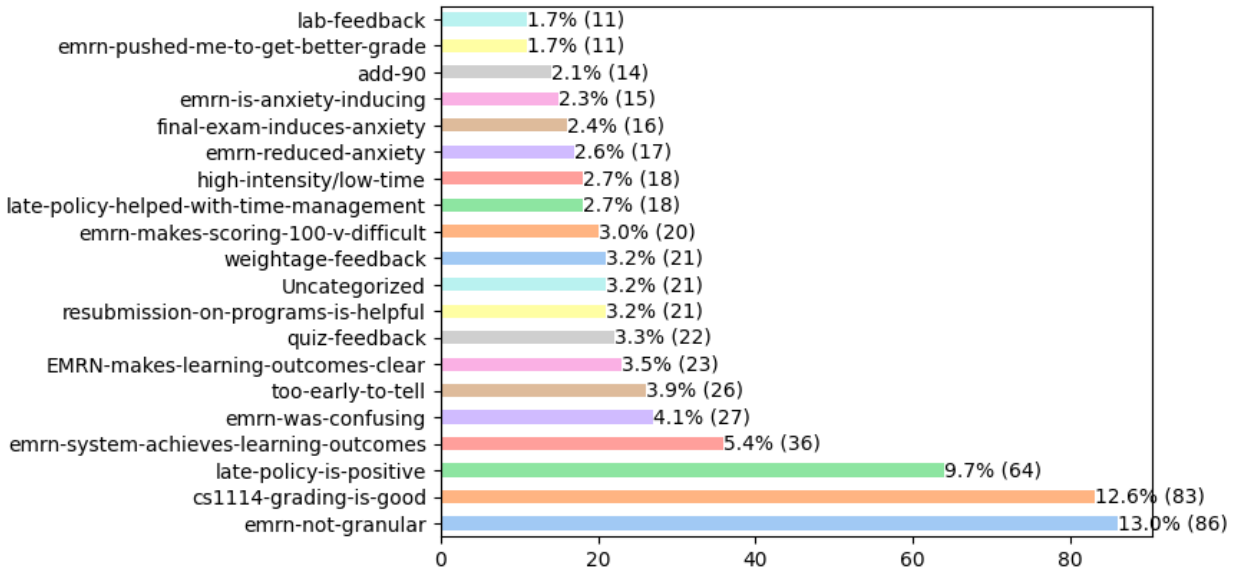


Figure 4.8: Free response questions - most frequent codes observed for Spring 2024 semester (Semester Week 2)

Figure 4.8 provides a breakdown of the feedback received for the Spring 2024 semester. Each bar represents a distinct code corresponding to a specific type of feedback. It is evident that the highest frequency of responses, amounting to 13.6% (86 responses), pertains to the "emrn-not-granular" code, indicating concerns regarding the granularity of the

EMRN grading scheme. Following this, 12.6% (83 responses) of the feedback suggests that "cs1114-grading-is-good", showing a positive reception to the grading system in the course. The "late-policy-is-positive" code also stands out, with 9.7% (64 responses), reflecting a favorable view of the late submission policy. Again, in conducting this analysis, only feedback codes that accumulated at least 10 responses were included, thereby focusing on the most prevalent issues. Feedback falling below this response threshold was excluded to ensure clarity and relevance in the results presented.

4.3.6 Sentiment Analysis

1. Fall 2023 semester

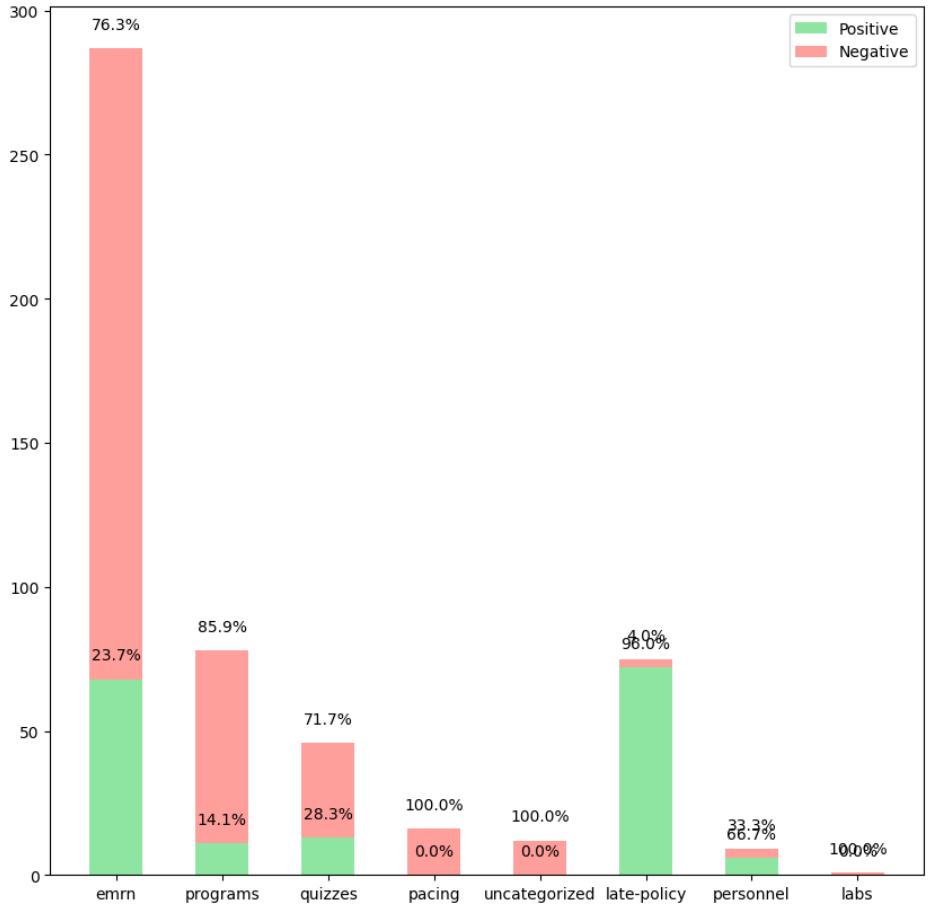


Figure 4.9: Survey free response questions Fall 2023 (Semester Week 13) - thematic sentiment analysis

The sentiment analysis for the Fall 2023 semester, as presented in the bar chart, shows the distribution of responses towards various thematic categories. Notably, the EMRN grading scheme theme displays a negative sentiment with 76.3% compared to the 23.7% of positive sentiment, indicating predominant dissatisfaction among the respondents in this area. As discussed in Section 4.3.5, a major cause of this is because of the complaint that EMRN makes scording an "E" very difficult and also that the scheme is not granular enough with its miniscule number of grades, which is 4.

Similarly, 'programs' reveal a majority of negative feedback at 85.9%, signifying a high

rate of dissatisfaction, although 14.1% of responses were positive. The sentiment regarding ‘quizzes‘ exhibits a substantial negative sentiment of 71.7%, with only 28.3% positive responses, suggesting room for improvement in this aspect.

A standout is the ‘late policy‘ theme which has a substantial 96% approval from students, with only 4% of the students responding negatively to it.

In stark contrast, the ‘personnel‘ category exhibits a significant negative sentiment at 66.7%, which is more than double the positive sentiment at 33.3%, highlighting concerns that may require immediate attention. On a positive note, ‘labs‘ receive unanimous accolades with 100% of sentiments being positive, showcasing a highly satisfactory experience in this area. The remaining categories contain a smaller number of responses to discuss.

2. Spring 2024 semester

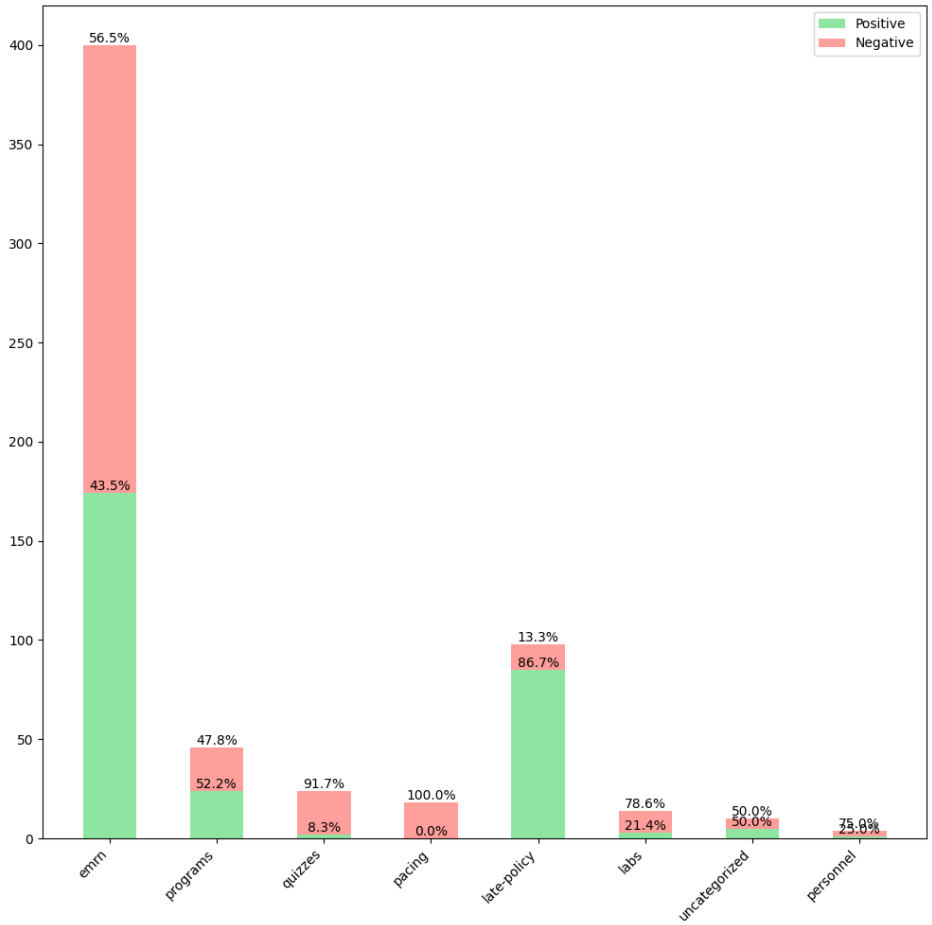


Figure 4.10: Survey free response questions Spring 2024 (Semester Week 2) - thematic sentiment analysis

The EMRN grading scheme theme presents a nearly balanced sentiment with a slight leaning towards negative at 56.5%, compared to 43.5% positive. In stark contrast, ‘programs‘ are viewed unfavorably, with a 91.7% negative sentiment overshadowing the 8.3% positive sentiment.

In the case of ‘pacing‘, there is an overwhelming positive sentiment, hitting the absolute of 100%, indicating a unanimous satisfaction within this theme. The ‘late policy‘ theme experiences a resounding positive tilt with 86.7% of the sentiments being positive. Again, the other categories consist of fewer responses to comment upon.

Chapter 5

Discussion

This chapter presents various insights into student perceptions, derived from the results section discussed earlier. It encompasses views on the EMRN grading scheme, late policy, Test Anxiety, and the relationships that the MSLQ subscales have with student perceptions, concluding with a discussion on threats to validity. It is crucial to note that our study concentrates on student perceptions of specific iterations of the diverse EGPs present in pedagogy. These include the EMRN grading scheme, a 7-day late policy extension with a maximum attainable grade, and the substitution of midterms with eight smaller quizzes.

It is important to contextualize the ensuing discussion within certain constraints. Throughout this chapter, we compare results across different semesters, comparing pre-semester (Spring 2024) and post-semester (Fall 2023) outcomes. The available data did not permit a direct comparison between semesters, which would have facilitated a more thorough analysis. The compared results pertain to different populations, which may exhibit seasonal variations, include students retaking the course, and differ in the number of students enrolled each semester.

5.1 EMRN Grading Scheme

5.1.1 Unfamiliarity and Acclimation

The EMRN grading scheme is one of the kinds of Equitable Grading Practices (EGPs) employed in CS-Educational research. It is a relatively recently developed approach and is mostly employed in courses where research is being conducted. Traditional grading practices are still prevalent in most classrooms.

This system was novel to a lot of students. In fact, 5 out of the 6 students who were interviewed mentioned that they had not been exposed to a specification based grading system in the past and had been graded using traditional grading in high school. The one student who did experience an alternative grading scheme mentioned that the scheme was weighted and was the closest thing to a specification grading system. Additionally, we observe codes such as "emrn-was-confusing", "assessment-instructions-unclear-undetailed" and "emrn-doesnt-relate-to-learning-outcomes" from the interviews.

When it comes to survey, the EMRN grading scheme subscale includes the question: "I am very familiar with the EMRN grading scale". The results of this question across the surveys are depicted in figure 4.2.

Two patterns can be observed from this graph. First, if we sum up the students who responded with 4 or 5 on the Likert scale, we get 62.7% for students from the Fall 2023 semester and 43.1% for students from the Spring 2024 semester. Conversely, students unfamiliar or somewhat familiar (those who marked 1, 2, 3 on the Likert scale) account for 37.3% for students from the Fall 2023 semester and 56.9% for students from the Spring 2024 semester. Thus, a significant number (at least a third) of students from both semesters reported being unfamiliar with the system. This is particularly relevant for the Fall 2023

semester's students since this was a post-survey and at the time of survey deployment they had spent a significant time being exposed to the grading system.

Second, we observe a difference in familiarity between the Fall 2023 and Spring 2024 semesters. Students from the Spring 2024 semester were less familiar with the grading system, with more than half of the students reporting unfamiliarity, compared to a third of the students from the Fall 2023 semester. This is relevant because, as mentioned earlier, the Fall 2023 survey is a post-survey, and the Spring 2024 survey is a pre-survey.

Successfully navigating this grading scheme requires not only an understanding of it, but also a shift in paradigm from the perspective of students. They need to focus on achieving learning outcomes rather than exhibiting behaviors that earn them grades. Although the EMRN grading scheme is simpler in terms of grade boundaries, it is more complicated in other aspects.

Firstly, apart from the course content, students are additionally required to understand the specifications within the grading scheme, in order to understand what they need to do get an "E" grade for example. This includes reading through the assessment instructions, understanding how these assessment tie in to each of the "E", "M", "R" and "N" grades, asking clarifying questions if they have any doubts before the assessment deadline itself and having discussions post the assessment to clarify any discrepancies or to improve performance. Compared to a traditional grading scale where for example, a program may be graded from 0 to 100% based on the number of test-cases their program can pass (this can be weighted), the EMRN grading scheme consists of additional steps from the student's end to familiarize themselves with the grading scheme.

Secondly, after receiving an EMRN grade within the assessment, the EMRN grading scheme maps grades from the EMRN notation onto a 100-point scale, which is discretized into tra-

ditional letter grades (A through F). The mapping process while transparent, adds a layer of complexity, potentially leading to confusion and frustration.

A student has to be able to reverse associate their final grade target onto the EMRN grading scale. For example, say a student intends to get a "B" grade at the minimum in this course. How many "M" or "R" grades could they afford to get away with to achieve this? These calculations of "affordable penalties" are also present in traditional grading schemes. However, since many students who took the course experienced this system for the first time, this mental calculation can prove to be challenging and as aforementioned, introduce a level of unintended complexity.

The unfamiliarity with the EMRN grading scheme has effects which are unintentional. These include an increased sense of confusion and anxiety, perceptions of harshness (because of a lack of understanding), and perceptions of unfairness. This can also lead to the student not being able to understand the relationship between the EMRN grading scheme and learning outcomes, as discussed in 5.1.3. To get past these issues, students require time, effective communication and support, and hence there exists a "period of acclimation" of the EMRN Grading Scale.

Thus, we can conclude that acclimation to a novel, alternative grading scheme such as the EMRN grading scheme employed within this course takes some time and different students adapt differently to it. This acclimation can also take more than a course where such a system is employed for the students to get used to it, i.e. more than a semester. Thus, it is important to facilitate this acclimation and ensure that is smooth and adequate support is provided to students. This novelty can also cause confusion and anxiety and also create difficulty for students to correlate it to learning outcomes. Moreover, clear and detailed communication can help alleviate some of these issues and help resolve this confusion. This involves detailed instructions around assessments in form of documentation and

also additional support from teaching assistants and instructors. Any future employers of EGPs, specifically specifications-based grading should be cognizant of the fact that acclimation to these novel practices takes time as well as cognitive effort on the part of the students in terms of gaining understanding as well as navigating stress and confusion.

5.1.2 Dysfunctional Expectations

When students transition from a traditional grading scheme to the EMRN scheme, they may carry over some expectations from the former. These expectations can become dysfunctional and reduce student engagement. For instance, students complained about the difficulty of scoring an "Exceeds Expectations" (E) grade or 100 points in this course on an assessment. They felt that to achieve this, they would have to do everything "perfectly". An exeplar quote is as follows:

> *'I personally find the EMRN to be completely ineffective. If I were to be at a 90% in a program, and it is due in less than an hour, I could spend that whole time making no progress. And due to the fact I either get an 80 or a 100, it makes no sense for me (as a student with other priorities) to even attempt getting the extra 10% on a chance I can succeed. If I do fail to get a 100, then I'm stuck with an 80% and a large amount of wasted time, along with some strong feelings of anger and regret. Obviously, I could have started earlier or gone to office hours, but I still shouldn't be punished for wanting to learn more instead of just doing the bare minimum. I hold very strong feelings about this because I have always been interested in computer science, however this class at times makes me consider not pursuing more computer science related courses.'*
> - A survey respondent from the Fall 2023 semester

This has two causes of dissatisfaction within it. The first one being that since everything has to be perfect, which is difficult to accomplish, the consequences of "not being perfect" is an 80 or a B/B- grade (assuming a student consecutively only scores "M" grades on assessments). The second cause is that some students that in the real world or within industry, a written program can never be perfect because of various reasons such as lack of resources and time, uncertainty, competing concerns, etc. As a result, expecting students to turn in a perfect program in order to score an 'E' grade for an assessment is also not a fair ask. It is important to mention that both these points are a cause of anxiety for students.

In relation to the first point, it can be argued that penalizing students to move down two or three letter grades (for example, A to B-) for the loss of less than five points on an assessment can be considered unfair. The introduction of a more granular scale could aid remedy this situation. However, with respect to the second point, the argument falls short as an assessment is a much more controlled environment (albeit not fully controlled), than a workplace setting.

The dysfunctional expectations include an important omission in the perception of students is the fact that many students that do not realize that scoring 100 points in a traditional 100-point scale graded assessments is much harder to accomplish than in an EMRN system. This is because unlike the traditional scale, the label of "Exceeds Expectations" is much more relaxed when compared to making zero mistakes and scoring 100 points under a traditional grading scheme. The EMRN scale is flexible in that it does allow minuscule errors, whereas a traditional scale may not. This is a subtle point and is easy to miss. Perhaps, making this information a part of the syllabus or announced by the instructor may reduce dissatisfaction toward the system.

The mapping of such specification grading schemes to traditional 100-point scale complicates this matter further. Anecdotally, instructors have mentioned difficulties in mapping

this scheme to a 100-point scale, since the overall pedagogical grading scheme in most universities happens to be a 100-point scale divided up into various letter grades. Using a 100-point scale probably contributes to dysfunction. Incorporating a specification grading scheme like the EMRN grading scheme into a 100-point scale in a manner that reduces this friction is part of our future work.

Another consequence of a simpler grading scheme are perceptions of inequity, particularly among high-performing students. Unlike traditional points-based grading systems that offer a finer differentiation of grades, the EMRN grading scheme's broader categories may not clearly distinguish between levels of performance. This lack of granularity can be perceived as inequitable by high-achieving students, as their exceptional performance may not be as recognized or rewarded compared to more traditional grading systems where there is significantly more partial grading. This perception could stem from a sense that the system fails to sufficiently acknowledge the efforts and accomplishments of high-performing students, potentially impacting their motivation and engagement. In fact, a number of students in the interviews alluded to this fact. Comments like my roommate barely works hard enough but we have the same grade were observed.

Another carried over expectation from students is that the grading scale must reward the effort put in by them, as opposed to skill. While there can be a positive correlation between effort and skill, achieving higher grades under the EMRN grading scheme is determined to a much greater degree by skill. Under a traditional grading system, students in the past may have been able to rely upon effort, rewarded behaviors, attendance, etc. to inflate their grades and overestimate their academic achievement, but these provisions do not exist under the employed EMRN grading scheme.

5.1.3 Better Achievement of Learning Outcomes

The EMRN grading scheme is meant to be a better academic medium over traditional grading practices by tying the learning outcomes with grades in a more effective manner as well as by eliminating the problems present in traditional grading practices. These include student appraisal for specific behaviors, instructor bias, grade-grubbing, etc. In this section, we aim to answer this question by evaluating the results from Chapter 4 in this direction. We come to the conclusion that the students' perspectives lean slightly to the positive on this scheme achieving academic outcomes better. Thereafter, we discuss the causes for this distribution.

The Likert scale survey results appear to be split for Fall 2023 semester and positive for the Spring 2024 semester. We report from figure 4.2 specifically the questions: "EMRN grading is a reliable way of assessing my skills and knowledge", "EMRN grading causes me to focus more on what I should learn in each assignment", "EMRN grading means I don't have to focus on earning each individual point on an assignment".

With the free response questions we see some student comments on the EMRN grading scheme achieving learning outcomes better. We see codes such as "cs1114-grading-is-good", "emrn-system-achieves-learning-outcomes", "EMRN-makes-learning-outcomes-clear", "emrn-pushed-me-to-get-a-better-grade", etc.

With the interviews, we notice codes such as "emrn-system-achieves-learning-outcomes", "EMRN-makes-learning-outcomes-clear", "EMRN-represents-skill-well", "CS1114-assessments-induce-learning" mentioned by at least 4 students.

These pieces of evidence point to the fact that a substantial amount of the student population does think that the EMRN grading scheme ties in well to achievement and explainability of learning outcomes. This can potentially increase student engagement with the

course material as this scheme facilitates better goal setting and planning, which in turn leads to more effective Performance and Self-Reflection phase as shown in figure 2.1.

However, almost a third of the student population does not feel this way. Causes of this could be unfamiliarity with the scale, which is discussed in Section 5.1.1, approaching this scale with the mental model of the traditional grading scale, which is discussed in Section 5.1.4. Approaching the EMRN grading scheme like a traditional grading scale can be problematic as students carry expectations from being judged using that scale into being judged by the EMRN grading scheme. Hence, student behaviors and techniques that were effective in the traditional scale, may or may not be effective in the EMRN grading scheme. This contributes to a greater degree of dissatisfaction with the scheme, and we speculate distracts students from focusing on the learning outcomes that the scheme tries to promote.

Expectation management can thus be a crucial step in acclimation of these novel practices. Effective communication of the key difference between the two grading schemes, conveyance of such by the instructors and TAs as well setting aside of special time to discuss student complaints could help soften the impact of novelty and help better acclimate the students.

5.1.4 Increased Grade Boundary Sensitivity

The primary intention behind minimizing the number of grades as suggested by Rapaport [32] and not having 0 as a failure grade as suggested by Feldman [11], is to lessen "grade-grubbing" and to decrease the overall failure rates (both discussed in Section 2.1). These approaches are grounded in the belief that fewer grade distinctions can simplify the evaluation process, potentially making learning environments less stressful and more focused on

actual learning rather than just achieving high grades. However, this simplification comes at a cost.

A frequent and common complaint among students across the surveys and interviews was that the EMRN system is not granular enough. A few quotes follow as examples:

> *"The fact that you can only get a 100, 80, 50, or zero, meaning if you have one small mistake you get an 80, or even a few little issues you get an automatic 50."* - A survey respondent from the Fall 2023 semester

> *"differences between different grades are negligible and feel unfair"* - A survey respondent from the Spring 2024 semester

As mentioned previously, this system breaks down the traditional 100-point letter grading scale (A, B, B-, through F) into broader categories of "Exceeds Expectations" (100 points), "Meets Expectations" (80 points), "Requires Improvement" (50 points) and "Not Assessable" (0 points), i.e. for each assessment, it is only possible for them to score one of these four points. A natural side-effect of such a system is that if a student makes mistakes such that it places them in the vicinity of these discrete points e.g. scoring a 82 or a 78, it could lead to them to either score those points (e.g. a 80), or get penalized heavily and be sent down to the next level (e.g. sent to 50 from a 80). The problem is that while the points lost by the student can be minuscule, the penalty can be severe. We surmise that as a result, we saw students complaint about the absence of grades between the 100 and 80 levels and between the 80 and 50 points. However, it is important to point out that we saw no complaints about the levels between 50 and 0 points.

The shift to fewer grade boundaries inherently magnifies the importance of each grade within the EMRN grading scheme. This can inadvertently escalate the stakes associated

with each grading decision. This heightened sensitivity to grade boundaries can significantly increase student anxiety and perceptions of harshness. When a single assessment or grade carries disproportionate weight, students might perceive the grading system as not only punitive but also arbitrary, especially if their performance in one aspect of a course disproportionately affects their overall grade in a way that seems misaligned with their understanding of their performance.

This system's potential to induce anxiety and a sense of unfairness is further complicated by its impact on student behavior. Some students may become overly fixated on minor errors or the pursuit of grade perfection, fearing that small mistakes could have disproportionately large effects on their final grades. While such behaviors might be less prevalent in a traditional grading system with more granular distinctions, the EMRN system's reduced granularity and increased grade boundary sensitivity can exacerbate these tendencies.

Additionally, while some of the assessments purely use the EMRN grading scheme, other assessments like the final exam use a traditional 100-point scale. Thus, the EMRN grading scheme is not used exclusively for the assessments. This can lead to a situation where students anticipate poor performance in the final exam, and to make up for it rely on the other assessments. In such a case, a student who misses a higher grade by a close margin because of the lack of granularity can experience a higher level of dissatisfaction. Thus, future work should involve a better reconciliation of these tow different designs.

Therefore, while the reduction in grade boundaries aims to address certain educational challenges, it introduces a complex array of unintended consequences. The increased importance of each grade boundary, combined with the significant jumps across traditional letter grades, not only heightens student anxiety but also shifts the focus away towards system harshness. This shift undermines the educational value of grading as a tool for

feedback and learning progression, suggesting that any implementation of reduced grading scales must be approached with careful consideration of these broader impacts.

5.2 Late Policy

In this section, we evaluate the results related to RQ2: "We aim to gather information about their perception of the employed *late policy*".

5.2.1 Enhanced Executive Function

The implementation of a seven day late submission policy, as described before, appears to significantly enhance students' executive function, which encompasses skills critical for planning, prioritization, task initiation, and emotional regulation. This finding aids in understanding the broader cognitive and behavioral benefits of EGPs.

One of the core components of executive function is cognitive flexibility, the ability to adapt behaviors and thoughts in response to changing environments and demands (Diamond, 2013). The late policy provided students with a more adaptable framework within which they could maneuver; this flexibility allowed them to reassess and adjust their workload based on a better understanding of their academic and personal commitments. By enabling students to submit work up to seven days late with penalty, the policy not only accommodated diverse student needs but also encouraged strategic planning and prioritization—key aspects of cognitive flexibility. At the same time, since the policy enforced a 20-point deduction if turning things in late, it limited the amount of slack students were given.

Moreover, the option to delay submission deadlines likely supported better planning behaviors. Students reported using the additional time to deliberate more effectively on their

work, indicating an enhancement in their planning skills. Such improvements are crucial for academic success, as they directly relate to students' ability to manage future tasks and deadlines more efficiently.

The flexibility inherent in the late policy also had significant implications for students' emotional regulation, another vital aspect of executive function. The reduced pressure to meet rigid deadlines likely diminished stress and anxiety, which are known to impair decision-making and cognitive performance (Sapolsky, 2004). Students noted a decrease in stress levels, which not only improved their well-being but also allowed them to engage in more thoughtful, less hurried work. This environment, where stress is minimized, supports the development of emotional regulation skills by providing students with the opportunity to manage their academic tasks without the imminent pressure of deadlines.

Finally, the late policy promoted better time management, a skill intricately linked to executive function. With the added flexibility, students expressed that they were better able to distribute their efforts effectively across different tasks and deadlines. This shift towards more proactive and strategic management of time can be seen as an improvement in task initiation and sustained attention, components of executive function that are essential for academic success. The development of these skills not only aids students during their academic careers but also prepares them for professional environments where time management is crucial.

We must mention here that the students whom we interviewed reported not being affected largely by the late policy. While many of them did report a reduction in anxiety and so on, much of them also used the late policy sparingly.

5.2.2 Reliance on Late Policy

The implementation of a seven day late submission policy, while largely beneficial in numerous respects, also results in responses about self-regulation, with some students expressing reservations about relying too heavily on it. This section explores the restraint and cognizance that students demonstrated towards the late policy.

In the student interviews, despite the flexibility offered by the late policy, we observed that a considerable number of students chose to use this option sparingly. The probable reason for these observations stems from the observations of the recurrent concern among students was that the regular use of the late policy might cultivate a habit of tardiness, affecting not just their academic submissions but potentially spilling over into other areas of personal and professional life. The apprehension about developing a lax attitude towards deadlines reflects an acute awareness of the importance of punctuality as a professional virtue. Furthermore, students were conscious of the opportunity costs associated with utilizing the late submission policy. The primary cost identified was the encroachment on time available for subsequent assignments. Hence, this behavior may be indicative of students' concerns about their self-regulation capabilities. The availability of a late submission option could potentially lead to procrastination if used without control. Students who recognized this possibility often opted to adhere to the original deadlines to maintain a disciplined approach to their coursework. This finding suggests that while the policy provides necessary flexibility, it also requires students to exercise a higher degree of self-regulation to avoid the pitfall of delayed submissions becoming a habitual response to all assignments. This was evident with the students interviewed, however not much evidence was found about this in the survey free-response questions.

In the Likert scale responses however, the student responses were found to be neutral on

the questions: "The late policy makes it easier for me to delay working on assignments until the last minute" and "The late policy encourages me to start working on assignments earlier" according to figure 4.4. More students responded on the negative here rather than the affirmative, suggesting that the Likert scale response are inclined more with the interviews in that the cost of the late policy is not lost on the students. Anecdotally, there were students who always turned their work in late. Future work would involve investigating how many students habitually turned things in late.

Further interviews may be required to investigate the perceptions of students who do indeed feel that the late policy promotes procrastination. To reiterate the final point mentioned in Section 5.2.1, the students whom we interviewed reported not being affected largely by the late policy. While many of them did report a reduction in anxiety and so on, much of them also used the late policy sparingly.

In sum, while the late submission policy undoubtedly offers crucial flexibility and support, it also poses challenges that necessitate careful consideration by students. The successful implementation of such policies not only relies on providing students with options but also on guiding them to develop robust self-regulation skills. This ensures that students can make judicious use of the flexibility afforded to them without compromising their academic integrity or future professional habits. Moving forward, it may be beneficial for educators to provide more structured guidance on how to effectively balance the use of late submissions with the development of strong self-regulatory practices.

5.3 Quizzes vs Exams

In this section, we evaluate the results related to RQ3: "Students' perception of being assessed using small quizzes vs bigger exams such as mid-terms and finals".

5.3.1 Better Testing Mechanisms

In exploring equitable grading practices, one significant shift was the transition from fewer, larger exams to more frequent, smaller quizzes. This approach appears to offer several pedagogical benefits that align with the principles of fairness and inclusivity in assessment. The feedback from students provides valuable insights into why quizzes may serve as more effective and equitable instruments for evaluation compared to traditional exams.

One of the primary advantages of quizzes noted by students is that they require shorter performance times. This characteristic significantly reduces the pressure typically associated with longer exam periods, which can exacerbate anxiety and negatively impact student performance [3]. Shorter quizzes minimize the duration of stress, allowing students to focus better on demonstrating their knowledge without the burden of prolonged anxiety. This shift is particularly beneficial for students who may struggle with test anxiety, providing a more accommodating environment that can lead to a truer representation of their capabilities.

Students also appreciated that quizzes allowed for testing across a broader range of topics. This frequent and varied assessment helps ensure that the learning process is continuous and comprehensive. By covering more topics in smaller segments, quizzes can provide a more detailed map of student understanding and progress throughout the course. This method contrasts with traditional exams that may only assess a limited range of knowledge due to time constraints, potentially overlooking areas where some students may excel.

A significant shift away from heavy reliance on memorization was another advantage cited by students in the quiz format. Quizzes often focus on understanding and applying knowledge rather than recalling large amounts of information from memory. This approach not only supports deeper learning but also levels the playing field for students who may have

difficulties with memorization due to various learning differences. By emphasizing comprehension and application, quizzes can more accurately assess a student's ability to integrate and utilize knowledge in practical contexts.

Perhaps one of the most critical features of the quiz format is the opportunity it offers for students to learn from their mistakes and improve over time. Unlike traditional exams where a single poor performance can have a disproportionate relationship with the final grade, quizzes provide multiple opportunities to assess and enhance understanding. This iterative process encourages a growth mindset, where students view assessments as opportunities for learning and improvement rather than as final judgments on their capabilities [8].

5.3.2 Correlation between Quizzes and Test Anxiety

We notice higher levels of test anxiety when analyzing the results of the Likert scale. Firstly, we notice in table 4.5 that when compared to the baseline mean of the test anxiety subscale in the study conducted by Pintrich et al. in 1991 [30], the test anxiety appears to be higher for our student populations across both semesters. While this may at the outset suggest that practices applied in this course are a greater source of test anxiety, it is important to recognize that Pintrich et al.'s study was conducted on a greatly different population of students in a very different area of pedadogical practices.

Secondly, we use the word "test" in all of our questions in the Likert scale because the MSLQ subscale used the same wording. Moreover, these questions are framed generally when taking tests, and not taking tests within the context of this course. In this course, we employed quizzes and have used the word quiz when referring to a closed form of assessment in the syllable. Use of the word test could mean that students may be respond-

ing (on the Likert scale) about tests in general and not about the quizzes in this course. We did not use the word quizzes to reduce ambiguity in student answers.

The students in the free response questions responded negatively while commenting on the level of their test anxiety. Many students responded with an increased level of anxiety when mentioning quizzes. Conversely, students in the interviews when asked about their feeling of anxiety in the context of taking quizzes responded in the positive. They gave responses which resulted in codes such as "Exams-induce-greater-test-anxiety" and "Small-quiz-better-than-exam", suggesting that they prefer quizzes over exams and that they indeed feel a lesser degree of anxiety when appearing for multiple quizzes as opposed to fewer exams, even though the overall assessment time reserved in the course is higher for quizzes.

This dichotomy of opinion between the survey and the interviews has multiple explanations. Firstly, the students interviewed may or may not represent the overall population accurately. Secondly, as mentioned under Question-4 in the interview script (Appendix E), the interviews ask students to compare their test anxiety in the context of smaller quizzes vs bigger exams, a conscious comparison which the surveys do not contain. Thus, additional work may be required to ascertain the cause of the high test anxiety experienced by the larger population of students in the survey by either interviewing more students so that they more accurately represent the overall student population, or by including the aforementioned conscious comparison between smaller quizzes vs bigger exams in the survey.

Additionally, modification of the MSLQ further to reword "test" as "quiz" in the questions as well as adding in more qualifying questions, tying in test anxiety with quizzes more directly may help us in the future to identify the correlation better.

5.4 MSLQ subscale Relationship with perceptions

This section discusses the relations between the MSLQ subscales and their relationships with the EMRN Grading Scale, late policy and Test Anxiety. The table below summarizes the most important relationships between the MSLQ subscales and the subscales of EMRN grading scheme, late policy and Test Anxiety, which is dirived from tables 4.7, 4.9 and 4.11.

Subscale	EMRN Grading Scheme	Late Policy	Test Anxiety
Self-Regulation	**0.29 (p < 0.01)**	**0.32 (p < 0.01)**	**0.39 (p < 0.01)**
Peer Learning	**0.13 (p < 0.01)**	**0.13 (p < 0.01)**	0.23 (p = 0.81)
Self-efficacy	**0.14 (p < 0.01)**	-0.02 (p = 0.6)	**-0.35 (p < 0.01)**
Help Seeking	**-0.1 (p = 0.04)**	-0.08 (p = 0.08)	**0.17 (p < 0.01)**
Intrinsic Goal Motivation	0.06 (p = 0.06)	-0.02 (p = 0.6)	**0.17 (p = 0.02)**

Table 5.1: A summary of the significant relationships between MSLQ subscales and EMRN grading scheme, Late Policy and Test Anxity subscales (Threshold - $p < 0.05$

5.4.1 Relationships with the EMRN Grading Scheme

The EMRN grading scheme is a specification-based grading system that allows students multiple opportunities to achieve their desired grade by meeting specific criteria laid out for each assessment. This method of grading focuses on achieving clear, predefined standards rather than competing against peers, which shifts the educational focus to mastery and personal improvement.

1. **Self-Regulation and EMRN:** Self-regulation is positively correlated with favorable perceptions of the EMRN grading scheme. Students who excel in self-regulation likely find this grading system beneficial as it allows them to plan, monitor, and evaluate their learning strategies in alignment with clearly defined benchmarks. These

students can iteratively refine their work to move from "Requires Improvement" to "Meets Expectations" or even "Exceeds Expectations", applying their self-regulatory skills to effectively meet specified criteria. The ability to revise submissions based on feedback makes EMRN particularly appealing to these learners, fostering a proactive approach to education.

2. **Peer Learning's Role in EMRN:** Peer learning also shows a positive relationship with the EMRN scheme. Since this grading system encourages continuous improvement through re-assessment, students who engage in peer learning benefit from discussing and refining their understanding with others. This collaborative learning process is invaluable in a system like EMRN, where feedback can be incorporated before resubmitting work. Students can leverage their social learning networks to elevate their work from "Requires Improvement" to higher standards, enhancing both individual and group learning outcomes.

3. **Impact of Self-Efficacy on EMRN Perceptions:** Students with high self-efficacy, who are confident in their academic abilities, tend to view the EMRN grading scheme positively. For these students, the opportunity to revise and improve their work confirms their capability to meet and exceed academic standards. Each revision cycle under the EMRN system is seen as a chance to demonstrate competence and achieve "Exceeds Expectations", reinforcing their self-confidence and motivation towards academic tasks.

5.4.2 Relationships with the Late Policy

The late policy in educational settings often serves as a buffer for students, allowing additional time to complete assignments. This policy is particularly favorable among students

who exhibit strong self-regulation and engage actively in peer learning. The flexibility afforded by the late policy can be crucial in helping students manage their academic tasks more effectively and leverage collaborative opportunities.

1. **Self-Regulation and Its Impact on late policy Perception:** Self-regulation, which involves the ability to manage one's time, effort, and learning strategies efficiently, aligns well with the advantages of a late policy. Students who are adept at self-regulation tend to plan their work with an understanding of potential setbacks. The late policy provides a safety net, enabling these students to adjust their plans and strategies without the immediate pressure of impending deadlines. As a result, self-regulated learners appreciate the late policy as it complements their strategic approach to learning, giving them confidence that they can recover from any academic delays without penalty.

2. **Peer Learning Enhancements via Late Policy:** Peer learning—where students collaborate, share resources, and support each other's learning journeys—also benefits significantly from the late policy. This policy extends the time available for collaborative learning activities, which can be particularly beneficial before critical submissions. Students who rely on peer interactions to enhance their understanding and performance find that the late policy enables more comprehensive and less rushed collaborative sessions. This additional time not only helps solidify their knowledge through discussion and mutual assistance but also fosters a more supportive learning environment.

5.4.3 Relationships with Test Anxiety

Test anxiety is a significant concern in educational settings as it can have a strong relationship with student performance. Understanding how different aspects of the Motivated Strategies for Learning Questionnaire (MSLQ) influence test anxiety can help educators develop targeted interventions to alleviate this stress.

1. **Self-Efficacy's Role in Reducing Test Anxiety:** The analysis of parameter estimates shows a clear negative correlation between self-efficacy and test anxiety. High self-efficacy, which is the belief in one's own ability to succeed in specific situations, helps reduce test anxiety by counteracting self-doubt. Students with strong self-efficacy feel more competent and in control, which decreases their anxiety levels during tests. This finding underscores the importance of enhancing students' confidence in their academic abilities as a strategy to combat test anxiety.

2. **Metacognitive Self-Regulation and Increased Anxiety:** Surprisingly, metacognitive self-regulation correlates positively with test anxiety. While self-regulation skills are typically associated with better academic outcomes, they may also lead to increased anxiety. This could be due to the heightened self-awareness and constant self-monitoring that come with metacognitive practices. Students who frequently evaluate their own understanding and progress might become overly critical or anxious about their performance, particularly in high-stakes testing scenarios. This aspect of self-regulation may inadvertently heighten feelings of anxiety by making students more aware of the stakes involved and their own potential for failure.

3. **Help-Seeking and Its Association with Test Anxiety:** There is also a notable positive correlation between help-seeking behavior and test anxiety. Students who often seek help may be more aware of their academic limitations, which is a healthy

recognition in most learning environments. However, during tests, the typical sources of support—such as peers, tutors, or educators—are not available. This lack of support can exacerbate anxiety for those who rely heavily on external assistance for academic success. Understanding this dynamic is crucial for developing support systems that encourage more independent problem-solving skills in preparation for tests.

5.5 Threats to Validity

The following threats to validity exist for our discussion before:

1. The use of ANOVA to analyze the relationship between the Motivated Strategies for Learning Questionnaire (MSLQ) and the subscales of EMRN grading scheme and late policy, which are not validated raises concerns about the validity of the results. While both of these are intended to be a subscale, their inclusion without validation could lead to misinterpretation of the data.

2. Possible demographic changes between the Fall and Spring semesters, such as differing proportions of first-time and repeat course takers, could confound the results. Additionally, changes in instructor, textbooks, and course settings may have influenced student experiences and perceptions.

3. Comparing pre-survey and post-survey results from different populations introduces a potential threat to the validity of the study. Differences in these populations may confound the results and limit the ability to draw meaningful conclusions.

4. Conducting interviews towards the beginning and towards the end of the course could introduce biases in the responses. The timing of the interviews may influence

participants' recollection and perception of their experiences, potentially skewing the data.

5. The limited number of interviews conducted may restrict the generalizability of the findings. A larger sample size would enhance the robustness of the study's conclusions.

6. The grading of EMRN on a 100-point scale may have influenced student perceptions. This approach conflates perceptions about the grading scale with perceptions about the EMRN itself, making it difficult to separate the two.

7. The inconsistency in grading, with some aspects on a 100-point scale and others on the EMRN scale, further complicates the interpretation of student perceptions. A uniform grading system would have provided clearer insights into student opinions.

8. The use of Cohen's D metric, despite the subscales not being normally distributed, raises concerns about the appropriateness of the statistical analysis. Non-parametric methods may be more suitable for comparing metrics in this context.

9. The surveys' focus on general test anxiety, rather than quiz-specific anxiety, may limit the relevance of the results to the study's specific context. Similarly, the lack of questions comparing quizzes to larger exams in the free-response section may limit the depth of insights gained from the survey data.

Chapter 6

Conclusions

This chapter presents conclusions derived from the results and discussions sections of the thesis. It is divided into sections, each addressing the following research questions:

1. **RQ1**: We aim to understand student perceptions of the employed specification grading scheme, called the 'EMRN grading scheme'

2. **RQ2**: We aim to gather information about their perception of the employed *late policy*

3. **RQ3**: Students' perception of being assessed using small quizzes vs bigger exams such as mid-terms and finals

4. **RQ4**: How does students' self-regulated learning beliefs affect their perceptions of EGPs.

This chapter focuses on student perceptions of the specific versions of the EGPs implemented in two instances of CS-1114 in Fall 2023 and Spring 2024. These include the EMRN grading scheme, a 7-day late policy extension with a maximum attainable grade, and the use of 8 smaller quizzes instead of midterms.

6.1 Student Perceptions of the EMRN Grading Scheme

The analysis of student perceptions regarding the EMRN grading scheme has revealed a complex landscape of opinions, showcasing a mixture of challenges and potential advantages. This section aims to provide a clearer synthesis of these findings, detailing both the immediate issues presented by the grading scheme and offering practical recommendations for addressing these concerns.

Positive Impacts of the EMRN grading scheme:

1. Alignment with Learning Outcomes: Despite the challenges, some students perceive the EMRN grading scheme as more closely aligned with their actual learning outcomes. However, this perception may be influenced by the issues with the grading scale. Further research is needed to verify this potential benefit. A detailed study focusing on specific learning outcomes and their correlation with the grading criteria under the EMRN system would provide clearer insights.

2. Longitudinal Studies: Longitudinal studies tracking the impact of the EMRN grading scheme over several academic cycles could provide deeper insights into its effectiveness and reception over time. Such studies could help determine the long-term benefits or detriments of the system, providing a comprehensive basis for its evaluation and possible adjustment.

Challenges Presented by the EMRN grading scheme:

1. Sensitivity of Grade Boundaries: A primary concern expressed by students revolves around the sensitivity of grade boundaries, particularly between the “E” and “M” grades. This sensitivity has resulted in heightened anxiety and dissatisfaction among

students, as small differences in performance can lead to significant jumps in final grades. To address this issue, introducing an additional grade between "E" and "M" could help smooth these transitions, making the grading process feel more fair and less arbitrary.

2. Human Factors and Acclimation Challenges: The negative perceptions of the EMRN grading scheme are also influenced by human factors, including unfamiliarity with the new grading scale and misaligned expectations. Many students are accustomed to more traditional grading scales, and the novelty of the EMRN system can lead to confusion and resistance.

3. The mapping of specification grading schemes to traditional 100-point scales further complicates this issue. Anecdotally, instructors have mentioned difficulties in mapping this scheme to a 100-point scale, as most universities use a 100-point scale divided into various letter grades. Using a 100-point scale may contribute to dysfunction. Future work will involve incorporating a specification grading scheme like the EMRN grading scheme into a 100-point scale in a way that reduces this friction.

4. Additionally, while some assessments exclusively use the EMRN grading scheme, others, like the final exam, use a traditional 100-point scale. This can lead to a situation where students anticipate poor performance in the final exam and rely on the other assessments to compensate. In such a case, a student who narrowly misses a higher grade due to the lack of granularity may experience a higher level of dissatisfaction. Thus, future work should involve better reconciliation of these two different designs.

Future Work:

1. Enhanced Communication and Additional Support: To mitigate the challenges associated with the novelty of the EMRN grading scheme, institutions should invest in comprehensive communication strategies. This could involve detailed orientations for new students, ongoing educational campaigns, clear and accessible resources explaining how the grading system works and what students can expect and also detailed communication plans between the instructors/TAs and students. Properly setting and managing expectations can significantly ease the transition and reduce resistance to new grading practices.

2. Incremental Introduction of Changes: Implementing gradual changes to the grading system could help ease student acclimation. By introducing one aspect of the new grading scale at a time, students and faculty can adjust more comfortably, reducing the shock and confusion associated with a complete overhaul.

6.2 Student Perceptions of the Late Policy

This section presents the conclusions that we draw on the late policy's relationship with student experiences and behaviors, focusing on its impact on academic performance and self-regulation. The students' generally positive perceptions of the late policy underscore its potential benefits for fostering a supportive academic environment. However, the complexities associated with its use reveal areas for cautious consideration and further investigation.

Positive Impacts of the Late Policy:

1. Improved Executive Function: The late policy has been viewed favorably by many students for its role in enhancing executive function. By allowing assignments to be

submitted late without immediate penalty, the policy reduces anxiety around strict deadlines, affording students the flexibility to better manage their time and adapt to unforeseen circumstances. This adaptability is crucial for students juggling multiple responsibilities or facing unexpected life events, as it helps them maintain their academic performance without undue stress.

2. Support for Better Schedule Integration: Students have noted that the late policy helps in integrating academic responsibilities more smoothly into their personal schedules. This flexibility is particularly beneficial for non-traditional students, who may be balancing studies with work or family commitments, thereby making higher education more accessible and manageable.

3. Recognition of Opportunity Costs: Interviews revealed a sophisticated awareness among students of the potential "slippery slope" of relying too heavily on the late policy. Many students expressed concerns about the opportunity costs involved, such as reduced time available for subsequent assignments and the potential for accumulating workloads. This awareness suggests that while the policy offers immediate relief and flexibility, it may also lead to greater pressures down the line if not managed carefully.

Challenges Presented by the Late Policy:

1. Promotion of Tardiness and Procrastination: Despite its benefits, the late policy is not without its drawbacks. According to responses from the Likert scale, a significant number of students perceive that the policy may encourage tardiness and procrastination. This perception highlights the potential risk of diminishing the urgency with which students approach deadlines, which could adversely affect their long-term academic habits and professional preparedness.

Future Work:

1. Impact on Diverse Student Groups: Future studies should aim to capture a broader spectrum of experiences, particularly from students who feel that the late policy may have negatively influenced their academic behaviors or exacerbated tendencies toward procrastination. Understanding these diverse impacts can help in refining the policy to better support all students.

2. Longitudinal Studies: Implementing longitudinal research could provide deeper insights into how the late policy affects student behavior and performance over longer academic periods. This approach would help in assessing the long-term benefits and potential drawbacks of such policies in academic settings.

3. Development of Support Structures: Educational institutions might consider developing support structures that can help students make the most of the late policy without falling into patterns of procrastination. Workshops on time management, personalized academic coaching, and increased awareness campaigns about the judicious use of such policies could mitigate some of the risks while enhancing the benefits.

6.3 Student Perceptions of Smaller Quizzes vs Bigger Exams

This section presents substantial insights into the benefits and effectiveness of employing quizzes as part of equitable grading practices, based on students' perceptions of being assessed using small quizzes versus larger exams such as mid-terms and finals. It draws

conclusions from the discussed results and proposes recommendations for educational practices and further research. It is important to reiterate the concerns around the wording and the structure of the test anxiety subscale in the survey, as mentioned in section 5.3.2. We used the word "test" in Likert scale questions of the Test Anxiety subscale as opposed to using the word quizzes. This asks the student to respond as though they are giving their opinions on general test-taking, rather than focusing on this course's quizzes. Additionally, the interviews prompt students to compare test anxiety between smaller quizzes and bigger exams, to which the free-response questions of the survey do not call students' attention.

Positive Impacts of Smaller Quizzes:

1. Reduced Anxiety and Enhanced Performance: The transition to smaller, more frequent quizzes has been positively received by students, primarily due to the reduced anxiety associated with shorter assessment times. This finding is significant as it highlights quizzes' role in creating a less stressful assessment environment, which can lead to a more accurate representation of student capabilities.

2. Comprehensive Assessment and Learning Continuity: Quizzes allow for a broader and more continuous assessment of student knowledge. By covering more topics in smaller segments, quizzes provide a comprehensive overview of students' understanding and progress, offering a contrast to the potentially narrow focus of traditional exams.

3. Focus on Understanding Over Memorization: The quiz format's emphasis on understanding and applying knowledge rather than rote memorization has proven beneficial, particularly for students with learning differences. This shift supports deeper learning and provides all students with a fair chance to demonstrate their true un-

derstanding.

4. Iterative Learning and Improvement: The frequent and iterative nature of quizzes offers students multiple opportunities for feedback and improvement. This approach fosters a growth mindset, encouraging students to view assessments as opportunities for learning rather than final judgments on their academic abilities.

Based on these findings, we can conclude that there is value in integrating more frequent quizzes into assessment strategies to replace or complement traditional high-stakes exams. This approach not only aligns with equitable grading practices but also enhances the educational experience by supporting continuous learning and reducing student anxiety.

Future Work: Further research should explore the long-term impacts of quiz-based assessment strategies on various academic outcomes, including student retention and overall performance. Additionally, modifying existing measurement tools, such as the MSLQ, to better differentiate between quizzes and exams could provide clearer insights into the specific impacts of different assessment types on student anxiety and performance along with asking students to consciously compare their test anxiety between smaller quizzes and bigger exams would shed more light on their perceptions.

6.4 Relationships between MSLQ subscales and Relevant Subscales

Equitable grading practices are essential for fair assessment in education. These practices are designed to ensure that all students have an equal opportunity to succeed, regardless of their backgrounds or personal challenges. This research examines the relationship between students' beliefs in self-regulated learning—including self-efficacy, peer learning, and

metacognitive self-regulation—and their views on grading systems like the EMRN grading scheme and late policy. By exploring how these beliefs influence perceptions of fairness in grading, the study aims to provide insights that could help educators improve assessment methods to support every student's academic progress.

1. EMRN Grading Scheme and Equitable Grading: Self-regulated learning significantly shapes students' perceptions of the EMRN grading scheme, a specification-based grading system that emphasizes mastery and personal improvement over peer competition. Students with high self-regulation skills find this system particularly beneficial as it aligns with their abilities to plan, monitor, and evaluate their progress against clear benchmarks. This capability enables them to iterate on their submissions to meet or exceed expectations, highlighting how equitable grading practices can support diverse learning strategies and promote fairness in academic assessments. Peer learning and self-efficacy also play crucial roles by fostering collaborative environments and reinforcing confidence in meeting academic standards, respectively.

2. Late Policy as a Facilitator of Equitable Learning: The late policy is viewed positively by students who are adept in self-regulation and peer learning. This policy, by providing extra time for assignment completion, acts as an equalizer, mitigating the disadvantages that might arise from unexpected personal or academic challenges. It allows students to manage their tasks effectively without penalty for delays, which is crucial for maintaining equity in grading. The extension of deadlines under this policy is particularly beneficial in collaborative learning settings, enabling students to maximize their peer resources and refine their collective outputs.

3. Impact on Test Anxiety and Perceptions of Fairness: Test anxiety, which is inversely

related to self-efficacy and positively correlated with metacognitive self-regulation and help-seeking behaviors, presents a nuanced challenge to equitable grading practices. While self-efficacy reduces anxiety by boosting confidence, metacognitive self-regulation may inadvertently increase stress due to heightened awareness of one's own learning process and potential shortcomings. Additionally, students who rely on help-seeking may feel disadvantaged during tests where such supports are unavailable, suggesting that equitable grading practices must consider these dynamics to truly level the playing field.

The findings suggest that students' beliefs in self-regulated learning not only influence their personal academic performance but also their perceptions of the fairness and effectiveness of grading practices. Educational policies that incorporate flexibility, such as the EMRN grading scheme and late policy, and that acknowledge the varied impacts of self-efficacy, self-regulation, and peer learning can lead to more equitable outcomes. By accommodating diverse learning behaviors and providing multiple avenues for demonstrating knowledge and skills, institutions can foster an educational environment that is truly inclusive and equitable.

Future Work:

Educational institutions should consider implementing and promoting grading schemes that accommodate diverse learning strategies and provide clear, achievable standards for all students. Training for teachers on the nuances of self-regulated learning and its impact on student perceptions of fairness could further enhance the effectiveness of these practices. Additionally, support systems should be bolstered to aid students who exhibit high levels of test anxiety, ensuring that all students have a fair chance to succeed regardless of their individual learning challenges.

Bibliography

[1] American College Health Association. American college health association-national college health assessment ii: Reference group executive summary spring 2013. *Hanover, MD: American College Health Association*, 2013.

[2] Judy Cameron and W David Pierce. The debate about rewards and intrinsic motivation: Protests and accusations do not alter the results. *Review of educational research*, 66(1):39–51, 1996.

[3] Jerrell C Cassady. The influence of cognitive test anxiety across the learning–testing cycle. *Learning and instruction*, 14(6):569–592, 2004.

[4] Hongxuan Chen, Ang Li, Geoffrey Challen, and Kathryn Irene Cunningham. Implementation of split deadlines in a large cs1 course. In *SIGCSE 2024 - Proceedings of the 55th ACM Technical Symposium on Computer Science Education*, SIGCSE 2024 - Proceedings of the 55th ACM Technical Symposium on Computer Science Education, pages 193–199, United States, March 2024. Association for Computing Machinery. doi: 10.1145/3626252.3630873. Publisher Copyright: © 2024 ACM.; 55th ACM Technical Symposium on Computer Science Education, SIGCSE 2024 ; Conference date: 20-03-2024 Through 23-03-2024.

[5] Adrienne Decker, Stephen H. Edwards, Brian M. McSkimming, Bob Edmison, Audrey Rorrer, and Manuel A. Pérez Quiñones. Transforming grading practices in the computing education community. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, SIGCSE 2024, page 276–282, New York,

NY, USA, 2024. Association for Computing Machinery. ISBN 9798400704239. doi: 10.1145/3626252.3630953. URL https://doi.org/10.1145/3626252.3630953.

[6] Adrienne Decker, Stephen H Edwards, Brian M McSkimming, Bob Edmison, Audrey Rorrer, and Manuel A Pérez Quiñones. Transforming grading practices in the computing education community. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, pages 276–282, 2024.

[7] Pleasanton Unified School District and Joe Feldman. Equitable grading practices - pleasanton unified school district. URL https://www.pleasantonusd.net/departments/teaching-learning/curriculum-instruction/equitable-grading-practices.

[8] Carol S Dweck. *Mindset: The new psychology of success*. Random house, 2006.

[9] Stephen Edwards, Manuel Pérez-Quiñones, Adrienne Decker, and Audrey Rorrer. Award abstract # 2235337 transforming grading practices in the computing education community. 2022. URL https://www.nsf.gov/awardsearch/showAward?AWD_ID=2235337&HistoricalAwards=false.

[10] Fara Elikai and Peter Schuhmann. An examination of the impact of grading policies on students' achievement. *Issues in Accounting Education*, 25(4):677–693, 11 2010. doi: 10.2308/iace.2010.25.4.677. URL https://publications.aaahq.org/iae/article-abstract/25/4/677/7746.

[11] Joe Feldman. *Grading for Equity: What It Is, Why It Matters, and How It Can Transform Schools and Classrooms*. Corwin Press, Thousand Oaks, CA, 2019.

[12] Andy Fleenor, Sarah Lamb, Jennifer Anton, Todd Stinson, and Tony Donen. The grades game., February 2011. URL https://eric.ed.gov/?id=EJ916326.

[13] The Derek Bok Center for Teaching and Learning. Beyond "the grade": Alternative approaches to assessment. URL https://bokcenter.harvard.edu/beyond-the-grade. Accessed on March 12, 2024.

[14] B. Glaser and A. Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Research. Mill Valley.* CA: Sociology Press, 1967.

[15] Ray Hembree. Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, 58(1):47–77, 1988. doi: 10.3102/00346543058001047. URL https://doi.org/10.3102/00346543058001047.

[16] John R. Hott. Analyzing student performance with free late submission days. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2*, SIGCSE 2024, page 1682–1683, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400704246. doi: 10.1145/3626253.3635562. URL https://doi.org/10.1145/3626253.3635562.

[17] Patricia Hutton. Understanding student cheating and what educators can do about it. *College Teaching*, 54, 01 2006. doi: 10.3200/CTCH.54.1.171-176.

[18] Caesar R. Jackson. Validating and adapting the motivated strategies for learning questionnaire (mslq) for stem courses at an hbcu. *AERA Open*, 4(4): 2332858418809346, 2018. doi: 10.1177/2332858418809346. URL https://doi.org/10.1177/2332858418809346.

[19] Russell Jones. Gender-Specific differences in the perceived antecedents of academic stress. *Psychological Reports*, 72(3):739–743, 6 1993. doi: 10.2466/pr0.1993.72.3.739. URL https://doi.org/10.2466/pr0.1993.72.3.739.

[20] Joseph Kim and Christian Murphy. Student perspectives on assignment deadline policies in computer science courses. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, SIGCSE 2024, page 659–665, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400704239. doi: 10.1145/3626252.3630932. URL https://doi.org/10.1145/3626252.3630932.

[21] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621, 1952.

[22] Wonki Lee, Jeffrey F Rhoads, Edward J Berger, and Jennifer Deboer. Wip: Validating a motivated strategy for learning questionnaire (mslq) in an active, blended, and collaborative (abc) dynamics learning environment. In *2020 ASEE Virtual Annual Conference Content Access*, 2020.

[23] Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.

[24] Robert J Marzano, Barbara B Gaddy, and Ceri Dean. What works in classroom instruction. 2000.

[25] Travis Mitchell. Most U.S. teens see anxiety, depression as major problems | Pew Research Center, 5 2020. URL https://www.pewresearch.org/social-trends/2019/02/20/most-u-s-teens-see-anxiety-and-depression-as-a-major-problem-among-their-peers/.

[26] Nadia Najjar and Debarati Basu. Implementation and evaluation of equitable practices in an introductory cs course. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2*, pages 1758–1759, 2024.

[27] Linda B. Nilson and Claudia J. Stanny. *Specifications grading*. 6 2023. doi: 10.4324/9781003447061. URL https://doi.org/10.4324/9781003447061.

[28] Ken O'Connor. Reforming grading practices in secondary schools. *Principal's Research Review*, 4(1):1–7, 2009.

[29] Great Schools Partnership. Research Supporting Proficiency-Based Learning: Grading + Reporting | Great Schools Partnership, 6 2018. URL https://www.greatschoolspartnership.org/proficiency-based-learning/research-evidence/research-supporting-ten-principles-grading-reporting/.

[30] P. R. Pintrich, D. Smith, T. Garcia, and W. McKeachie. *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, MI, 1991.

[31] Paul R. Pintrich and Elisabeth Vialpando De Groot. Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1):33–40, 3 1990. doi: 10.1037/0022-0663.82.1.33. URL https://doi.org/10.1037/0022-0663.82.1.33.

[32] William J. Rapaport. A triage theory of grading. *Teaching Philosophy*, 34(4):347–372, 1 2011. doi: 10.5840/teachphil201134447. URL https://doi.org/10.5840/teachphil201134447.

[33] Douglas B Reeves. *Leading change in your school: How to conquer myths, build commitment, and get results*. Ascd, 2009.

[34] G. E. Samson, Elizabeth Graue, T. Weinstein, and H. Walberg. Academic and Occupational Performance: a quantitative synthesis, 1984. URL https://www.semanticscholar.org/paper/Academic-and-Occupational-Performance%3A-A-Synthesis-Samson-Graue/6c075f423eada917677de37dd3549722d416428d.

[35] Deborah Selby and Sharon Murphy. Graded or degraded: Perceptions of letter-grading for mainstreamed learning-disabled students. *BC Journal of Special Education*, 16(1):92–104, 1992.

[36] Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611, 1965.

[37] Lorrie A Shepard. Ambitious teaching and equitable assessment: A vision for prioritizing learning, not testing. *American Educator*, 45(3):28, 2021.

[38] Katie M. Silaj, Shawn T. Schwartz, Alexander L. M. Siegel, and Alan D. Castel. Test anxiety and metacognitive performance in the classroom. *Educational psychology review*, 33(4):1809–1834, 3 2021. doi: 10.1007/s10648-021-09598-6. URL https://doi.org/10.1007/s10648-021-09598-6.

[39] Diantha Soemantri, Geoff McColl, and Agnes E. Dodds. Measuring medical students' reflection on their learning: modification and validation of the motivated strategies for learning questionnaire (MSLQ). *BMC Medical Education*, 18(1), 11 2018. doi: 10.1186/s12909-018-1384-y. URL https://doi.org/10.1186/s12909-018-1384-y.

[40] Rodney Y. Stutzman and Kimberly H. Race. EMRF: Everyday Rubric Grading., 1 2004. URL https://eric.ed.gov/?id=EJ717675.

[41] Mary Beth Townsend. Why I stopped giving zeros, 10 2022. URL https://www.edutopia.org/article/why-i-stopped-giving-zeros.

[42] Matt Townsley. What does the research say about standards-based grading?, 6 2016. URL http://mctownsley.net/standards-based-grading-research/.

[43] Frank Vahid, Ashley Pang, and Kelly Downey. Towards grading for equity in a large cs1 class: An experience with flexible deadlines and resubmissions. In *Proceedings of*

the 2023 Conference on Innovation and Technology in Computer Science Education V. 2, pages 646–646, 2023.

[44] Lyanda Vermeulen and Henk Schmidt. Learning environment, learning process, academic outcomes and career success of university graduates. *Studies in Higher Education - STUD HIGH EDUC*, 33:431–451, 08 2008. doi: 10.1080/03075070802211810.

[45] Nathaniel von der Embse and Ramzi Hasson. Test anxiety and high-stakes test performance between school settings: Implications for educators. *Preventing School Failure: Alternative Education for Children and Youth*, 56(3):180–187, 2012. doi: 10.1080/1045988X.2011.633285. URL https://doi.org/10.1080/1045988X.2011.633285.

[46] Carol Larson Wiekhorst. *Relationships between Sources of School Related Pressure and Academic Achievement*. PhD thesis, University of Illinois at Urbana-Champaign, 1973. URL https://www.proquest.com/docview/302646613. ProQuest Dissertations & Theses Global.

[47] Rick Wormeli. Accountability: Teaching through assessment and feedback, not grading. *American secondary education*, pages 14–27, 2006.

[48] Barry J. Zimmerman. *Attaining Self-Regulation*. 1 2000. doi: 10.1016/b978-012109890-2/50031-7. URL https://doi.org/10.1016/b978-012109890-2/50031-7.

Appendices

Appendix A

IRB Approval Letter

Division of Scholarly Integrity and Research Compliance
Institutional Review Board
North End Center, Suite 4120 (MC 0497)
300 Turner Street NW
Blacksburg, Virginia 24061
540/231-3732
irb@vt.edu
http://www.research.vt.edu/sirc/hrpp

MEMORANDUM

DATE: November 13, 2023

TO: Stephen H Edwards, Bob Edmison Jr, Anmol Shukla

FROM: Virginia Tech Institutional Review Board (FWA00000572)

PROTOCOL TITLE: Collaborative Research: Transforming Grading Practices in the CS Education Community

IRB NUMBER: **22-916**

Effective November 13, 2023, the Virginia Tech Human Research Protection Program (HRPP) determined that this protocol meets the criteria for exemption from IRB review under 45 CFR 46.104 (d) category(ies) 1,2(ii),3(i)(B),4(ii).

Ongoing IRB review and approval by this organization is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities impact the exempt determination, please submit an amendment to the HRPP for a determination.

This exempt determination does not apply to any collaborating institution(s). The Virginia Tech HRPP and IRB cannot provide an exemption that overrides the jurisdiction of a local IRB or other institutional mechanism for determining exemptions.

All investigators (listed above) are required to comply with the researcher requirements outlined at:

https://secure.research.vt.edu/external/irb/responsibilities.htm

(Please review responsibilities before beginning your research.)

PROTOCOL INFORMATION:

Determined As:	**Exempt, under 45 CFR 46.104(d) category(ies) 1,2(ii),3(i)(B),4(ii)**
Protocol Determination Date:	**November 10, 2023**

ASSOCIATED FUNDING:

The table on the following page indicates whether grant proposals are related to this protocol, and which of the listed proposals, if any, have been compared to this protocol, if required.

SPECIAL INSTRUCTIONS:

This is an amendment, submitted on 11/13/23, to add the collection of the email address.

Date*	OSP Number	Sponsor	Grant Comparison Conducted?
10/06/2022	PXGZMEJY	National Science Foundation (Title: Collaborative Research: Transforming Grading Practices in the CS Education Community)	Not required (Exempt approval)

* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this protocol is to cover any other grant proposals, please contact the HRPP office (irb@vt.edu) immediately.

Appendix B

Survey Content

Survey: Opinions on Course Grading Policies

This voluntary survey includes questions regarding your opinions on the grading policies in this course and how you approach learning in this course. We will use this information to understand better how you are affected by these policies so that we can improve the course.

The results from this survey will be used for research purposes and for improving the policies used in other courses. Please complete all items, even if you feel that some are redundant. This may require 30 minutes of your time. Usually it is best to respond with your first impression, without giving a question much thought. Your answers will remain confidential, and will not affect your grade in any way.

If you are a minor (under 18 years old), please skip this survey.

Your E-mail Address

Please enter your email address to ensure you receive credit for submitting this survey:
_____

EMRN Grading

Some assignments in this course use an EMRN grading scale (Excellent, Meets Expectations, Revision Needed, or Not Assessable). For the following questions, mark the rating on the 5-point scale for how strongly you **agree or disagree**:

	Strongly Disagree 1	2	3	4	**Strongly Agree** 5
Earning a good grade is a strong motivating factor for me	○	○	○	○	○
I am very familiar with the EMRN grading scale	○	○	○	○	○
EMRN grading is a reliable way of assessing my skills and knowledge	○	○	○	○	○
EMRN grading causes me to focus more on what I should learn in each assignment	○	○	○	○	○
EMRN grading increases my anxiety in this course	○	○	○	○	○
EMRN grading means I don't have to focus on earning each individual point on an assignment	○	○	○	○	○
I prefer EMRN grading to using regular points for grading	○	○	○	○	○
I expect my course grade to be higher because of EMRN grading	○	○	○	○	○

Late Policy

Consider the late policy used on the assignments in this course. For the following questions, mark the rating on the 5-point scale for how strongly you **agree or disagree**:

	Strongly Disagree 1	2	3	4	**Strongly Agree** 5
The late policy reduces my stress about assignment deadlines	○	○	○	○	○
The late policy gives me the flexibility I need to deal with unanticipated situations	○	○	○	○	○
The late policy is unfair	○	○	○	○	○
The late policy encourages me to start working on assignments earlier	○	○	○	○	○
The late policy makes it easier for me to delay working on assignments until the last minute	○	○	○	○	○
The late policy helps me succeed in this course	○	○	○	○	○

Please rate the following items based on your behavior in this class. Your rating should be on a 7- point scale where 1 = not at all true of me to 7 = very true of me:

	Not at All True of Me						**Very True of Me**
	1	**2**	**3**	**4**	**5**	**6**	**7**
In this course, I am more interested in understanding the material than getting a good grade	○	○	○	○	○	○	○
When studying for this course, I try to determine which concepts I don't understand well	○	○	○	○	○	○	○
I often find myself questioning things I hear or read in this course to decide if I find them convincing	○	○	○	○	○	○	○
When I take a test I think about how poorly I am doing compared to other students	○	○	○	○	○	○	○
I'm confident I understand the most difficult learning material presented in this course	○	○	○	○	○	○	○
The most satisfying thing for me in this course is reaching an understanding of the content as thoroughly as possible	○	○	○	○	○	○	○
When I become confused about something I'm reading for this course, I go back and try to sort it out	○	○	○	○	○	○	○
I treat the learning material as a starting point and try to develop my own ideas about it	○	○	○	○	○	○	○
In a course like this, I prefer material that arouses my curiosity, even if it is difficult to learn	○	○	○	○	○	○	○
When I take a test I think about other items on other parts of the test I can't answer	○	○	○	○	○	○	○
I'm confident I can do well on the assessment in this course	○	○	○	○	○	○	○
When studying for this course, I often try to explain the material to a classmate or a friend	○	○	○	○	○	○	○
I ask myself questions to make sure I understand the material I have been studying in this course	○	○	○	○	○	○	○
Before I study new course material thoroughly, I often skim it to see how it is organized	○	○	○	○	○	○	○
Even if I have trouble learning the material in this course I try to work things out for myself	○	○	○	○	○	○	○
When I take tests I think of the consequences of failing	○	○	○	○	○	○	○
When studying for this course, I often set aside time to discuss course material with a group							
When a theory, interpretation, or conclusion is presented in the course, I try to find if there is good supporting evidence	○	○	○	○	○	○	○

	Not at All True of Me 1	2	3	4	5	6	Very True of Me 7
I'm confident I can learn the basic concepts taught in this course	○	○	○	○	○	○	○
If I get confused taking notes or learning clinical skills in the course, I make sure I sort it out afterwards	○	○	○	○	○	○	○
In a course like this, I prefer material that really challenges me so I can learn new things	○	○	○	○	○	○	○
I have an uneasy, upset feeling when I take an exam	○	○	○	○	○	○	○
I believe I will receive an excellent grade in this course	○	○	○	○	○	○	○
When I can't understand the material in this course, I ask another student in this course for help	○	○	○	○	○	○	○
If course materials are difficult to understand, I read them in a different way	○	○	○	○	○	○	○
When reading for this course, I generate questions to help focus my reading	○	○	○	○	○	○	○
I try to play around with ideas of my own related to what I am learning in this course	○	○	○	○	○	○	○
I'm confident I can understand the most complex material presented by the teachers in this course	○	○	○	○	○	○	○
I try to identify students in this course whom I can ask for help if necessary	○	○	○	○	○	○	○
I try to change the way I study in order to fit the course requirements	○	○	○	○	○	○	○
Whenever I read or hear an assertion or conclusion in this course, I think about possible alternatives	○	○	○	○	○	○	○
I'm certain I can master the skills being taught in this course	○	○	○	○	○	○	○
When studying for this course, I try to think through a topic and decide what I am supposed to learn from it rather than just reading it over	○	○	○	○	○	○	○
I try to work with other students from this course to complete the course assignments	○	○	○	○	○	○	○
I feel my heart beating fast when I take an exam	○	○	○	○	○	○	○
When I study for this course, I set goals for myself in order to direct my activities in each study period	○	○	○	○	○	○	○
I expect to do well in this course	○	○	○	○	○	○	○

What was most beneficial aspect of the grading policies in this course?

What was the most anxiety-provoking aspect of the grading policies in this course?

How can we improve the grading policies in this course?

INFORMED CONSENT FORM FOR INVESTIGATIVE PROJECTS
Virginia Tech

Title of Project: Collaborative Research: Transforming Grading Practices in the CS Education
Study Investigators: Dr. Stephen Edwards and Dr. Bob Edmison, Virginia Tech

1. **Purpose of the Study:** The purpose of this study is to determine the extent to which mastery-based grading policies and practices in computer science courses affects the attitudes and beliefs of students. The results may be presented to others and published in scholarly publications.
2. **Procedures:** By participating in this study you agree to allow the study investigators to access any of the work (and associated grades) that you complete as part of your regular course participation. This consent form applies only to this specific course to be used for research purposes, not for any other course. Participating in this study does not require any of your time beyond your normal participation in the course. Your course instructor will not know whether you agree to participate until after you have completed the course and final grades have been turned in. You will earn assignment credit for submitting this survey, whether or not you agree to participate in the study or decline to answer some or all of the questions.
3. **Discomforts and Risks:** There are no known risks in participating in this research beyond those typically experienced as a student in a college course.
4. **Benefits:** The results may be used to improve the instruction in this course and/or other courses in the future.
5. **Duration:** Your participation will require no more of your time than what is asked of normal classroom activities.
6. **Statement of Confidentiality:** The researchers will not share any identifying information with anyone outside of the research team. In published material or presentations, no personal data will be included. Only the study investigators will have access to your identifying information.
7. **Right to Ask Questions:** You may ask questions about this research by contacting Stephen Edwards (540-353-6850). In addition, for questions about your rights as a research participant, you may contact the Virginia Tech Institutional Review Board at irb@vt.edu or (540) 231-3732.
8. **Compensation:** There is no compensation for your participation.
9. **Voluntary Participation:** Your decision to participate in this research is voluntary, and whether or not you participate will not affect your grade in this course in any way. You can stop at any time.
10. **Contact Information:** Principal Investigator: Dr. Stephen Edwards, Email: edwards@cs.vt.edu, Phone: 540-353-6850; Co-Investigator: Dr. Bob Edmison, Email: bedmison@vt.edu, Phone: 540-231-3957.

Participant's Permission: You must be 18 years or older to participate in this study. If are 18 years or older and you agree to participate in this research study, please select "Yes I give consent" below. If you

do not want to participate, please select "I would like to have my data excluded from the broader study of student attitudes and beliefs."

Consent to Participate: Do you give consent to participate in this study?

☐ Yes, I give consent
☐ I would like to have my data excluded from the broader study of student attitudes and beliefs

Appendix C

Interview Recruitment Email

Hi,

As a student in this course, you are being invited to participate in a research interview about your experiences with the course's grading policies and practices. This interview is part of a study titled "Collaborative Research: Transforming Grading Practices in the CS Education" and the purpose is to assess the impacts of mastery-based grading practices on students. A more complete description of the study is provided at the beginning of the interview.

At this time, we're asking you to let us know if you would be willing to participate in this interview with a member of our research team (not your instructor). Your participation will require you to talk with the interviewer for about 50-60 minutes. A light snack will be available. If you agree, we will schedule a time for the interview that is convenient for you.

If you have any questions, please contact me or another one of the project investigators listed here:

Dr. Stephen Edwards, Email: edwards@cs.vt.edu, Phone: 540-353-6850;

Dr. Bob Edmison, Email: bedmison@vt.edu, Phone: 540-231-3957.

Thank you for your consideration!

Appendix D

Interview Sign Up Google Form

Study Sign up form - Equitable grading practices

Hi,

As a student in this course, you are being invited to participate in a research interview about your experiences with the course's grading policies and practices. This interview is part of a study titled "Collaborative Research: Transforming Grading Practices in the CS Education" and the purpose is to assess the impacts of mastery-based grading practices on students. A more complete description of the study is provided at the beginning of the interview.

At this time, we're asking you to let us know if you would be willing to participate in this interview with a member of our research team (not your instructor). Your participation will require you to talk with the interviewer for about 50-60 minutes. A light snack will be available. If you agree, we will schedule a time for the interview that is convenient for you.

If you have any questions, please contact me or another one of the project investigators listed here:

Dr. Stephen Edwards, Email: edwards@cs.vt.edu, Phone: 540-353-6850;

Dr. Bob Edmison, Email: bedmison@vt.edu, Phone: 540-231-3957.

Thank you for your consideration!

* Indicates required question

1. Email *

2. Would you like to participate in this research study? *

Mark only one oval.

- Yes
- No

Skip to question 3

Select Time slot

Please select up to 3 preferred time slots that work the best for you. Please note that slots will be allotted on a first come-first serve basis.

Please use the following Google calendar appointment scheduler to book a slot:
Book time with Shukla, Anmol: Equitable Grading Practices - User study

Interview length is 50 mins - 1 hour.

3. Which date and time slot did you book? (MM/DD - HH:MM)

Note

You will be contacted with a confirmation of your RSVP via email.

Appendix E

Interview Script

Equitable Grading Practices Interview Script

Details

Interview type: One-on-one with each student
Interview duration: 50-60 minutes
Topics to be discussed: Below
Time period: 2/3 – 3/4 of the way through the semester

[Purpose]

We are conducting an educational research project on equitable grading practices.

Participation in this interview is purely voluntary—you will not receive any course credit, and your comments will not affect any course grade in any way. For completion of the interview, you will be offered some snacks that you are free to have.

While we hold our discussion, I'll take notes on the comments that are made. However, I will not use any names or write down any personally identifying information, so everything is kept anonymous.
Please read the consent form completely and decide if you wish to participate.

If you don't wish to participate, that is no problem—you are free to leave for any reason you choose, at any time. Also, if you are a minor, we cannot include you in our research, so please decline to participate now.
[Give time to participants to respond, or for participants to elect to leave.]

Script

Topics

1. Personal History, experiences and perception with respect to grading
    a. What are your views on grading?
        i. How reliable a measure is it to determine skill?
        ii. How reliable a measure is it to determine knowledge?
        iii. In your experience, if you've gotten a good grade in a course, has that correlated with you gaining knowledge/skill too?
    b. How much of your time/effort is spent on mastering a skill/gaining relevant knowledge vs getting the best grade

    c. Have you been graded in a course using a specification grading (EMRN) system before? If so, which course and time period?
    d. How much of a motivating factor is getting a good grade for you?
    e. What is the grade that you strive to get when you start out with a course?
2. EMRN grading system (or other fixed-value mastery-based grading system)
    a. How clear or convoluted did the EMRN grading system make the learning outcomes of this course?
    b. Did you feel stress/relaxation when you were informed about the changed grading policy for this course?
    c. How often did you have questions about any or all of the specifications, and how often did you have to reach out to the professor(s)/TA(s) to understand better?
    d. Which of E, M, R or N did you find yourself gravitating towards achieving, as the semester went on?
3. Late policy
    a. Did the late policy induce anxiety or help in relieving it?
        i. What was the degree of this feeling?
    b. How did your relationship and opinions with the late policy change as the semester went on?
    c. Decision making
        i. How did the late policy affect your decision making in terms of your submission timeline
        ii. How did the late policy affect your decision making in terms of your submission quality
    d. When using the late policy, did you use the extra time on the assignment or as buffer time to relax the intensity?
    e. In your opinion, did the late policy help you be on time for submissions in other courses?
4. Smaller quizzes vs bigger exams
    a. What do you find more challenging, 10 smaller assignments or 3 big ones spread out over the semester?