Examining Student Performance and Reception towards Interactive Feedback in Formative Assessments

Ishaq Chowdhury

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2025

Abstract

Examining Student Performance and Reception towards Interactive Feedback in Formative Assessments

Ishaq Chowdhury

Formative assessments are indispensable tools utilized to evaluate student learning trajectories, facilitate continuous improvement, and deepen comprehension. Central to the efficacy of formative assessments is the provision of feedback, which serves as a catalyst for students' progression towards mastery of the subject matter. However, not all feedback is equally effective in improving learning goals and performance on assessments. Additionally, the effectiveness of certain feedback types might not consistently align with the preferences students have when engaging with feedback. The main purpose of this dissertation is to examine how interactive and non-interactive feedback influence student performance on online formative assessments. It also investigates how students engage and respond to feedback, along with what type of learner characteristics might influence student feedback-seeking behavior.

This study uses a Pre Test, an assessment with feedback based on the condition, and a Post Test to measure student performance and test duration, while using questionnaire surveys to measure learner characteristics and reception to the feedback. Results from the study indicate that interactive feedback can have a greater impact on improving student scores and reducing their ratings of cognitive strain. Although it does not reduce the duration to complete assessments and requires more time investment to use the feedback, student who were exposed to the interactive feedback also reported it to be more helpful than those who were exposed to the non-interactive feedback.

Table of Contents

List of Tables and Figures iv

Acknowledgments v

Introduction 1

Chapter 1: Literature Review 4

1.1 Types of Feedback 4

1.2 Effects on Performance and Preferences based on Types of Feedback 7

1.3 Theoretical Framework & Research Questions 17

Chapter 2: Pilot Study Methods 20

2.1 Participants 20

2.2 Materials & Procedures 20

2.3 Scoring & Coding 23

Chapter 3: Pilot Study Results 25

3.1 Learner Characteristics 25

3.2 Proportion Correct 26

3.3 Duration 31

3.4 Reception 35

Chapter 4: Pilot Study Conclusions & Directions 41

4.1 Pilot Study Conclusions 41

Chapter 5: Main Study Methods 45

5.1 Participants 45
5.2 Materials & Procedures 45
5.3 Scoring & Coding 48
Chapter 6: Main Study Results 50
6.1 Learner Characteristics 50
6.2 Proportion Correct 53
6.3 Duration 59
6.4 Reception 63
6.5 Cognitive Load 71
Chapter 7: Conclusion 75
7.1 Proportion Correct 75
7.2 Reception 77
7.3 Duration 91
7.4 Learner Characteristics 96
7.5 Cognitive Load 99
7.5 Limitations and Future Considerations 102
References 107
Appendix A 113
Appendix B 114
Appendix C 115
Appendix D 116
Appendix E 117
Appendix F 118

List of Tables and Figures

Table 1. Learner Characteristics Descriptives and Reliability 26

Table 2. Pre Test (Q) and Post Test (PQ) Question Item Difficulty 27

Table 3. Point-Biserial Correlations for Pre Test (Q) and Post Test (PQ) 28

Table 4. Test Score Descriptive Statistics by Condition 30

Table 5. Duration Descriptive Statistics by Condition 35

Table 6. Reception Descriptive statistics by condition 39

Figure 1: Feedback Interactivity Preferences by Condition 40

Table 7. Learner Characteristics Descriptives and Reliability 51

Table 8. Pre Test, Formative Feedback Test, and Post Test Question Item Difficulty 54

Table 9. Point-Biserial Correlations for Pre Test, Formative Feedback Test and Post Test 55

Table 10. Test Score Descriptive statistics by condition 58

Table 11. Duration Descriptive Statistics by Condition 62

Table 12. Reception Descriptive Statistics by Condition 66

Figure 2: Feedback Interactivity Preferences by Condition 70

Figure 3: Feedback Interactivity Preferences by Assessment Scores 71

Table 13. Cognitive Load Descriptive Statistics by Condition 72

Acknowledgments

I would like to acknowledge my advisor, Dr. John Black, along with the rest of the Cognition Lab at Teachers College for all their aid and guidance throughout the entire process. I would also like to acknowledge my friends and family for their support and comfort during this endeavor.

Additionally, text from multiple chapters of my dissertation was drawn directly from my first author manuscripts in preparation for publication.

Introduction

In the discourse surrounding educational assessment practices, formative assessments emerge as essential tools used to gauge student learning trajectories, encourage continuous improvement, and deepen comprehension of acquired knowledge. Central to the value of formative assessments is the inclusion of feedback, which serves as a catalyst for students' progression towards mastery of the subject matter. The significance of feedback mechanisms within formative assessment frameworks has prompted researchers to delve into the complexities of optimal design considerations.

The educational landscape highlights the role of feedback in shaping students' cognitive processes, metacognitive awareness, and self-regulatory strategies. As such, investigations into the nature, delivery, and impact of feedback mechanisms within formative assessment contexts have been a topic of interest in educational research. Scholars have investigated the multifaceted dimensions of feedback, including its content, timing, format, and mode of delivery. By investigating the relationship between feedback and learning outcomes, researchers endeavor to unveil insights that inform the refinement of instructional practices and the enhancement of student engagement and achievement. Experimental studies, longitudinal analyses, and qualitative inquiries converge to explain the differential effects of various feedback modalities on student motivation, self-efficacy, and academic performance. Through examination of feedback interventions across diverse educational settings and learner populations, researchers aim to distill evidence-based recommendations for educators and instructional designers, facilitating the implementation of pedagogical strategies that optimize student learning experiences.

Studies have examined a myriad of factors related to feedback, encompassing aspects such as the timeliness of delivery (Van der Kleij, 2015; Williams, 2012; DeLucenay, Conn, &

Corigliano, 2017; Sinha & Glass, 2015; Attali & Van der Kleij, 2017; Metcalfe, Kornell, & Finn, 2009), student attitudes (Ashford & Cummings, 1983; Bempechat et al., 1991; Brandl, 1995; Butler, 1993; Rowe & Wood, 2008; VandeWalle *et al*., 2000; VandeWalle, 2003), and the typology of formative feedback (Attali & Van der Kleij, 2017; Collins, Carnine, & Gersten, 1987; Jaehnig & Miller 2007; Mandernach, 2005; Máñez et al., 2019; Mory, 2004; Petrovic, Pale, & Jeren, 2017; Slowaik & Lakowske, 2017; Smits et al, 2008; Van der Kleij et al., 2012; Van der Kleij, 2015). Elaborative Feedback (EF) emerges as a prominent focal point. However, the conceptual breadth of Elaborative Feedback poses a challenge, as it encompasses a spectrum of feedback types that aid students with supplementary instructional insights.

Given the diverse options for EF, there exists a rich terrain for scrutinizing and juxtaposing its various iterations. A facet of EF that has been under examination is Interactive Feedback, distinguished by its capacity to foster student engagement within formative assessment contexts, thereby potentially facilitating deeper comprehension of the subject matter (Crisp & Ward, 2008; Rowe & Wood, 2008). The interactive nature of this feedback modality holds promise for enhancing the learning experience by promoting active cognitive engagement and facilitating knowledge construction.

In the exploration of feedback efficacy, another pivotal facet that merits careful consideration is the subjective perception of students towards the feedback they receive. The way students perceive and interpret feedback is integral, as it may diverge from objective assessments of its value and effectiveness, consequently exerting an influence on their learning trajectories. The complexity of students' perceptions is further compounded by the design attributes inherent in the feedback mechanism itself, emphasizing the relationship between instructional design and student receptivity.

The following review endeavors to highlight the distinct relationship between feedback, student learning outcomes, and the perceptual dimensions of feedback engagement. Recognizing the intricate dynamics at play, the study aims to delineate the disparities in student learning outcomes among cohorts exposed to different types of feedback mechanisms. Additionally, the investigation seeks to unravel the pathways through which students' perceptions of feedback—shaped by its design attributes and contextual variations—may modulate their engagement with the feedback process. Central to the study's objectives is the interrogation of how students' subjective interpretations of feedback quality, relevance, and utility intersect with their cognitive and affective responses, thereby mediating the translation of feedback into tangible learning gains. Moreover, the study will examine the diverse array of factors—individual differences, motivational orientations, testing behavior—that underpin students' feedback perceptions and responses.

Chapter 1: Literature Review

1.1 Types of Feedback

Feedback interventions constitute a foundational mechanism by which external agents provide learners with information pertaining to various facets of task performance (Ammons, 1956). At the heart of feedback provision lies the delineation of distinct modalities, with Knowledge of Result (KR) Feedback emerging as the most basic form characterized by the binary revelation of correctness or incorrectness (Schute, 2008). For instance, within the context of a multiple-choice assessment equipped with KR Feedback, students are apprised solely of the accuracy or inaccuracy of their responses, devoid of elucidation regarding the correct option.

Knowledge of Correct Response (KCR) Feedback follows as a level of complexity above KR. In the paradigm of KCR Feedback, students are not only provided with the correct answer but are also afforded the right answer to the question (Schute, 2008). Consequently, students receive dual-layered feedback with KR Feedback revealing the alignment of their responses, coupled with the provision of the accurate answer.

Elaborative Feedback (EF) emerges as a comprehensive construct encompassing feedback modalities that is more complex or provides more information than KR or KCR Feedback (Schute, 2008). Distinguished by its multifaceted nature, EF transcends conventional feedback paradigms by providing learners with nuanced insights and instructional scaffolding aimed at augmenting comprehension and performance. The expansive scope of EF manifests through a diverse array of formats and mechanisms catered to the factors of learners' cognitive processes and learning objectives.

Typically, EF manifests in some form of explanation aimed at enriching students' understanding of the subject matter and explaining the pathways to arrive at the correct response.

For instance, an EF intervention may encompass a succinct narrative demonstrating the rationale behind the correct response, thereby illuminating the underlying principles and conceptual frameworks (Kulhavy & Stock, 1989). Moreover, EF can also include justification as to why each of the incorrect responses fail to answer the question properly.

Furthermore, EF encompasses a spectrum of interventions ranging from supplementary resources to customized instructional materials aimed at augmenting learning outcomes. For instance, learners may be directed towards supplementary study materials—ranging from general topic overviews to targeted instructional modules—designed to address gaps in comprehension and fortify foundational knowledge. This is called topic-contingent EF (Schute, 2008). EF interventions may adopt multimedia formats, including interactive simulations, instructional videos, or textual resources, to cater to diverse learning preferences and modalities. In alignment with instructional objectives, EF interventions may be calibrated to align with the specific nature of the question posed. For instance, learners may be directed towards instructional resources delineating the procedural steps required to solve questions of a particular type, thereby fostering skill acquisition and metacognitive awareness. This goes further in providing the student with feedback that is catered towards the question's specific nature, by using an isometric question as the example. The feedback serves as a demonstration for the student to turn to use as a guide to answer similar questions of that manner moving forward.

A similar variant of EF manifests in the form of worked solutions, characterized by their detailed, step-by-step explanation of problem-solving processes tailored to the exact problem that the student answered incorrectly (Glogger-Frey et al., 2015). This distinctive form of feedback not only provides learners with a model solution for the exact problem they answered incorrectly but also facilitates a direct comparison between their own process and the correct procedure, thereby enabling pinpoint identification of errors.

At the zenith of personalized feedback experiences lies the customization of worked solutions to address the specific errors exhibited by individual students during problem-solving endeavors. This approach, referred to as response-contingent EF, entails the generation of multiple iterations of feedback, each catered to rectify errors encountered at different junctures or stemming from distinct conceptual misunderstandings (Schute, 2008). For instance, in the context of multiple-choice questions, each incorrect option may be accompanied by a written explanation explaining the rationale behind its inadequacy, thereby enabling students to discern the differences of correct versus erroneous responses. Similarly, in scenarios necessitating multi-step problem-solving processes, worked solutions offer a dynamic avenue for pinpointing errors and guiding students towards corrective action. By identifying the precise juncture at which errors occurred, worked solutions can provide students with targeted instructional guidance aimed at correcting conceptual misunderstandings and reinforcing procedural fluency. Notably, this adaptive approach to feedback provisioning necessitates a comprehensive framework for the dynamic delivery of customized feedback experiences.

An alternative manifestation of worked solutions is encapsulated in the concept of partial-worked solutions, which imbue learners with agency in the problem-solving process by presenting incomplete step-by-step guides that necessitate active engagement and critical thinking (Manson & Ayres, 2021). In this format, learners are tasked with filling in the missing components of the solution pathway, thereby fostering deeper levels of cognitive engagement and metacognitive awareness. By bridging conceptual ideas and leveraging prior knowledge to navigate the gaps in the solution pathway, learners are empowered to construct their understanding iteratively, thereby fostering enduring comprehension and skill acquisition.

An additional dimension of importance pertains to the timing of feedback delivery. A burgeoning body of literature has highlighted the impact of temporal factors on the efficacy of feedback provision within assessment contexts, demonstrating the role of timing in shaping learning trajectories and fostering cognitive development. Immediate Feedback entails the prompt delivery of feedback to students immediately following their responses to assessment items (Hathalia et al., 2023). In contrast, Delayed Feedback includes a lag between students' responses and the subsequent provision of feedback (Williams, 2012). The period of Delayed Feedback can encompass a broad timespan, ranging from immediate post-assessment feedback to protracted delays spanning days or weeks. For instance, Delayed Feedback may be provided upon completion of discrete segments of the assessment, or aggregated to coincide with the culmination of the assessment in its entirety. Alternatively, feedback may be deferred to a later period of time, such as a day or a week after the completion of the assessment.

The temporal dimension of feedback delivery engenders considerations that extend beyond the immediacy of corrective guidance. Immediate Feedback, characterized by its timely provision, offers students the opportunity for real-time error correction and conceptual reinforcement, thereby fostering immediate comprehension and skill acquisition. Conversely, Delayed Feedback introduces a temporal buffer that affords students the opportunity for reflective processing, enabling deeper cognitive engagement and metacognitive reflection.

1.2 Effects on Performance and Preferences based on Types of Feedback

Numerous studies have probed the efficacy of diverse feedback modalities within educational contexts (Van der Kleij et al., 2015; Collins, Carnine, & Gersten, 1987; Corbalan, Kester, & Van Merriënboer, 2009; Corbalan, Paas, & Cuypers, 2010; Epstein, 1997; Jaehnig & Miller, 2007; Kim & Phillips, 1991; Kornell & Metcalfe, 2014; Huelser & Metcalfe, 2012), revealing insights into the dynamics of feedback provision and its impact on student learning outcomes. Of particular significance is the seminal meta-analysis that was conducted by Van der Kleij et al. (2015), which synthesized findings from 40 studies examining the differential effects of feedback mechanisms employed in computer-based assessments. The meta-analysis indicated that Elaborative Feedback was more effective than Knowledge of Result (KR) and Knowledge of Correct Response (KCR) feedback, especially in fostering higher-order learning outcomes. By providing learners with comprehensive, instructional insights and scaffolding, Elaborative Feedback emerges as a potent catalyst for cognitive engagement and knowledge construction within assessment contexts.

Furthermore, the meta-analysis delved into the impact of feedback timing, revealing patterns in the efficacy of immediate versus delayed feedback across varying learning outcomes. Consistent with prior research, the findings emphasize the superiority of immediate feedback in bolstering lower-order learning outcomes. However, the meta-analysis also unveiled noteworthy limitations inherent in the extant literature, including the prevalence of studies characterized by modest sample sizes and a scarcity of investigations exploring additional variables such as multimedia feedback, school level, and long-term effects. This dearth of empirical inquiry highlights the need for future research endeavors aimed at examining the interplay between these variables and feedback efficacy within diverse educational contexts.

Nevertheless, the findings of the meta-analysis demonstrate the transformative potential of Elaborative Feedback and immediate feedback interventions in enhancing student learning outcomes within online assessment frameworks. By attending to the effective integration of these feedback modalities, educators can cultivate a learning environment conducive to iterative improvement and cognitive advancement. Ultimately, the efficacy of feedback interventions hinges on learners' attentiveness and receptivity, with those who actively engage with effective feedback poised to reap the benefits of their inclusion in the learning process.

A divergent perspective emerges in the scholarly discourse, challenging the assertions put forth by the previous research, as evidenced by studies suggesting the absence of significant disparities among various feedback mechanisms in online assessment contexts (Mandernach, 2005; Slowaik & Lakowske, 2017; Van der Kleij et al., 2012). However, these divergent findings offer valuable insights into the multifaceted dynamics of feedback provision and its impact on student engagement and preference within educational settings.

Of particular significance are studies such as that conducted by Mandernach (2005), which interrogate the differential effects of diverse feedback conditions on student learning outcomes within computer-based online assessments. Mandernach's investigation encompassed an array of feedback modalities, ranging from KR and KCR feedback to topic-contingent and response-contingent feedback. Moreover, the study examined the synergistic effects of computer-generated feedback and feedback from independent online tutors, thereby affording a comprehensive appraisal of feedback mechanisms. In the study, 210 students were administered online assessments, with each assigned to a distinct feedback condition. Subsequent to the assessment, students in certain conditions were afforded the opportunity for review sessions with online tutors. Despite the absence of statistically significant differences in performance improvement across feedback conditions, Post Test satisfaction surveys revealed students found feedback more helpful if it was KCR or response-contingent feedback, expressing a heightened sense of satisfaction and perceived utility in comparison to counterparts who received no feedback, KR feedback, or topic-contingent feedback. These findings demonstrate that the perceived helpfulness and motivational value of feedback modalities can differ regardless of whether it improves their performance outcomes.

In further exploration of feedback preferences among students, Slowaik and Lakowske (2017) conducted an experimental inquiry involving 36 undergraduate and graduate students, aiming to discern their inclinations towards various feedback patterns. The experiment had participants engaging in a task simulating medical data entry across multiple sessions, with feedback conditions systematically manipulated across treatment groups. Specifically, participants received feedback structured according to three distinct patterns: PCP (positive, corrective, positive), PPC (positive, positive, corrective), or CPP (corrective, positive, positive). Despite the absence of statistically significant differences in performance outcomes across the feedback conditions, post-study questionnaires revealed that the majority of participants expressed a preference for the CPP feedback structure over alternative configurations, suggesting a preference for corrective feedback preceded by positive reinforcement. Furthermore, a prevailing preference for positive feedback over corrective feedback emerged, highlighting the motivational significance of feedback valence in influencing student engagement and receptivity.

In a complementary vein, Van der Kleij et al. (2012) conducted a study that examined the effects of written feedback within a computer-based assessment framework among 152 Commercial Economics undergraduate students. The study administered pretests and Post Tests to gauge student performance improvements, with feedback conditions systematically manipulated across treatment groups. Participants were exposed to immediate KCR feedback coupled with EF, delayed KCR+EF feedback, or delayed KR feedback. While no statistically significant differences in student achievement were discerned across feedback conditions, participant preferences and perceptions indicated that participants had a preference for immediate feedback over delayed feedback modalities, with immediate KCR+EF feedback perceived as more beneficial compared to delayed alternatives. This finding highlights the instrumental role of temporal immediacy in shaping feedback efficacy and student receptivity within educational settings.

Furthermore, the convergence of findings from Slowaik and Lakowske (2017) and Van der Kleij et al. (2012) emphasizes the difference between feedback preferences and learning outcomes. While feedback mechanisms may not invariably yield differential effects on performance outcomes, the identification of student preferences for specific feedback structures and valences underscores the salience of tailoring feedback interventions to align with individual learner needs and preferences. By embracing a nuanced understanding of student feedback preferences, educators can foster a collaborative learning environment conducive to enhanced motivation, engagement, and cognitive advancement.

In alignment with prior investigations, Smits et al. (2008) contributed to the discourse surrounding feedback efficacy by delving into the differential effects of Global Feedback and Elaborative Feedback modalities on student learning outcomes within a secondary school context. The study sample comprised 156 Dutch students in their fifth year of secondary education, who engaged in tasks involving genetics problems accompanied by feedback interventions. Participants were assigned to either the Global Feedback condition, characterized by a blend of KCR feedback and a systematic problem-solving approach, or the Elaborative Feedback condition, incorporating KCR feedback, systematic problem-solving strategies, fully worked-out solution steps, and explanatory rationales for correct answers. Additionally, participants received immediate feedback following each task and delayed feedback after every alternate task.

Upon analysis, the findings revealed insights into the effects of feedback modalities on student learning outcomes, contingent upon the participants' prior knowledge levels. Specifically, students with high prior knowledge exhibited greater improvement under the Global Feedback condition, emphasizing the efficacy of streamlined feedback interventions in fostering performance gains among cognitively adept learners. However, data gleaned from Post Test questionnaires indicated a preference among students for the Elaborative Feedback condition, despite its lesser effectiveness in promoting performance enhancements. This phenomenon demonstrates the complex relation between feedback preferences, cognitive load considerations, and learning outcomes, explaining the intrinsic tensions inherent in feedback provision within educational settings.

Despite its diminished efficacy in performance improvement, the observed preference for Elaborative Feedback among students may be attributed to its perceived utility in alleviating cognitive load and minimizing the cognitive effort required for comprehension. Students with lower prior knowledge levels may find the explanatory rationales inherent in Elaborative Feedback particularly beneficial, as it obviates the need for independent reasoning and facilitates comprehension without taxing cognitive resources excessively. Conversely, the Global Feedback modality, while more effective in fostering performance gains, may impose a greater cognitive burden due to its emphasis on independent problem-solving and critical reasoning skills.

Further corroborating evidence from Máñez et al. (2019) reaffirms the prevalence of student preferences for simpler feedback modalities over more elaborate alternatives. In their investigation, 52 eighth-grade students were tasked to complete a reading comprehension assessment on an online software program that included KR, KCR, and elaborative feedback. Máñez et al. observed a pronounced inclination towards KR and KCR feedback modalities, indicative of students' heightened attention to response accuracy and goal attainment. This predilection for simplicity over comprehensiveness highlights the primacy of performance outcomes in driving student engagement with feedback mechanisms, often to the detriment of deeper comprehension and metacognitive awareness. Students are simply motivated to reach the answer in the easiest way possible, without taking into consideration whether it will help them improve their understanding of the material.

In the pursuit of investigating the dynamics of student-feedback behavior, many studies have endeavored to discern the influence of various variables, including student traits such as achievement levels, on feedback preferences and engagement. Among these endeavors, Brandl's (1995) investigation stands as a notable contribution, shedding light on the differential feedback-seeking behaviors exhibited by high- and low-achievement students within the context of German language learning. Brandl's study enlisted the participation of twenty-one intermediate-level undergraduate students enrolled in German language courses, who underwent an achievement assessment to ascertain their academic proficiency levels. Subsequently, participants engaged in a grammatical task involving the conversion of German sentences from active to passive voice, wherein they received feedback interventions tailored to different facets of error correction. Specifically, participants were presented with options to access KR feedback, KCR feedback, feedback highlighting the location of errors in the sentence, or elaborative feedback offering grammatical hints pertinent to error rectification.

The findings gleaned from Brandl's study revealed that high-achievement students demonstrated a preference for accessing KR feedback, indicative of their propensity for independent error correction and solution attainment. Conversely, low-achievement students exhibited a heightened preference for KCR feedback, reflecting their inclination towards accessing correct answers with minimal cognitive effort. This dichotomy in feedback-seeking behaviors demonstrates the relationship between achievement levels and motivational factors in shaping student engagement with feedback mechanisms. High-achievement students' preference for independent problem-solving and solution attainment aligns with the findings of previous investigations, echoing the propensity for simpler feedback modalities among students motivated by performance outcomes.

The findings of Brandl's study align with prior research by Smits et al. (2008) and Máñez et al. (2019), which underscored the prevalence of student preferences for feedback modalities imposing lesser cognitive load, irrespective of their efficacy in fostering comprehension and mastery. Notably, these findings converge with the assertions put forth by Cohen (1987), who elucidated the pivotal role of achievement levels in shaping students' motivation and engagement with feedback mechanisms.

Alternatively, Bempechat et al. (1991) investigates how students' implicit theories of intelligence play a role in the feedback they prefer, yet the results are also in line with the previous findings. The study enlisted the participation of thirty-six 5th-6th grade students, who underwent an initial orientation session aimed at instilling either an entity orientation (fixed-mindset) or an incremental orientation (growth-mindset) regarding intelligence. Drawing on the theoretical framework proposed by Dweck and Bempechat (1983), which posits that children harbor implicit beliefs regarding the malleability of intelligence—viewing it as either fixed and immutable or as capable of improvement through effort—the study aimed to discern the differential effects of these orientations on students' goal-setting behaviors and feedback preferences. Participants were subsequently afforded the opportunity to select their goal for the task at hand, with options delineated into learning-oriented goals involving more challenging tasks aimed at increasing competence and performance-oriented goals entailing easier tasks designed to showcase existing competence. Subsequently, participants engaged in matrix problem-solving tasks, with certain problems deliberately rendered insoluble to manipulate controlled failure.

The findings derived from the study yielded insights into the interplay between students' implicit theories of intelligence, goal-setting behaviors, and feedback preferences. Specifically, students indoctrinated with the entity theory of intelligence demonstrated a preference for learning-oriented goals, whereas those imbued with the incremental theory of intelligence evinced a heightened preference for performance-oriented goals. This dichotomy in goal-setting behaviors suggests that students holding beliefs in fixed intelligence exhibit a greater motivation to engage in tasks that affirm their existing abilities rather than challenge them to strive for improvement.

This observed propensity for self-affirming tasks among students embracing fixed intelligence beliefs aligns with the proclivity of many low-achievement students to gravitate towards KR and KCR feedback modalities over more elaborate feedback options conducive to cognitive growth and comprehension enhancement. Such findings resonate with the assertions put forth by Bandura and Dweck (1985), stressing the pivotal role of belief systems and self-perceptions in shaping students' motivational orientations and engagement with feedback mechanisms.

An examination of the existing literature reveals a notable gap in our understanding of various factors influencing student performance and perceptions regarding feedback within the context of formative assessments. While elaborative feedback emerges as a predominant method for enhancing student learning outcomes, its effectiveness is contingent upon numerous dimensions that often yield conflicting results. Moreover, research on the timing of feedback delivery demonstrates the efficacy of both immediate and delayed feedback in disparate educational settings, with divergent findings contingent upon the duration of delayed feedback periods.

Effective feedback mechanisms aim to achieve two primary objectives: augmenting student engagement with feedback and mitigating the cognitive load and effort required for feedback processing. Partially-worked solutions, characterized by their interactive nature and step-by-step guidance, offer a promising avenue for delivering elaborative feedback. By enabling students to interact directly with feedback components, such solutions facilitate active engagement and comprehension, thereby reducing cognitive strain.

Furthermore, within the context of formative assessments, immediate feedback emerges as a valuable strategy for fostering skill acquisition and comprehension. Immediate feedback provisions afford students the opportunity to practice material iteratively while receiving real-time guidance on error resolution and application of appropriate strategies. Through consistent reinforcement and error correction, immediate feedback serves as a catalyst for enhancing student comprehension and mastery of subject matter.

In summary, while the literature underscores the efficacy of elaborative feedback and immediate feedback in formative assessment contexts, significant gaps remain in our understanding of optimal feedback delivery methods and their impact on student learning outcomes. Future research endeavors should aim to explore the dynamics of feedback mechanisms, thereby informing the development of evidence-based instructional strategies conducive to fostering student engagement, comprehension, and skill acquisition within educational settings.

1.3 Theoretical Framework & Research Questions

The primary aim of this study is to examine the influence of interactive elaborative feedback on student performance and engagement. By incorporating partially-worked examples, the study seeks to demonstrate the effectiveness of elaborative feedback in encouraging student engagement and learning. Partially-worked examples serve as a form of scaffolding by providing students with initial steps, thereby enabling them to continue problem-solving independently. This approach is designed to mitigate cognitive load by outlining the procedural steps, facilitating the development of a systematic problem-solving strategy. Additionally, scaffolding as feedback has be found to be more effective for long-term retention compared to other forms of feedback as well (Finn & Metcalfe, 2010). This study investigates several key research questions to explore the relationship between interactive elaborative feedback and various aspects of student performance and perception.

Research Question 1

The first research question is in regards to student performance, and posits that students receiving interactive feedback will exhibit superior performance on Post Tests compared to those receiving no interaction. The rationale is that increased interactivity enhances engagement with the material, fostering a deeper understanding of the mathematical concepts. This research question will be tested by analyzing the differences in Pre Test and Post Test scores across different feedback conditions.

Research Question 2

The second research question concerns student reception, suggesting that participants will respond more positively to the interactive feedback compared to no-interaction. It is hypothesized that no interaction fails to sufficiently engage students, while very high interaction imposes excessive cognitive demands, potentially leading to increased stress. Therefore, an intermediate level of interaction is expected to strike an optimal balance, resulting in the highest satisfaction levels. This research question will be evaluated through post-questionnaire data, focusing on students' ratings of the feedback received.

Research Question 3

The third research question addresses the duration of task completion, proposing that students with interactive feedback will spend less time on Post Tests than those with no-interaction. Enhanced interactive feedback is anticipated to equip students with a stronger grasp of the material, thus reducing the time required to complete subsequent tasks. This research question will be assessed by comparing the time spent on Post Tests across the different feedback conditions.

Research Question 4

The fourth research question examines the influence of participant characteristics on engagement with feedback. It is posited that students with higher need for achievement, intrinsic value in mathematics, growth mindset, or self-efficacy will engage more thoroughly with the feedback provided. This is because these students may have a greater motivation to learn the material and to improve their understanding of the topic, which would in-turn encourage a greater attention to the feedback provided in the study. This research question will be investigated by correlating pre-questionnaire data on learner characteristics with the duration of time spent on feedback pages during the feedback phase.

This study aims to provide comprehensive insights into the effects of interactive elaborative feedback on student learning. By examining performance, perception, time-on-task, and participant characteristics, the research seeks to identify optimal feedback strategies that enhance educational outcomes. The findings are expected to contribute to the development of more effective instructional techniques, ultimately improving student engagement and achievement in educational settings.

Chapter 2: Pilot Study Methods

2.1 Participants

A total of 36 graduate students from Columbia University – Teachers College were recruited for this study. Participants were recruited through flyers posted in online course pages for Psychology classes. The flier included information regarding the study and contact information of the principal investigator for the potential participants to schedule a timeslot to participate in the study. Participants were eligible to participate if they were a student of Teachers College and are unfamiliar with the process of determining area/perimeter of Recursive Fractals. Participants were awarded two credits of research participation, which may be required or used as extra credit for certain classes.

2.2 Materials & Procedures

Participants emailed the primary investigator to schedule a 90-minute time slot to participate in the study. The study took place in a lab at Teachers College - Columbia University. Participants were seated at a desk with a desktop computer and provided with a pen, two sheets of blank paper, and basic pocket calculator. Participants were informed that they could use the provided materials to work out problems. Participants were asked to avoid talking to other participants in the room, engaging in any distractions from their phone, or opening other tabs on the computer. Once they were given their instructions, participants were sent a link to start the study on *Qualtrics*.

Participants were first asked if they would like to consent to take part in the study, which would include a series of self-paced questionnaires and assessments. Upon agreeing, participants were asked to complete a 2-item demographics survey inquiring about their gender and race/ethnicity. Following this, participants were then asked to complete a 20-item questionnaire regarding learner characteristics in which they chose how much they agree with various statements (see Appendix A). The choices consist of 5-point Likert scales ranging from strongly disagree to strongly agree. There are four statements for each of the following categories: Implicit Theory of Intelligence (IT), Self-Efficacy (SE), Intrinsic Value of Math (IV), Test Anxiety (TA), and Need for Achievement (NA). At least one question in each category were reverse-coded. The statements are presented in a randomized order five items at a time across four pages.

After the questionnaire, participants were given a short introduction to view that included text and visuals covering the concept of fractal shapes. Once they have seen the introduction, they will move on to the Acquisition Phase of the study. They were presented with two short mini-lessons about finding the area and finding the perimeter of fractal shapes. Each mini-lesson utilized a model shape to serve as an example to teach the participant how to use the general formula. The lesson included definitions of each variable and a sample worked-solution to demonstrate the process of using the formula (see Appendix C). They were allowed to spend up to 5 minutes reviewing each lesson, and would not be able to return to the lesson page after that section.

The next step is the Formative Feedback Phase. In this section, participants were tasked with solving a series of 10 multiple-choice questions. Each question has 5 choices and participants were allowed to use the calculator provided and to use the scrap paper. All students received the questions in the same order, with the first five questions related to area and last five questions related to perimeter. For the five questions from each topic, the first two questions use the same shape as the example provided in the mini-lesson, while the next two questions use a slight variation of the shape, and the final question use a different shape (see Appendix B). All of the questions were solvable using the formulas provided in the mini-lessons. All ten questions were equivalent in format, with different numbers used for each of the variables.

Participants in the Non-Interaction Condition were given Knowledge of Result (KR) + Knowledge of Correct Response (KCR) feedback, along with a fully-worked solution as a step-by-step guide demonstrating how to solve the question they had just answered (see Appendix E). Participants in the High-Interaction Condition were given a partial-worked solution. In these cases, certain portions of the worked solution required the participant to make choices or provide input in the guide. This involved using multiple-choice format to ask participants which formula should be used and how the variables would be plugged into the formula. This also includes fill-in-the-blank questions asking the participant to type in what value to input for each variable of the formula (see Appendix D). Participants in the Low-Interaction Condition were given partial-worked solutions that only included the multiple choice questions and not the fill-in-the-blank questions from the High-Interaction Condition. Participants were given the feedback regardless of whether they answered the questions correctly or incorrectly. Participants were allowed to spend as much time as they want with the feedback before moving on to the next question. Participants in the control group were not given any feedback between each question.

After the Formative Feedback Phase, participants moved on to the Post Test Phase. In this section they were tasked with answering a series of 10 more multiple-choice questions. These problems were equivalent in format to the set from the Formative Feedback Phase, however they were all shuffled in a randomized order for this assessment. All participants were given no feedback after each question.

Once they complete the assessment, all participants were given a Post Test Questionnaire. The first portion of the questionnaire asked participants how they felt about the difficulty of the topic using 5-point Likert scales and their perceived prior/current knowledge regarding the topic using 10-point Likert scales. It also included a 7 items related cognitive load using a 5-point Likert scale. The second portion of the questionnaire was only given to the participants who were in the non-interactive, low-interactive, or high interactive conditions and was related to their reception towards the feedback. The first item is a 5-point Likert scale asking them how helpful they found the explanation guides, the second item is a 5-point Likert scale asking them how difficult it was for them to use the explanation guides, and the third item is a multiple choice question asking them if they would have preferred the explanation guides to be "More Interactive", "Less Interactive", or "It is fine as it is".

After completing the Post Test questionnaire, they were asked to provide details in order to contact their professors to confirm their participation and award them the research credit. Participants were reminded that this information would not be used in the analysis for the study.

2.3 Scoring & Coding

A composite score was created for Implicit Theory of Intelligence (IT), Self-Efficacy (SE), Intrinsic Value of Math (IV), Test Anxiety (TA), and Need for Achievement (NA) by calculating each of the averages of their respective items from the Learner Characteristics Questionnaire. A composite score was created for Cognitive Load (CL) by calculating the average of seven items. A Formative Feedback Test score was created by grading the number of correct answers from the Formative Feedback phase out of ten. A Post Test score was created by grading the number of correct answers from the Post Test Phase out of ten. Each of the questions in the Formative Feedback Test and Post Test were weighted equally. A Delta score was calculated by subtracting the Formative Feedback Test score from the Post Test score. A Feedback Score was created for the participants in the Interactive condition based on the number of correct choices and inputs they made in their feedback. The Feedback Score was graded based on the two multiple choice questions weighted equally and the four fill-in-the-blank questions weighted equal to the two multiple choice questions.

Chapter 3: Pilot Study Results

There were 9 participants in each of the four conditions: Control, No-Interaction (NI), Low-Interaction (LI), and High-Interaction (HI). The participants were 63.9% female and 36.1% male and comprised 83.3% Asian/Pacific Islander, 13.9% White/Caucasian, and 2.8% Hispanic/Latino. SPSS software (Version 28.0.0) was used to perform all statistical analysis of the data.

3.1 Learner Characteristics

The composite scores from the four items representing each of the learner characteristics were out of a scale of five. Descriptive data and reliability analysis are reported in Table 1. Cronbach's Alphas were calculated in for each of the characteristics as follows: Implicit Theory of Intelligence ($\alpha = .870$), Self-Efficacy ($\alpha = .906$), Intrinsic Value of Math ($\alpha = .704$), Testing Anxiety ($\alpha = .720$), and Need for Achievement in Math ($\alpha = .622$). The mean score for Implicit Theory of Intelligence was 3.14 ($SD = 0.966$). The mean score for Self-Efficacy was 3.06 ($SD = 1.047$). The mean score for Intrinsic Value of Math was 3.71 ($SD = 0.798$). The mean score for Testing Anxiety was 3.28 ($SD = 0.780$). The mean score for Need for Achievement in Math was 3.85 ($SD = 0.713$).

Pearson correlation tests were conducted between each of the characteristics, Post Test Scores, and Test Score Deltas. These correlations were conducted to determine if there were any trends between the learner characteristics or any patterns related to the assessment scores. Self-Efficacy had a moderate positive correlation with Intrinsic Value of Math ($r = .588$, $p < .001$). Testing Anxiety had a weak positive correlation with Post Test scores ($r = .331$, $p = .048$). Need for Achievement had a weak positive correlation with Post Test scores ($r = .356$, $p = .033$).

Implicit Theory of Intelligence was not found to have a significant correlation with any of the other characteristics or test scores.

Table 1. Learner Characteristics Descriptives and Reliability

	Mean	SD	Cronbach's Alpha
Implicit Theory (IT)	3.14	0.966	0.870
Self-Efficacy (SE)	3.06	1.047	0.906
Intrinsic Value (IV)	3.71	0.798	0.704
Test Anxiety (TA)	3.28	0.780	0.720
Need for Achievement (NA)	3.85	0.713	0.622

Note: n=4 items for each category

3.2 Proportion Correct

Descriptive statistics of the Pre Test and Post Test question scores are reported in Table 2. The mean scores for each of the 10 multiple-choice question in the Pre Test are as follows: Q1 (m = 0.83, SD = 0.378), Q2 (m = 0.89, SD = 0.319), Q3 (m = 0.89, SD = 0.319), Q4 (m = 0.94, SD = 0.232), Q5 (m = 0.94, SD = 0.232), Q6 (m = 0.83, SD = 0.378), Q7 (m = 0.94, SD = 0.232), Q8 (m = 0.89, SD = 0.319), Q9 (m = 0.94, SD = 0.232), Q10 (m = 0.69, SD = 0.467). The mean scores for each of the 10 multiple-choice question in the Post Test are as follows: PQ1 (m = 0.94, SD = 0.232), PQ2 (m = 0.89, SD = 0.319), PQ3 (m = 0.92, SD = 0.280), PQ4 (m = 0.97, SD = 0.167), PQ5 (m = 0.89, SD = 0.319), PQ6 (m = 1.00, SD = 0.000), PQ7 (m = 0.92, SD = 0.280), PQ8 (m = 0.92, SD = 0.280), PQ9 (m = 0.94, SD = 0.232), PQ10 (m = 0.94, SD = 0.232). The mean score for the Pre Test was 88.1% (SD = 0.165) and the mean score for the Post Test was

93.3% (*SD* = 0.153). The mean difference between the Pre Test and the Post Test was an increase of 5.3% (*SD* = 0.116).

Table 2. Pre Test (Q) and Post Test (PQ) Question Item Difficulty

	Mean	SD		Mean	SD
Q1	0.83	0.378	**PQ1**	0.94	0.232
Q2	0.89	0.319	**PQ2**	0.89	0.319
Q3	0.89	0.319	**PQ3**	0.92	0.280
Q4	0.94	0.232	**PQ4**	0.97	0.167
Q5	0.94	0.232	**PQ5**	0.89	0.319
Q6	0.83	0.378	**PQ6**	1.00	0
Q7	0.94	0.232	**PQ7**	0.92	0.28
Q8	0.89	0.319	**PQ8**	0.92	0.28
Q9	0.94	0.232	**PQ9**	0.94	0.232
Q10	0.69	0.476	**PQ10**	0.94	0.232
Overall	0.881	0.165	**Overall**	0.933	0.153
			Post-Pre Delta	0.053	0.116

Note: Means indicate the accuracy rate of each question, with 1.00 representing 100% of the participants answering correctly

For each question, point-biserial correlation coefficients were calculated to assess the strength and direction of the relationship between correct/incorrect responses and the total test score (Table 3). The results for the Pre Test questions with the Pre Test scores are as follows: Q1 ($r = 0.678, p < 0.001$), Q2 ($r = 0.663, p < 0.001$), Q3 ($r = 0.663, p < 0.001$), Q4 ($r = 0.789, p < 0.001$), Q5 ($r = 0.417, p = 0.011$), Q6 ($r = 0.450, p = 0.006$), Q7 ($r = 0.641, p < 0.001$), Q8 ($r = 0.066, p = 0.701$), Q9 ($r = 0.566, p < 0.001$), Q10 ($r = 0.476, p = 0.003$). All of the item responses had a significant positive correlation with the total Pre Test score except for Question 8. The results for the Post Test questions with the Post Test scores are as follows: PQ1 ($r = 0.696, p < 0.001$), PQ2 ($r = 0.722, p < 0.001$), PQ3 ($r = 0.799, p < 0.001$), PQ4 ($r = 0.821, p < 0.001$), PQ5 ($r = 0.722, p < 0.001$), PQ7 ($r = 0.533, p < 0.001$), PQ8 ($r = 0.599, p < 0.001$), PQ9

($r = 0.295$, $p = 0.081$), PQ10 ($r = 0.696$, $p < 0.001$). The item-total score correlation for Question 6 could not be calculated since it was a constant in which all participants chose the correct response. All of the remaining item responses had a significant positive correlation with the total Post Test score except for Question 9. This indicates that each item is consistent with the overall scores for the respective assessments.

Table 3. Point-Biserial Correlations for Pre Test (Q) and Post Test (PQ)

	Correlation (*r*)	sig.		Correlation (*r*)	sig.
Q1	0.678	0.001**	**PQ1**	0.696	0.001**
Q2	0.663	0.001**	**PQ2**	0.722	0.001**
Q3	0.663	0.001**	**PQ3**	0.799	0.001**
Q4	0.789	0.001**	**PQ4**	0.821	0.001**
Q5	0.417	0.011*	**PQ5**	0.722	0.001**
Q6	0.45	0.006**	**PQ6**	N/A	N/A
Q7	0.641	0.001**	**PQ7**	0.533	0.001**
Q8	0.066	0.701	**PQ8**	0.599	0.001**
Q9	0.566	0.001**	**PQ9**	0.295	0.081
Q10	0.476	0.003**	**PQ10**	0.696	0.001**

* Correlation is significant at the .05 level (2-tailed)

** Correlation is significant at the .01 level (2-tailed)

The normality of the test scores were assessed using the Shapiro-Wilk test. It revealed that the Pre Test Scores ($w = 0.691$, $df = 36$, $p < .001$), the Post Test Scores ($w = 0.495$, $df = 36$, $p < .001$), and the Test Score Delta ($w = 0.858$, $df = 36$, $p < .001$) were not normally distributed. The homogeneity of variances were assessed using Levene's test. The results revealed that the Pre Test Scores did not have equal variances across the groups, $F(3, 32) = 5.144$, $p = .005$. However, Levene's test indicated that there were equal variances across the groups for the Post Test Scores, $F(3, 32) = 2.688$, $p = .063$. It also indicated that there were equal variances across groups for the Test Score Delta, $F(3, 32) = 1.410$, $p = .258$.

The Kruskal-Wallis test was conducted to compare the means of the Pre Test scores between each condition due to violations of the assumptions of equal variances and normal distribution. This test was conducted in order ensure that there were no significant differences between the two groups prior to any treatment condition, and that it would be reasonable to treat them as equal baselines. The Pre Test mean scores for each condition are as follows: Control (*m* = 0.789, *SD* = 0.276), No-Interaction (*m* = 0.911, *SD* = 0.093), Low-Interaction (*m* = 0.922, *SD* = 0.083), and High-Interaction (*m* = 0.900, *SD* = 0.123). The results revealed that there was not a significant difference between the high-interactive, low-interaction, and non-interactive conditions ($H(3) = 0.920$, $p = 0.821$).

A One-Way ANOVA was conducted to compare the means of the Post Test scores between each condition. The Post Test mean scores for each condition are as follows: Control (*m* = 0.867, *SD* = 0.260), No-Interaction (*m* = 0.978, *SD* = 0.067), Low-Interaction (*m* = 0.911, *SD* = 0.136), and High-Interaction (*m* = 0.978, *SD* = 0.044). The results revealed that there was not a significant difference between groups, $F(3) = 1.288$, $p = 0.295$.

A One-Way ANOVA was conducted to compare the means of the Test Score Deltas between each condition. This was conducted to see if there were any differences in which group may have improved the most, rather than solely looking at the final assessment scores alone. The Test Score Delta mean scores for each condition are as follows: Control (*m* = 0.078, *SD* = 0.139), No-Interaction (*m* = 0.067, *SD* = 0.112), Low-Interaction (*m* = -0.011, *SD* = 0.078), and High-Interaction (*m* = 0.078, *SD* = 0.120). The results revealed that there was not a significant difference between groups, $F(3) = 1.263$, $p = 0.304$.

Pearson correlation tests were conducted between the Pre Test scores, Post Test Scores, and Test Score Deltas. These correlations were conducted in order to see if there were trends between the baseline performances and later assessment outcomes and overall improvement. Pre Test scores had a strong positive correlation with Post Test scores ($r = .588$, $p < .001$) and moderate negative correlation with Test Score Deltas ($r = -.452$, $p = .006$).

Table 4. Test Score Descriptive Statistics by Condition

	Pre-Test Score				Post-Test Score				Post-Pre Delta	
	Means	**SD**	**Skew**	**Kurtosis**	**Means**	**SD**	**Skew**	**Kurtosis**	**Means**	**SD**
Control (n=9)	0.789	0.276	-1.505	1.611	0.867	0.260	-2.609	7.135	0.078	0.139
Worked Solution (n=9)	0.911	0.093	-1.470	3.281	0.978	0.067	-3.000	9.000	0.067	0.112
Interactive (Low) (n=9)	0.922	0.083	-0.501	-1.275	0.911	0.136	-1.771	3.033	-0.011	0.078
Interactive (High) (n=9)	0.900	0.123	-1.050	-0.286	0.978	0.044	-1.620	0.735	0.078	0.120
Overall	0.881	0.165	-2.484	7.748	0.933	0.153	-3.639	15.448	0.053	0.116

Note: n=36 participants, means represent average assessment scores on a scale of 0.00-1.00

Participants in the Low-Interaction and High-Interaction conditions were given feedback scores based on the proportion of correct answers in the feedback prompts. The mean score for the Low-Interaction feedback questions was .975 ($SD = 0.035$). The mean score for the High-Interaction feedback questions was .983 ($SD = 0.024$). The normality of the Feedback Score were assessed using the Shapiro-Wilk test. It revealed that the Low-Interaction Feedback Score ($w = 0.705$, $df = 18$, $p < .001$) and the High-Interaction Feedback Score ($w = 0.728$, $df = 9$, $p = .003$) were not normally distributed. The homogeneity of variances was assessed using Levene's test. The results revealed that there were equal variances across groups for the Low-Interaction Feedback Score, $F(1, 16) = 0.024$, $p = .880$. Only one group was given the High-Interaction

Feedback questions so there was no need to check for homogeneity of variances for the High-Interaction Feedback Score.

A One-Way ANOVA was conducted to compare the Low-Interaction Feedback Score between each condition. This was used to determine if the interactivity of the feedback may have impacted the proportion of correct responses participants had for the feedback prompts. The Low-Interaction Feedback Score for each condition are as follows: Low-Interaction (m = 0.972, SD = 0.036), and High-Interaction (m = 0.978, SD = 0.036). The results revealed that there was not a significant difference between groups, $F(1) = 0.105$, $p = 0.750$.

Pearson correlation tests were conducted between the Feedback Score, Post Test Scores, and Test Score Deltas. These correlations were conducted in order to determine if higher participants' scores in the feedback prompts were related to better performance on the assessments. Post Test scores had a moderate positive correlation with Low-Interaction Feedback Score ($r = .559$, $p = .016$) and strong positive correlation with High-Interaction Feedback Score ($r = .781$, $p = .013$). Additionally, Low-Interaction Feedback Score had a very high positive correlation with High-Interaction Feedback Score ($r = .948$, $p < .001$).

3.3 Duration

The normality of the durations was assessed using the Shapiro-Wilk test. The results revealed that Total Duration ($w = 0.933$, $df = 27$, $p = .083$), Acquisition Phase Duration ($w = 0.989$, $df = 27$, $p = .990$), Pre Test Duration ($w = 0.944$, $df = 27$, $p = .157$), and Pre-Post Test Duration Difference ($w = 0.938$, $df = 27$, $p = .107$) were all normally distributed. The test also revealed that Feedback Duration ($w = 0.802$, $df = 27$, $p < .001$) and Post Test Duration ($w = 0.860$, $df = 27$, $p = .002$) were not normally distributed.

The homogeneity of variances was assessed using Levene’s test. Levene's tests indicated that the assumption of homogeneity of variances was not violated for the following variables: Total Duration, $F(3, 32) = 0.902$, $p = .451$; Acquisition Phase Duration, $F(3, 32) = 0.540$, $p = .659$; Feedback Duration, $F(3, 32) = 1.930$, $p = .167$; Post Test Duration, $F(3, 32) = 1.227$, $p = .316$; and Pre-Post Test Duration Difference, $F(3, 32) = 0.150$, $p = .929$. Levene's tests indicated that the assumption of homogeneity of variances was violated for Pre Test Duration $F(3, 32) = 4.546$, $p = .009$.

A One-Way ANOVA was conducted to compare the means of the Total Duration between each condition. This was conducted in order to determine if participants would take less time to complete the study based on the type of feedback they were engaging with. The mean total duration of the experiment was 43.74 minutes ($SD = 12.03$). The Total Duration means for each condition are as follows: Control ($m = 41.42$, $SD = 8.866$), No-Interaction ($m = 42.24$, $SD = 10.708$), Low-Interaction ($m = 44.137$, $SD = 15.496$), and High-Interaction ($m = 47.067$, $SD = 13.364$). The results revealed that there was not a significant difference between groups, $F(3) = 0.361$, $p = 0.782$.

A One-Way ANOVA was conducted to compare the means of the Acquisition Phase Duration between each condition. This was conducted in order to confirm that participants across each group spent a similar amount of time viewing the acquisition material related to the content for the assessments. The mean duration participants spent on the Acquisition Phase was 4.80 minutes ($SD = 1.607$). The Acquisition Phase Duration means for each condition are as follows: Control ($m = 5.09$, $SD = 1.758$), No-Interaction ($m = 4.24$, $SD = 1.329$), Low-Interaction ($m = 5.24$, $SD = 1.401$), and High-Interaction ($m = 4.62$, $SD = 1.945$). The results revealed that there was not a significant difference between groups, $F(3) = 0.705$, $p = 0.556$.

A One-Way ANOVA was conducted to compare the means of the Feedback Duration between each treatment condition. This was conducted in order to determine if any of the feedback mechanisms require significantly more time to engage with compared to the other mechanisms. The mean of the total duration participants in the treatment conditions spent viewing the feedback pages was 4.01 minutes (*SD* = 2.986). The Feedback Duration means for each condition are as follows: No-Interaction (*m* = 2.05, *SD* = 1.224), Low-Interaction (*m* = 3.34, *SD* = 1.802), and High-Interaction (*m* = 6.62, *SD* = 3.427). The results revealed that there was a significant difference between groups, $F(2) = 9.086$, $p = 0.001$. A Tukey post-hoc test was conducted to assess specific group differences. The results indicated the High-Interaction condition had significantly higher mean time spent on the feedback compared to both the Low-Interaction condition ($p = .018$) and the Worked Solution condition ($p = .001$).

A One-Way ANOVA was conducted to compare the means of the Pre Test Duration between each condition. This was to confirm that participants between each condition did not differ in the amount of time they took to complete the initial assessment, which would reaffirm the basis of their comparison as equal baselines. The mean of the total duration participants spent on the Pre Test was 14.79 minutes (*SD* = 4.088). The Pre Test Duration means for each condition are as follows: Control (*m* = 15.32, *SD* = 1.699), No-Interaction (*m* = 14.77, *SD* = 3.414), Low-Interaction (*m* = 14.55, *SD* = 5.193), and High-Interaction (*m* = 14.51, *SD* = 5.581). The results revealed that there was not a significant difference between groups, $F(3) = 0.068$, $p = 0.977$.

A One-Way ANOVA was conducted to compare the means of the Post Test Duration between each condition. This was conducted in order to determine if the differences in feedback conditions would impact the amount of time participants needed to completed the final assessment. The mean of the total duration participants spent on the Post Test was 9.60 minutes

(*SD* = 3.762). The Post Test Duration means for each condition are as follows: Control (*m* = 11.40, *SD* = 4.643), No-Interaction (*m* = 9.40, *SD* = 2.854), Low-Interaction (*m* = 8.84, *SD* = 3.907), and High-Interaction (*m* = 8.74, *SD* = 3.434). The results revealed that there was not a significant difference between groups, $F(3) = 0.968$, $p = 0.420$.

A One-Way ANOVA was conducted to compare the means of the Pre-Post Test Duration Difference between each condition. This was to determine if there were any differences between the conditions regarding how much less time participants needed when comparing the amount of time they needed for the initial and final assessments. The mean difference of the total duration participants spent between the Pre Test and the Post Test was -5.19 minutes (*SD* = 3.513). The Pre-Post Test Duration Difference means for each condition are as follows: Control (*m* = -3.92, *SD* = 3.938), No-Interaction (*m* = -5.38, *SD* = 3.425), Low-Interaction (*m* = -5.71, *SD* = 4.042), and High-Interaction (*m* = -5.77, *SD* = 2.808). The results revealed that there was not a significant difference between groups, $F(3) = 0.526$, $p = 0.668$.

Pearson correlation tests were conducted between the Feedback durations, the Pre Test scores, Post Test Scores, and Test Score Deltas. This was to determine if there was consistency in how long participants spent to complete each phase. Total Feedback Duration had a moderate positive correlation with Test Score Deltas ($r = .411$, $p = .033$).

Table 5. Duration Descriptive Statistics by Condition

	Total Duration				Acquisition Phase Duration				Feedback Duration			
	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis
Control (n=9)	41.42	8.866	0.110	-0.971	5.09	1.758	1.401	0.424	N/A	N/A	N/A	N/A
Worked Solution (n=9)	42.24	10.708	0.166	-1.291	4.24	1.329	0.770	-0.837	2.05	1.224	1.894	3.520
Interactive (Low) (n=9)	44.137	15.496	1.805	3.738	5.24	1.401	1.106	1.120	3.34	1.802	2.502	7.008
Interactive (High) (n=9)	47.067	13.364	-0.244	-1.782	4.62	1.945	-0.363	-1.087	6.62	4.427	2.012	4.660
Overall	43.74	12.03	0.841	0.780	4.80	1.607	0.46	-0.077	4.01	2.986	2.066	5.702

	Pre Test Duration				Post Test Duration				Pre-Post Duration			
	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis
Control (n=9)	15.32	1.699	0.211	-1.630	11.40	4.643	0.572	-0.123	-3.92	3.938	0.629	0.426
Worked Solution (n=9)	14.77	3.414	-0.631	-0.567	9.40	2.854	2.404	0.717	-5.38	3.425	0.081	-1.412
Interactive (Low) (n=9)	14.55	5.193	0.747	-0.927	8.84	3.907	1.355	1.367	-5.71	4.042	-2.076	4.795
Interactive (High) (n=9)	14.51	5.581	0.437	-0.932	8.74	3.434	1.307	1.651	-5.77	2.808	0.430	-1.424
Overall	14.79	4.088	0.287	-0.323	9.60	3.762	1.112	0.588	-5.19	3.513	-0.284	1.308

Note: , n=36 participants, Duration is measured in minutes

3.4 Reception

The composite scores for the participants' reception towards the feedback and material were out of a scale of five. On average, the participants rated the difficulty of the topic as 2.08 (*SD* = .996). A cognitive load score was calculated for each participant by averaging their responses on seven 5-point Likert scale items. Cronbach's Alphas were calculated for the Cognitive Load scale (α = .818). On average, the participants rated their cognitive load during the test as 2.00 (*SD* = .624). For the participants who received feedback, they rated the helpfulness of the feedback as 4.26 (*SD* = .944). For the participants who received feedback, they rated the difficulty of using the feedback as 1.59 (*SD* = .636). Participant prior knowledge scores and post-survey knowledge scores were out of a scale of ten. On average, the participants rated their own prior knowledge of the topic as 1.36 (*SD* = 1.775). On average, the participants rated their own post-survey knowledge of the topic as 7.94 (*SD* = 1.985). On average, the participants increased their knowledge of the topic by 6.58 (*SD* = 2.568).

The normality of the post-survey questionnaire measures were assessed using the Shapiro-Wilk test. The results revealed that Topic Difficulty score ($w = 0.819$, $df = 36$, $p < .001$), Feedback Helpfulness score ($w = 0.740$, $df = 27$, $p < .001$), Feedback Difficulty Score ($w = 0.752$, $df = 27$, $p < .001$), Prior Knowledge score ($w = 0.753$, $df = 36$, $p < .001$), Post-Survey Knowledge score ($w = 0.878$, $df = 36$, $p < .001$), and Knowledge Score Delta ($w = 0.925$, $df = 36$, $p = .018$) were all not normally distributed. The test also revealed that the Cognitive Load score ($w = 0.753$, $df = 36$, $p = .150$) was normally distributed.

The homogeneity of variances were assessed using Levene's test. Levene's tests indicated that the assumption of homogeneity of variances was not violated for each of the variables: Topic Difficulty, $F(3, 32) = 0.990$, $p = .410$; Cognitive Load , $F(3, 32) = 1.078$, $p = .372$; Feedback Helpfulness, $F(2, 24) = 0.685$, $p = .514$; Feedback Difficulty, $F(2, 24) = 0.471$, $p = .630$; Prior Knowledge, $F(3, 32) = 0.475$, $p = .702$, Post-Survey Knowledge, $F(3, 32) = 1.037$, $p = .390$, and Knowledge Score Delta $F(3, 32) = 0.898$, $p = .453$.

A One-Way ANOVA was conducted to compare the Topic Difficulty scores between each condition. This was conducted in order to determine if there were differences in how difficult participants found the material to be based on the feedback condition they had. The Topic Difficulty scores for each condition are as follows: Control ($m = 2.44$, $SD = 1.236$), No-Interaction ($m = 2.33$, $SD = 1.118$), Low-Interaction ($m = 1.78$, $SD = 0.833$), and High-Interaction ($m = 1.75$ $SD = 0.667$). The results revealed that there was not a significant difference between groups, $F(3) = 1.163$, $p = 0.339$.

A One-Way ANOVA was conducted to compare the Cognitive Load scores between each condition. This was conducted in order to determine if participants perceived themselves to have less strain on their cognitive load based on the feedback condition they had. The Cognitive Load scores for each condition are as follows: Control (m = 2.16, SD = 0.587), No-Interaction (m = 2.19, SD = 0.770), Low-Interaction (m = 2.00, SD = 0.616), and High-Interaction (m = 1.67 SD = 0.435). The results revealed that there was not a significant difference between groups, $F(3)$ = 1.379, p = 0.267.

A One-Way ANOVA was conducted to compare the Feedback Helpfulness scores between each condition. This was conducted in order to determine if there were differences in how helpful participants found the feedback to be based on the condition they had. The Feedback Helpfulness scores for each condition are as follows: No-Interaction (m = 4.00, SD = 1.323), Low-Interaction (m = 4.44, SD = 0.726), and High-Interaction (m = 4.33 SD = 0.707). The results revealed that there was not a significant difference between groups, $F(2)$ = 0.520, p = 0.601.

A One-Way ANOVA was conducted to compare the Feedback Difficulty scores between each condition. This was conducted in order to determine if there were differences in how difficult participants found the feedback to be based on the condition they had. The Feedback Difficulty scores for each condition are as follows: No-Interaction (m = 1.78, SD = 0.667), Low-Interaction (m = 1.44, SD = 0.527), and High-Interaction (m = 1.56 SD = 0.726). The results revealed that there was not a significant difference between groups, $F(2)$ = 0.622, p = 0.545.

A One-Way ANOVA was conducted to compare the Prior Knowledge scores between each condition. This was conducted in order to establish that participants had similar baseline ratings of their understanding of the material to serve as equal comparisons. The Prior Knowledge scores for each condition are as follows: Control (m = 1.89, SD = 2.205), No-

Interaction (*m* = 1.67, *SD* = 1.581), Low-Interaction (*m* = 0.78, *SD* = 1.641), and High-Interaction (*m* = 1.11 *SD* = 1.691). The results revealed that there was not a significant difference between groups, $F(3) = 1.720$, $p = 0.548$.

A One-Way ANOVA was conducted to compare the Post-Survey Knowledge scores between each condition. This was conducted in order to determine if there were differences in how high participants rated their understanding of the material based on the feedback condition they had. The Post-Survey Knowledge scores for each condition are as follows: Control (*m* = 7.00, *SD* = 2.500), No-Interaction (*m* = 7.67, *SD* = 1.936), Low-Interaction (*m* = 8.33, *SD* = 1.936), and High-Interaction (*m* = 8.78 *SD* = 1.202). The results revealed that there was not a significant difference between groups, $F(3) = 1.433$, $p = 0.251$.

A One-Way ANOVA was conducted to compare the Knowledge Score Deltas between each condition. This was to determine if there were differences in how participants rated their improvement of their understanding of the material. The Knowledge Score Deltas means for each condition are as follows: Control (*m* = 5.11, *SD* = 2.713), No-Interaction (*m* = 6.00, *SD* = 2.398), Low-Interaction (*m* = 7.56, *SD* = 2.744), and High-Interaction (*m* = 7.67 *SD* = 1.732). The results revealed that there was not a significant difference between groups, $F(3) = 2.349$, $p = 0.091$.

Pearson correlation tests were conducted between the post-survey questionnaire measures, the Pre Test scores, Post Test Scores, and Test Score Deltas. This was to determine if there were any trends in how participants scored on assessments and their ratings of the content and the feedback. Pre Test scores had a moderate negative correlation with Topic Difficulty ratings ($r = -.562$, $p < .001$) and a moderate positive correlation with Post-Survey Knowledge ratings ($r = .458$, $p = .005$). Post Test scores had a moderate negative correlation with Topic

Difficulty ratings ($r = -.581$, $p < .001$) and a high positive correlation with Post-Survey Knowledge ratings ($r = .627$, $p < .001$). Additionally, Topic Difficulty ratings had a moderate negative correlation with Post-Survey Knowledge ratings ($r = -.518$, $p < .001$) and a moderate positive correlation with Cognitive Load ratings ($r = .584$, $p < .001$). Prior Knowledge ratings have a weak negative correlation with Cognitive Load ratings ($r = -.337$, $p = .044$). Post-Survey Knowledge ratings have a moderate negative correlation with Cognitive Load ($r = -.564$, $p < .001$), a moderate positive correlation with Feedback Helpfulness ratings ($r = .430$, $p = .025$), and a moderate negative correlation with Feedback Difficulty ratings ($r = -.426$, $p = .027$). Feedback Helpfulness ratings had a high negative correlation with Feedback Difficulty ratings ($r = -.650$, $p < .001$).

Table 6. Reception Descriptive statistics by condition

	Topic Difficulty				Cognitive Load				Feedback Helpfulness				Feedback Difficulty			
	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis
Control (n=9)	2.44	1.236	-1.439	-0.800	2.16	0.587	-0.964	0.513	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Worked Solution (n=9)	2.33	1.118	-0.537	-1.275	2.19	0.770	0.374	-1.360	4.00	1.323	-1.666	2.950	1.78	0.667	-0.254	-0.040
Interactive (Low) (n=9)	1.78	0.833	-0.501	-0.040	2.00	0.616	0.718	0.053	4.44	0.726	-1.014	0.185	1.44	0.527	-0.271	-2.571
Interactive (High) (n=9)	1.75	0.667	-0.254	1.522	1.67	0.435	0.273	-1.491	4.33	0.707	-0.606	-0.286	1.56	0.726	-1.014	0.185
Overall	2.08	0.996	-1.110	1.179	2.00	0.624	0.393	-0.636	4.26	0.944	-1.751	4.198	1.59	0.636	-0.594	-0.484

	Prior Knowledge				Post-Survey Knowledge				Post-Prior Knowledge			
	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis
Control (n=9)	1.89	2.205	1.169	0.175	7.00	2.500	-1.173	0.955	5.11	2.713	0.574	-0.365
Worked Solution (n=9)	1.67	1.581	1.193	1.543	7.67	1.936	-0.192	-1.384	6.00	2.398	0.000	-1.671
Interactive (Low) (n=9)	0.78	1.641	2.633	7.235	8.33	1.936	-0.738	-1.003	7.56	2.744	-1.128	0.688
Interactive (High) (n=9)	1.11	1.691	1.772	3.203	8.78	1.202	-0.574	-1.099	7.67	1.732	-1.010	1.897
Overall	1.36	1.1775	1.394	0.917	7.94	1.985	-1.033	0.929	6.58	2.568	-0.364	-1.053

Note: Topic Difficulty, Cognitive Load, and Feedback Helpfulness are measured on a scale from 1-5. Prior Knowledge and Post-Survey Knowledge are measured on a scale of 1-10

In regards to preferences towards the feedback they received, 22.2% of the participants preferred more interactive elements, 7.4% of the participants preferred less interactive elements, and 70.4% participants felt that the interactivity they received was fine as it was. In the Worked Solution condition, 33.3% of these participants preferred more interactive elements while 66.6% of them felt that the interactivity they received was fine as it was. In the Low-Interaction condition, 22.2% of these participants preferred more interactive elements, 22.2% of these participants preferred less interactive elements, and 55.6% of them felt that the interactivity they received was fine as it was. In the High-Interaction condition, 11.1% of these participants preferred more interactive elements while 88.9% of them felt that the interactivity they received was fine as it was.

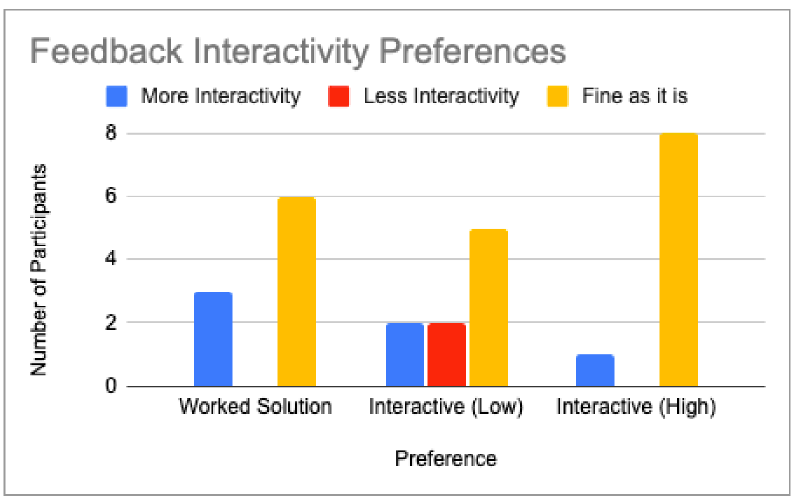


Figure 1: Feedback Interactivity Preferences by Condition

Chapter 4: Pilot Study Conclusions & Directions

4.1 Pilot Study Conclusions

The outcomes of this study hold significant potential for informing effective feedback practices and guiding future research in educational psychology.

Research Question 1: Interactive Feedback and Student Performance

If the first research question is validated, demonstrating that higher levels of interactive feedback result in better student performance, it will illuminate the importance of incorporating interactive elements into feedback mechanisms. This finding would support the notion that engaging students more deeply with feedback materials enhances their learning outcomes. Educational practitioners could leverage these insights to design feedback systems that maximize student interaction and engagement, potentially leading to improved educational outcomes.

Conversely, if lower levels of feedback interaction are found to be more effective, this would suggest a U-shaped relationship between feedback interaction and student learning. Excessive interaction might increase ratings of cognitive load, thereby diminishing students' ability to learn or apply new knowledge. This would indicate a need to balance interactivity to avoid overwhelming students. Should the no-interaction feedback condition prove most effective, it might reveal that interactive elements in feedback can sometimes distract and impede learning, prompting a reevaluation of when and how such elements are best used.

Research Question 2: Student Preferences for Feedback

The second research question focuses on student perceptions of different feedback styles. If students show a preference for the feedback style that also proves most effective in enhancing performance, this would highlight an optimal feedback approach for online formative assessments. Such alignment between effectiveness and preference would suggest that this feedback style not only enhances learning but also increases student satisfaction and engagement.

However, if there is no significant difference in student performance across feedback conditions, the preference data alone would guide the suggestion of which feedback method would still be the best option. Even if all methods are equally effective in terms of learning outcomes, the method preferred by students could foster a stronger connection to the material, promoting ongoing engagement. On the other hand, if students prefer a less effective feedback style, it would present an issue that could potentially be a difficult obstacle when considering the implementation of the feedback. Feedback that students find difficult or cumbersome may reduce their willingness to engage with the material. This would indicate a need to explore if students are willing to continue to use the feedback moving forward and how feedback can be adapted to meet both effectiveness and user-friendliness criteria.

Research Question 3: Feedback Efficiency

The third research question pertains to the time required to complete Post Tests under different feedback conditions and is similar in implications to the second research question. If the most effective feedback also results in the shortest Post Test completion times, it would demonstrate a beneficial synergy between efficiency and learning scores. This would advocate for the adoption of such feedback styles, especially in contexts where time efficiency is crucial.

Just as with student preference, even if there is no significant difference in student performance between the various conditions, then the results of the second research question may still indicate which feedback style serves to be most time-effective. In scenarios where feedback conditions yield similar learning outcomes, the feedback method requiring the least time would be preferable.

However, if the most effective feedback is also the most time-consuming in the Post Test, it suggests that such feedback may be better suited to contexts where students have the flexibility to invest more time in assessment. For time-constrained environments, it may necessitate finding a middle ground in performance improvement and time investment. Additionally, the duration of the feedback phase itself must be considered. Even if a condition has a shorter duration in the Post Test, it may require a longer duration during the feedback phase depending on the level of interaction. Because of this, high interactivity during this phase could offset time savings during the Post Test. This research question will thus shed light on the practical applicability of different feedback styles, potentially prompting further research into optimizing feedback for both effectiveness and efficiency.

Research Question 4: Learner Characteristics and Feedback Engagement

The fourth research question explores whether certain feedback styles are more effective for specific types of learners. If students with a high need for achievement, intrinsic value in mathematics, growth mindset, or self-efficacy engage more with interactive feedback, it would support the promotion of these characteristics in educational settings. Conversely, if certain learner types consistently avoid engaging with the feedback, it highlights the need to develop more accessible and appealing feedback styles for these students. This could lead to differentiated feedback approaches tailored to diverse learner profiles, enhancing the overall effectiveness of educational interventions.

In summary, the results of this study have the potential to significantly advance our understanding of effective feedback practices. By examining the relationship between feedback interactivity, student performance, perception, efficiency, and individual learner characteristics, this research aims to identify optimal strategies for enhancing educational outcomes. The findings will not only inform best practices in feedback design but also highlight areas for further investigation, ultimately contributing to the development of more effective and engaging learning environments.

Chapter 5: Main Study Methods

5.1 Participants

A total of 81 undergraduate students from Columbia and Barnard University were recruited for this study using the SONA recruitment platform. The recruitment flier was uploaded to the system and allowed participants to view information regarding the study and sign up for timeslots to participate in the study. Participants were eligible to participate if they are a student of Columbia or Barnard and are unfamiliar with the process of determining area/perimeter of Recursive Fractals. Participants were awarded two credits of research participation, which may be required or used as extra credit for certain classes.

5.2 Materials & Procedures

The study took place in private rooms in the Gottesman Library at Teachers College - Columbia University. Participants were seated at a table with a laptop and provided with a pen, two sheets of blank paper, and basic pocket calculator. Participants were informed that they could use the provided materials to work out problems. Participants were asked to avoid talking to other participants in the room, engaging in any distractions from their phone, or opening other tabs on the computer. Once they are given their instructions, participants were sent a link to start the study on *Qualtrics*.

Participants were first asked if they would like to consent to take part in the study, which would include a series of self-paced questionnaires and assessments. Upon agreeing, participants were asked to complete a 3-item demographics survey inquiring about their gender, race/ethnicity, and age. Following this, participants were then asked to complete a 20-item questionnaire regarding learner characteristics in which they choose how much they agree with various statements (see Appendix A). The choices consist of 5-point Likert scales ranging from strongly disagree to strongly agree. There are four statements for each of the following categories: Implicit Theory of Intelligence (IT), Self-Efficacy (SE), Intrinsic Value of Math (IV), Test Anxiety (TA), and Need for Achievement (NA). At least one question in each category were reverse-coded. The statements are presented in a randomized order five items at a time across four pages.

After the questionnaire, participants were given a short introduction to view that included text and visuals covering the concept of fractal shapes. Once they have seen the introduction, they moved on to the Pre Test phase. In this section, participants were tasked with solving a series of 10 multiple-choice questions. Each question had 5 choices and participants were allowed to use the calculator provided and to use the scrap paper. All students received the questions in the same order, with the first five questions related to area and last five questions related to perimeter. For the five questions from each topic, the first question uses the same shape as the example provided in the mini-lesson, while the next two questions use a slight variation of the shape, and the next two questions use a different shape (see Appendix B). All ten questions were equivalent in format, with different numbers used for each of the variables.

Once participants had completed the Pre Test they were given instructions regarding the Acquisition Phase of the study. They were presented with two short mini-lessons about finding the area and finding the perimeter of fractal shapes. Each mini-lesson utilized a model shape to serve as an example to teach the participant how to use the general formula. The lesson includes definitions of each variable and a sample worked-solution to demonstrate the process of using the formula (see Appendix C). They would be allowed to spend up to 5 minutes reviewing each lesson, and would not be able to return to the lesson page after that section.

The next step is the Formative Feedback Phase. In this section, participants were tasked with solving a series of 10 multiple-choice questions in the same format as the Pre Test phase. All of the questions were solvable using the formulas provided in the mini-lessons. Participants in the Non-Interaction Condition were given KR + KCR feedback, along with a fully-worked solution as a step-by-step guide demonstrating how to solve the question they had just answered (see Appendix E). Participants in the Interactive Condition were given a partial-worked solution. In these cases, certain portions of the worked solution required the participant to make choices or provide input in the guide. This involved using multiple-choice format to ask participants which formula should be used and how the variables would be plugged into the formula. This also includes fill-in-the-blank questions asking the participant to type in what value to input for each variable of the formula (see Appendix D). Participants in both conditions were given the feedback regardless of whether they answered the questions correctly or incorrectly. Participants were allowed to spend as much time as they want with the feedback before moving on to the next question.

After the Formative Feedback Phase, participants moved on to the Post Test Phase. In this section they were tasked with answering a series of 10 more multiple-choice questions. These problems were equivalent in format to the set from the Formative Feedback Phase, however they were all shuffled in a randomized order for this assessment. All participants were given no feedback after each question.

Participants were also given a cognitive load survey at Time 1 (after the Pre Test), Time 2 (after the formative feedback test), and Time 3 (after the post test). The cognitive load survey consisted of three items asking participants how strongly they agreed with various statements related to cognitive strain using a 5-point Likert scale.

Once they complete the assessments, all participants were given a Post Test Questionnaire (see Appendix F). The questionnaire asked participants how they felt about the difficulty of the topic, their perceived prior/current knowledge regarding the topic, how helpful they found the feedback, how difficult to use they found the feedback, how satisfied they were with their performance on the post test and how satisfied they were with how much they learned using 5-point Likert scales. The last item is a multiple choice question asking them if they would have preferred the explanation guides to be "Step-by-Step Worked-Out Solution", "Step-by-Step Worked-Out Solution with some multiple choice prompts", or "Step-by-Step Worked-Out Solution with some multiple choice & fill-in-the-blank prompts". For this question, participants were given examples of the feedback from the three options to use as a comparison.

After completing the Post Test questionnaire, they were asked to provide details in order to contact their professors to confirm their participation and award them the research credit. Participants will be reminded that this information would not be used in the analysis for the study.

5.3 Scoring & Coding

A composite score was created for Implicit Theory of Intelligence (IT), Self-Efficacy (SE), Intrinsic Value of Math (IV), Test Anxiety (TA), and Need for Achievement (NA) by calculating each of the averages of their respective items from the Learner Characteristics Questionnaire. A composite score was created for rating of Cognitive Load (CL) at each of the three timeframes by calculating the average of three items. A Pre Test score was created by grading the number of correct answers from the Pre Test phase out of ten. A Formative Feedback Test score was created by grading the number of correct answers from the Formative Feedback phase out of ten. A Post Test score was created by grading the number of correct answers from the Post Test Phase out of ten. Each of the questions in the Pre Test, Formative Feedback Test, and Post Test were weighted equally. Delta scores were calculated by subtracting the Pre Test score from the Post Test score, the Pre Test score from the Formative Feedback Test Score, and the Formative Feedback Test score from the Post Test score. A Feedback Score was created for the participants in the Interactive condition based on the number of correct choices and inputs they made in their feedback. The Feedback Score was graded based on the two multiple choice questions weighted equally and the four fill-in-the-blank questions weighted equal to the two multiple choice questions.

Chapter 6: Main Study Results

There were initially n=81 total participants in the study divided into the two conditions. There were 40 participants in the Non-Interactive Condition and 41 participants in the Interactive Condition. Data from four participants were used as available due to incompletion of the study. The participants were 55.7% female and 44.3% male and comprised 35.4% White/Caucasian, 31.6% Asian/Pacific Islander, 13.9% Black/African American, 13.9% Hispanic/Latino, and 5.1% other. SPSS software (Version 28.0.0) was used to perform all statistical analysis of the data.

6.1 Learner Characteristics

The composite scores from the four items representing each of the learner characteristics were out of a scale of five. Descriptive data and reliability analysis are reported in Table 7. Guttman's λ^2 were calculated in for each of the characteristics as follows: Implicit Theory of Intelligence (λ^2= .829), Self-Efficacy (λ^2= .887), Intrinsic Value of Math (λ^2= .649), Testing Anxiety (λ^2= .776), and Need for Achievement in Math (λ^2= .498). Implicit Theory of Intelligence, Self-Efficacy, and Testing Anxiety were found to be significantly reliable while Intrinsic Value of Math and Need for Achievement in Math were found to be unreliable. The mean score for Implicit Theory of Intelligence was 3.71 (*SD* = 0.826). The mean score for Self-Efficacy was 2.72 (*SD* = 1.040). The mean score for Intrinsic Value of Math was 3.89 (*SD* = 0.659). The mean score for Testing Anxiety was 3.53 (*SD* = 0.912). The mean score for Need for Achievement in Math was 3.52 (*SD* = 0.614).

Pearson correlation tests were conducted between each of the characteristics, the Pre Test scores, Formative Feedback Test Scores, Post Test Scores, Test Score Deltas, Feedback Score, and Feedback Duration. These correlations were conducted to determine if there were any trends between these factors to inform further regression analyses. Implicit Theory of Intelligence had a weak negative correlation with Post Test Scores (r = -.258, p = .023) and a weak negative correlation with Post-Pre Delta Scores (r = -.322, p = .004). Self-Efficacy had a weak negative correlation with Test Anxiety (r = -.313, p = .004), a weak positive correlation with Pre Test Scores (r = .244, p = .028), and a weak positive correlation with Post Test Scores (r = .244, p = .032). Testing Anxiety had a weak negative correlation with Pre Test Scores (r = -.294, p = .008), a weak negative correlation with Formative Feedback Test Scores (r = -.310, p = .005), and a weak negative correlation with Post Test Scores (r = -.268, p = .018). Intrinsic Value in Math had a weak positive correlation with Pre Test Scores (r = .244, p = .028), a weak positive correlation with Post Test Scores (r = .270, p = .017), and a weak positive correlation with Feedback Score (r = .328, p = .041). Need for Achievement was not found to have a significant correlation with any of the other characteristics or test scores.

Table 7. Learner Characteristics Descriptives and Reliability

	Mean	SD	Gutman's λ2
Implicit Theory (IT)	3.71	0.826	0.829
Self-Efficacy (SE)	2.72	1.040	0.887
Intrinsic Value (IV)	3.89	0.659	0.649
Test Anxiety (TA	3.53	0.912	0.776
Need for Achievement (NA)	3.52	0.614	0.498

Note: n=4 items for each category

A multiple regression was conducted to examine if Self-Efficacy, Intrinsic Value in Math, and Test Anxiety scores were predictors of Pre Test Scores. This was to determine if these learner characteristics were related to the participant's baseline performance on the assessment.

The results indicated that the regression model was significant, $F(3, 75) = 4.744$, $p = .004$, and explained 12.6% of the variance in Pre Test , Adjusted $R^2 = .126$. The results also indicated that Test Anxiety scores significantly predicted Pre Test Scores, $\beta = -0.309$, $t(75) = -2.489$, $p = .015$. Self-Efficacy did not significantly predict Pre Test Scores $\beta = 0.038$, $t(75) = 0.234$, $p = .815$. Intrinsic Value in math did not significantly predict Pre Test Scores $\beta = 0.147$, $t(75) = 1.011$, $p = .315$.

A simple linear regression was conducted to examine if Test Anxiety scores were predictors of Formative Feedback Test Scores. This was to determine if their anxiety might impact the participant's performance during the feedback phase. The results indicated that Test Anxiety scores significantly predicted Formative Feedback Test Scores, $\beta = -0.239$, $t(77) = 10.876$, $p < .001$. The regression model was significant, $F(1, 77) = 4.680$, $p = .034$, and explained 5.7% of the variance in Formative Feedback Test scores, $R^2 = .057$.

A multiple regression was conducted to examine if Implicit Theory, Self-Efficacy, Test Anxiety, and Intrinsic Value scores were predictors of Post Test Scores. This was to determine what learner characteristic factors might play a role in the participants' final outcome performance. The results indicated that the regression model was significant, $F(4, 73) = 2.601$, $p = .043$, and explained 7.7% of the variance in Post Test , Adjusted $R^2 = .077$. Implicit Theory did not significantly predict Post Test Scores $\beta = -0.111$, $t(73) = -0.991$, $p = .325$. Self-Efficacy did not significantly predict Post Test Scores $\beta = 0.227$, $t(73) = 1.317$, $p = .192$. Intrinsic Value in Math did not significantly predict Post Test Scores $\beta = 0.118$, $t(73) = 0.760$, $p = .450$. Test Anxiety did not significantly predict Post Test Scores $\beta = -0.033$, $t(73) = -0.256$, $p = .799$.

A simple linear regression was conducted to examine if Implicit Theory scores were predictors of Post-Pre Delta Scores. This was to determine if a participants' learner mindset would impact their overall improvement in scores. The results indicated that Implicit Theory scores did not significantly predict Post-Pre Delta Scores, $F(1, 77) = 0.294$, $p = .589$.

6.2 Proportion Correct

Descriptives of the Pre Test, Formative Feedback Test, and Post Test question scores are reported in Table 8. The mean scores for each of the 10 multiple-choice question in the Pre Test are as follows: Q1 (*m* = 0.52, *SD* = 0.503), Q2 (*m* = 0.62, *SD* = 0.489), Q3 (*m* = 0.69, *SD* = 0.465), Q4 (*m* = 0.69, *SD* = 0.465), Q5 (*m* = 0.84, *SD* = 0.369), Q6 (*m* = 0.37, *SD* = 0.486), Q7 (*m* = 0.40, *SD* = 0.492), Q8 (*m* = 0.68, *SD* = 0.470), Q9 (*m* = 0.52, *SD* = 0.503), Q10 (*m* = 0.63, *SD* = 0.486).

The mean scores for each of the 10 multiple-choice question in the Formative Feedback Test are as follows: Q1 (*m* = 0.78, *SD* = 0.418), Q2 (*m* = 0.79, *SD* = 0.414), Q3 (*m* = 0.91, *SD* = 0.286), Q4 (*m* = 0.82, *SD* = 0.384), Q5 (*m* = 0.91, *SD* = 0.286), Q6 (*m* = 0.77, *SD* = 0.422), Q7 (*m* = 0.73, *SD* = 0.445), Q8 (*m* = 0.70, *SD* = 0.463), Q9 (*m* = 0.71, *SD* = 0.457), Q10 (*m* = 0.76, *SD* = 0.430).

The mean scores for each of the 10 multiple-choice question in the Post Test are as follows: PQ1 (*m* = 0.82, *SD* = 0.388), PQ2 (*m* = 0.82, *SD* = 0.388), PQ3 (*m* = 0.88, *SD* = 0.323), PQ4 (*m* = 0.74, *SD* = 0.441), PQ5 (*m* = 0.88, *SD* = 0.323), PQ6 (*m* = 0.83, *SD* = 0.377), PQ7 (*m* = 0.92, *SD* = 0.268), PQ8 (*m* = 0.72, *SD* = 0.451), PQ9 (*m* = 0.80, *SD* = 0.406), PQ10 (*m* = 0.85, *SD* = 0.361).

The mean score for the Pre Test was 59.5% (SD = 0.211), the mean score for the Formative Feedback Test was 79.0% (SD = 0.205), and the mean score for the Post Test was 81.9% (SD = 0.225). The mean difference between the Pre Test and the Post Test was an increase of 23.5% (SD = 0.259). The mean difference between the Pre Test and the Formative

Feedback Test was an increase of 19.5% (SD = 0.239). The mean difference between the Formative Feedback Test and the Post Test was an increase of 3.5% (SD = 0.198).

Table 8. Pre Test, Formative Feedback Test, and Post Test Question Item Difficulty

Pre-Test			Formative Feedback Test			Post-Test		
	Mean	SD		Mean	SD		Mean	SD
Q1	0.52	0.503	Q1	0.78	0.418	Q1	0.82	0.388
Q2	0.62	0.489	Q2	0.79	0.414	Q2	0.82	0.388
Q3	0.69	0.465	Q3	0.91	0.286	Q3	0.88	0.323
Q4	0.69	0.465	Q4	0.82	0.384	Q4	0.74	0.441
Q5	0.84	0.369	Q5	0.91	0.286	Q5	0.88	0.323
Q6	0.37	0.486	Q6	0.77	0.422	Q6	0.83	0.377
Q7	0.40	0.492	Q7	0.73	0.445	Q7	0.92	0.268
Q8	0.68	0.470	Q8	0.70	0.463	Q8	0.72	0.451
Q9	0.52	0.503	Q9	0.71	0.457	Q9	0.80	0.406
Q10	0.63	0.486	Q10	0.76	0.430	Q10	0.85	0.361
Overall	0.60	0.210	Overall	0.79	0.205	Overall	0.82	0.225

Note: Means indicate the accuracy rate of each question, from a range of 0.00-1.00

For each question, point-biserial correlation coefficients were calculated to assess the strength and direction of the relationship between correct/incorrect responses and the total test score (Table 9). The results for the Pre Test questions with the Pre Test scores are as follows: Q1 ($r = 0.604$, $p < 0.001$), Q2 ($r = 0.480$, $p < 0.001$), Q3 ($r = 0.483$, $p < 0.001$), Q4 ($r = 0.534$, $p < 0.001$), Q5 ($r = 0.424$, $p < 0.001$), Q6 ($r = 0.520$, $p < 0.001$), Q7 ($r = 0.345$, $p < 0.002$), Q8 ($r = 0.060$, $p = 0.597$), Q9 ($r = 0.438$, $p < 0.001$), Q10 ($r = 0.545$, $p < 0.001$). All of the item responses had a significant positive correlation with the total Pre Test score except for Question 8.

The results for the Formative Feedback Test questions with the Formative Feedback scores are as follows: Q1 ($r = 0.579$, $p < 0.001$), Q2 ($r = 0.579$, $p < 0.001$), Q3 ($r = 0.422$, $p < 0.001$), Q4 ($r = 0.482$, $p < 0.001$), Q5 ($r = 0.313$, $p = 0.005$), Q6 ($r = 0.581$, $p < 0.001$), Q7 ($r =$

0.519, p < 0.001), Q8 (r = 0.522, p < 0.001), Q9 (r = 0.475, p < 0.001), Q10 (r = 0.569, p < 0.001). All of the item responses had a significant positive correlation with the total Formative Feedback Test score.

The results for the Post Test questions with the Post scores are as follows: Q1 (r = 0.553, p < 0.001), Q2 (r = 0.708, p < 0.001), Q3 (r = 0.528, p < 0.001), Q4 (r = 0.648, p < 0.001), Q5 (r = 0.714, p < 0.001), Q6 (r = 0.501, p < 0.001), Q7 (r = 0.261, p = 0.021), Q8 (r = 0.665, p < 0.001), Q9 (r = 0.640, p < 0.001), Q10 (r = 0.434, p < 0.001). All of the item responses had a significant positive correlation with the total Post Test score.

Table 9. Point-Biserial Correlations for Pre Test, Formative Feedback Test and Post Test

	Pre-Test			Formative Feedback Test			Post-Test	
	Correlation (*r*)	sig.		Correlation (r)	sig.		Correlation (r)	sig.
Q1	0.604	0.001**	**Q1**	0.579	0.001**	**Q1**	0.553	0.001**
Q2	0.480	0.001**	**Q2**	0.579	0.001**	**Q2**	0.708	0.001**
Q3	0.483	0.001**	**Q3**	0.422	0.001**	**Q3**	0.528	0.001**
Q4	0.534	0.001**	**Q4**	0.482	0.001**	**Q4**	0.648	0.001**
Q5	0.424	0.001**	**Q5**	0.313	0.005*	**Q5**	0.714	0.001**
Q6	0.520	0.001**	**Q6**	0.581	0.001**	**Q6**	0.501	0.001**
Q7	0.345	0.002*	**Q7**	0.519	0.001**	**Q7**	0.261	0.021*
Q8	0.060	0.597	**Q8**	0.522	0.001**	**Q8**	0.665	0.001**
Q9	0.438	0.001**	**Q9**	0.475	0.001**	**Q9**	0.640	0.001**
Q10	0.545	0.001**	**Q10**	0.569	0.001**	**Q10**	0.434	0.001**

*Correlation is significant at the .05 level (2-tailed)

**Correlation is significant at the .01 level (2-tailed)

The data meets the assumption of normality due to the central limit theorem, in that there are greater than 30 participants in each condition. The homogeneity of variances were assessed using Levene’s test. The results revealed that there were equal variances between the groups for the Pre Test Scores [$F(1, 77) = 0.197$, $p = .658$], Formative Feedback Test Scores [$F(1, 77) =$

0.525, $p = .471$], Post Test Scores [$F(1, 76) = 0.013$, $p = .991$], Post-Pre Delta Scores [$F(1, 75) = 2.018$, $p = .160$], Formative-Pre Delta Scores [$F(1, 77) = 0.217$, $p = .643$], and Post-Formative Delta Scores [$F(1, 75) = 0.221$, $p = .639$].

A two-tailed Student' *T*-test was conducted to compare the means of the Pre Test scores between the Interactive condition ($m = 0.56$, $SD = 0.205$) and the Non-Interactive condition ($m = 0.63$, $SD = 0.212$). This was to confirm that there were no differences between the groups at their baseline. The results revealed that there was not a significant difference between groups $t(77) = -1.623$, $p = 0.109$. A one-tailed Student' *T*-test was conducted to compare the means of the Formative Feedback Test scores between the Interactive condition ($m = 0.78$, $SD = 0.199$) and the Non-Interactive condition ($m = 0.80$, $SD = 0.212$). This was to determine if participants in the interactive condition scored higher than the participants in non-interactive condition for the formative feedback assessment. The results revealed that there was not a significant difference between groups $t(77) = -0.443$, $p = 0.670$. A one-tailed Student' *T*-test was conducted to compare the means of the Post Test scores between the Interactive condition ($m = 0.82$, $SD = 0.232$) and the Non-Interactive condition ($m = 0.82$, $SD = 0.221$). This was to determine if participants in the interactive condition scored higher than the participants in non-interactive condition for the final assessment. The results revealed that there was not a significant difference between groups $t(76) = -0.150$, $p = 0.559$.

A one-tailed Student's *T*-test was conducted to compare the means of the Post-Pre Delta scores between the Interactive condition ($m = 0.28$, $SD = 0.234$) and the Non-Interactive condition ($m = 0.19$, $SD = 0.278$). This was to determine if participants in the interactive condition improved more overall than the participants in non-interactive condition. The results revealed that there was not a significant difference between groups $t(75) = 1.479$, $p = 0.072$. A one-tailed Student's *T*-test was conducted to compare the means of the Formative-Pre Delta scores between the Interactive condition (m = 0.05, SD = 0.200) and the Non-Interactive condition (m = 0.03, SD = 0.198). The results revealed that there was not a significant difference between groups $t(77) = 1.036$, $p = 0.152$. A one-tailed Student' *T*-test was conducted to compare the means of the Post-Formative Delta scores between the Interactive condition (m = 0.22, SD = 0.240) and the Non-Interactive condition (m = 0.17, SD = 0.237). The results revealed that there was not a significant difference between groups $t(75) = 0.421$, $p = 0.337$.

A one-tailed Dependent Samples *T*-test was conducted to compare the means of the Pre Test Scores (m = 0.60, SD = 0.210) and the Post Test Scores (m = 0.82, SD = 0.225). This was to determine if the participants across both conditions improved in general regardless of the type of feedback they engaged with. The results revealed that there was a significant difference between groups $t(77) = -7.371$, $p < 0.001$. A one-tailed Dependent Samples *T*-test was conducted to compare the means of the Pre Test Scores (m = 0.60, SD = 0.210) and the Formative Feedback Test Scores (m = 0.79, SD = 0.205). The results revealed that there was a significant difference between groups $t(77) = -7.264$, $p < 0.001$. A one-tailed Dependent Samples *T*-test was conducted to compare the means of the Post Test Scores (m = 0.82, SD = 0.225) and the Formative Feedback Test Scores (m = 0.79, SD = 0.205). The results revealed that there was not a significant difference between groups $t(77) = -1.357$, $p = 0.089$.

Pearson correlation tests were conducted between the Pre Test scores, Formative Feedback Test Scores, Post Test Scores, and Test Score Deltas. This was to determine if there was consistency on how participants performed on each of the assessments. Pre Test scores had a weak positive correlation with Formative Feedback Test scores ($r = .341$, $p = .002$), a weak positive correlation with Post Test Scores ($r = .234$, $p = .039$), and a moderate negative correlation with Post-Pre Delta Scores ($r = -.585$, $p < .001$). Formative Feedback Test scores had a moderate positive correlation with Post Test scores ($r = .571$, $p < .001$). Post Test scores had a moderate positive correlation with Post-Pre Delta scores ($r = .619$, $p < .001$).

Table 10. Test Score Descriptive statistics by condition

	Pre Test				Formative Feedback Test				Post Test			
	Mean	**SD**	**Skew**	**Kurtosis**	**Mean**	**SD**	**Skew**	**Kurtosis**	**Mean**	**SD**	**Skew**	**Kurtosis**
Interactive	0.56	0.205	0.087	-0.724	0.78	0.199	-1.012	0.619	0.82	0.232	-1.511	1.126
Non-Interactive	0.63	0.212	-0.059	-0.978	0.80	0.212	-0.963	0.097	0.82	0.221	-1.161	0.187
Overall	0.60	0.210	0.026	-0.873	0.79	0.205	-0.953	0.220	0.82	0.225	-1.326	0.609

	Post-Pre Delta				Post-Formative Delta				Formative-Pre Delta			
	Mean	**SD**	**Skew**	**Kurtosis**	**Mean**	**SD**	**Skew**	**Kurtosis**	**Mean**	**SD**	**Skew**	**Kurtosis**
Interactive	0.28	0.234	-0.455	-0.016	0.05	0.200	-0.037	0.552	0.22	0.240	-0.227	-0.954
Non-Interactive	0.19	0.278	0.014	-0.593	0.03	0.198	0.265	1.133	0.17	0.237	0.458	-0.081
Overall	0.24	0.259	-0.235	-0.476	0.04	0.198	0.113	0.675	0.20	0.239	0.114	-0.710

Note: n=81 participants, means represent average assessment scores on a scale of 0.00-1.00

Participants in the Interactive condition were given feedback scores based on the proportion of their answers correct in the feedback prompts ($m = .941$, $SD = .097$). Pearson correlation tests were conducted between the Feedback Score, Pre Test scores, Formative Feedback Test Scores, Post Test Scores, and Post-Pre Delta Scores. This was to determine if participants' proportion of correct responses to feedback prompts were related to their performance on assessments. Pre Test scores had a weak positive correlation with Feedback Score ($r = .316$, $p = .050$). Formative Feedback Test scores had a moderate positive correlation with Feedback Score ($r = .676$, $p < .001$). Post Test scores had a strong positive correlation with Feedback Score ($r = .753$, $p < .001$). Post-Pre Delta scores had a moderate positive correlation with Feedback Score ($r = .474$, $p = .003$).

Simple linear regressions were conducted to examine if Feedback Score is a predictor of Formative Feedback Test Scores, Post Test Scores, and Post-Pre Delta Scores. The results indicated that Feedback Score significantly predicted Formative Feedback Test Scores, $\beta = 0.676$, $t(77) = -2.250$, $p = .030$. The regression model was significant, $F(1, 77) = 31.185$, $p < .001$, and explained 45.7% of the variance in Formative Feedback Test scores, $R^2 = .457$. The results also indicated that Feedback Score significantly predicted Post Test Scores, $\beta = 0.753$, $t(77) = -3.605$, $p < .001$. The regression model was significant, $F(1, 77) = 48.347$, $p < .001$, and explained 56.6% of the variance in Post Test scores, $R^2 = .566$. Additionally the results indicated that Feedback Score significantly predicted Post-Pre Delta Scores, $\beta = 0.474$, $t(77) = 3.229$, $p = .003$. The regression model was significant, $F(1, 77) = 10.426$, $p = .003$, and explained 22.5% of the variance in Post Test scores, $R^2 = .225$.

6.3 Duration

The mean score for the Total Test Duration was 57.59 minutes (*SD* = 16.05). The mean score for the Total Pre Test Duration was 16.79 minutes (*SD* = 9.96). The mean score for the Total Formative Feedback Test Duration was 13.88 minutes (*SD* = 6.49). The mean score for the Total Post Test Duration was 8.43 minutes (*SD* = 4.27). The mean score for the Total Feedback Duration was 3.61 minutes (*SD* = 3.06).

The data meets the assumption of normality due to the central limit theorem, in that there are greater than 30 participants in each condition. The homogeneity of variances were assessed using Levene’s test. The results revealed that there were equal variances between the groups for the Total Duration [$F(1, 77) = 0.545$, $p = .463$], Total Pre Test Duration [$F(1, 77) = 0.698$, $p = .406$], Formative Feedback Test Duration [$F(1, 77) = 0.001$, $p = .980$], Post-Pre Delta Duration [$F(1, 75) = 2.861$, $p = .095$], Formative-Pre Delta Duration [$F(1, 77) = 1.965$, $p = .165$], and

Post-Formative Delta Duration [$F(1, 75) = 0.635$, $p = .428$]. However Levene's Test of Homogeneity also revealed that Total Post Test Duration [$F(1, 75) = 5.696$, $p = .019$], and Total Feedback Duration [$F(1, 77) = 7.139$, $p = .009$] did not meet the assumption of homogeneity of variances.

A one-tailed Student' *T*-test was conducted to compare the means of the Total Test duration between the Interactive condition ($m = 57.6$, $SD = 17.02$) and the Non-Interactive condition ($m = 57.5$, $SD = 15.3$). This was to determine if participants required more time in the interactive condition due to the feedback compared to the non-interactive condition. The results revealed that there was not a significant difference between groups $t(77) = 0.027$, $p = 0.511$. Welch's test was conducted to compare the means of the Total Post Test Duration between the Interactive condition ($m = 8.04$, $SD = 3.33$) and the Non-Interactive condition ($m = 8.83$, $SD = 5.05$). This was to determine if participants in the interactive and non-interactive conditions differed in how much time they needed to complete the final assessment. The results revealed that there was not a significant difference between groups $F(1,65.8) = -0.815$, $p = 0.209$.

A one-tailed Student' *T*-test was conducted to compare the means of the Post-Pre Delta Durations between the Interactive condition ($m = -6.51$, $SD = 9.80$) and the Non-Interactive condition ($m = -9.35$, $SD = 12.54$). This was conducted in order to determine if there were any differences between the two conditions in regards to how much less time participants needed comparing their initial assessment and the final assessment. The results revealed that there was not a significant difference between groups $t(75) = 1.102$, $p = 0.863$. A one-tailed Student' *T*-test was conducted to compare the means of the Post-Formative Delta Durations between the Interactive condition ($m = -5.02$, $SD = 3.49$) and the Non-Interactive condition ($m = -5.36$, $SD = 4.86$). The results revealed that there was not a significant difference between groups $t(75) =$

0.344, $p = 0.634$. Welch's test was conducted to compare the means of the Total Feedback Durations between the Interactive condition ($m = 5.57$, $SD = 3.11$) and the Non-Interactive condition ($m = 1.71$, $SD = 1.33$). The results revealed that there was a significant difference between groups $F(1,51.2) = 7.134$, $p < 0.001$.

A two-tailed Student' *T*-test was conducted to compare the means of the Total Pre Test duration between the Interactive condition ($m = 14.8$, $SD = 9.23$) and the Non-Interactive condition ($m = 18.7$, $SD = 10.4$). This was conducted in order to confirm that the two groups had similar baselines and could serve as equal comparison groups. The results revealed that there was not a significant difference between groups $t(77) = -1.763$, $p = 0.082$. A two-tailed Student' *T*-test was conducted to compare the means of the Total Formative Feedback Test duration between the Interactive condition ($m = 13.4$, $SD = 6.15$) and the Non-Interactive condition ($m = 14.4$, $SD = 6.85$). This was conducted in order to determine if there amount of time participants would need to complete the feedback phase would differ based on the type of feedback they engaged with. The results revealed that there was not a significant difference between groups $t(77) = -0.682$, $p = 0.497$. A two-tailed Student' *T*-test was conducted to compare the means of the Formative-Pre Delta duration between the Interactive condition ($m = -1.45$, $SD = 10.6$) and the Non-Interactive condition ($m = -4.34$, $SD = 12.8$). The results revealed that there was not a significant difference between groups $t(77) = 1.094$, $p = 0.277$.

A one-tailed Dependent Samples *T*-test was conducted to compare the means of the Pre Test Duration ($m = 16.79$, $SD = 9.949$) and the Post Test Duration ($m = 8.431$, $SD = 4.265$). This was conducted in order to determine if participants across both groups overall needed less time in to complete the assessments as they progressed through the study. The results revealed that there was a significant difference between groups $t(77) = 6.323$, $p < 0.001$. A one-tailed

Dependent Samples *T*-test was conducted to compare the means of the Pre Test duration (m = 16.79, SD = 9.949) and the Formative Feedback Test duration (m = 13.88, SD = 6.489). The results revealed that there was a significant difference between groups $t(77) = 2.200$, $p = 0.015$. A one-tailed Dependent Samples *T*-test was conducted to compare the means of the Post Test Scores (m = 8.431, SD = 4.265) and the Formative Feedback Test Scores (m = 13.88, SD = 6.489). The results revealed that there was a significant difference between groups $t(77)$ = 10.728, $p < 0.001$.

Table 11. Duration Descriptive Statistics by Condition

	Total Duration				Pre-Test Duration				Formative Feedback Test Duration				Post Test Duration			
	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis
Interactive	57.6	17.02	0.356	-0.254	14.8	9.23	1.922	6.393	13.4	6.15	0.450	-0.934	8.04	3.33	0.100	-0.501
Non-Interactive	57.5	15.3	0.453	-0.156	18.7	10.4	1.110	1.332	14.4	6.85	0.911	0.963	8.83	5.05	0.782	-0.055
Overall	57.59	16.05	0.392	-0.246	16.79	9.96	1.404	2.710	13.88	6.49	0.728	0.281	8.43	4.27	0.757	0.490

	Post-Pre Delta Duration				Post-Formative Delta Duration				Formative-Pre Delta Duration				Feedback Duration			
	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis
Interactive	-6.51	9.80	-1.965	6.669	-5.02	3.49	-0.505	-0.899	-1.45	10.6	-1.754	-0.358	5.57	3.11	2.162	4.838
Non-Interactive	-9.35	12.54	-0.582	0.360	-5.36	4.86	-1.086	1.902	-4.34	12.8	5.966	0.418	1.71	1.33	1.402	1.134
Overall	-7.95	11.29	-1.089	1.923	-5.19	4.21	-0.985	1.702	-2.91	11.77	-0.898	1.846	3.61	3.06	1.944	5.306

Note: n=81 participants

Pearson correlation tests were conducted between the Total Feedback duration and the Pre Test Scores, Formative Feedback Scores, Post Test Scores, and Feedback Score. These correlations were conducted in order to see if there were any trends that could provide insight for further analyses. Total Feedback Duration had a weak negative correlation with Formative Feedback Test Scores ($r = -.353$, $p = .001$), a weak negative correlation with Post Test Scores ($r = -.327$, $p = .004$), and a moderate negative correlation with Feedback Score ($r = -.538$, $p < .001$). There was a nonsignificant negative correlation with Feedback Duration and Pre Test Scores ($r = -.152$, $p < .018$).

A simple linear regression was conducted to examine if Feedback Duration is a predictor of Formative Feedback Test Scores and Post Test Scores. The results indicated that Feedback Duration scores significantly predicted Formative Feedback Test Scores, $\beta = -0.353$, $t(77) = 25.982$, $p < .001$. The regression model was significant, $F(1, 77) = 10.981$, $p = .001$, and explained 12.5% of the variance in Formative Feedback Test scores, $R^2 = .125$. The results also indicated that Feedback Duration scores significantly predicted Post Test Scores, $\beta = -0.327$, $t(77) = 24.036$, $p < .001$. The regression model was significant, $F(1, 77) = 9.080$, $p = .004$, and explained 10.7% of the variance in Post Test scores, $R^2 = .107$.

6.4 Reception

The composite scores for the participants' reception towards the feedback and material were out of a scale of five. On average, the participants rated the difficulty of the topic as 3.17 (*SD* = 1.163). On average, the participants rated the helpfulness of the feedback as 4.09 (*SD* = 1.015) and the difficulty of using the feedback as 4.30 (*SD* = 0.933). On average, the participants rated their own prior knowledge of the topic as 1.74 (*SD* = 0.865) and their own post-survey knowledge of the topic as 3.58 (*SD* = 0.978). On average, the participants rated their satisfaction of how much they learned as 4.12 (*SD* = 0.858) and their satisfaction of how well they performed on the post test as 3.79 (*SD* = 1.056).

The data meets the assumption of normality due to the central limit theorem, in that there are greater than 30 participants in each condition. The homogeneity of variances were assessed using Levene's test. The results revealed that there were equal variances between the groups for the Topic Difficulty [$F(1, 75) = 0.269$, $p = .606$], Prior Knowledge [$F(1, 75) = 0.199$, $p = .657$], Post-Survey Knowledge [$F(1, 75) = 0.033$, $p = .856$], Helpfulness of Feedback [$F(1, 75) = 0.916$, $p = .342$], Difficulty of Feedback [$F(1, 75) = 0.025$, $p = .876$], and Satisfaction of Learned [F(1,

75) = 1.191, p = .279]. However Levene's Test of Homogeneity also revealed that Satisfaction of Post Test Performance [$F(1, 75) = 10.923$, $p = .001$] did not meet the assumption of homogeneity of variances.

A one-tailed Student' *T*-test was conducted to compare the means of the Topic Difficulty ratings between the Interactive condition ($m = 3.26$, $SD = 1.13$) and the Non-Interactive condition ($m = 3.08$, $SD = 1.20$). This was conducted to determine if participants in the interactive feedback condition perceived the content to be less difficult than those in the non-interactive condition. The results revealed that there was not a significant difference between groups $t(75) = 0.700$, $p = 0.757$. A two-tailed Student' *T*-test was conducted to compare the means of the Prior Knowledge ratings between the Interactive condition ($m = 1.68$, $SD = 0.809$) and the Non-Interactive condition ($m = 1.80$, $SD = 0.923$). This was conducted to confirm that the participants in each condition had similar baseline ratings of their knowledge of the content and could serve as equal comparisons. The results revealed that there was not a significant difference between groups $t(75) = -0.559$, $p = 0.578$. A one-tailed Student' *T*-test was conducted to compare the means of the Post-Survey Knowledge ratings between the Interactive condition ($m = 3.76$, $SD = 0.943$) and the Non-Interactive condition ($m = 3.41$, $SD = 0.993$). This was conducted to determine if participants in the interactive feedback condition rated their understanding of the material higher than those of the non-interactive feedback condition. The results revealed that there was not a significant difference between groups $t(75) = 1.599$, $p = 0.057$. A one-tailed Dependent Samples *T*-test was conducted to compare the means of the Prior Knowledge ratings ($m = 1.74$, $SD = 0.865$) and the Post-Survey Knowledge ratings ($m = 3.58$, $SD = 0.978$). This was conducted to see if the participants perceived their understanding of the material to improve overall, regardless of which condition they were in. The results revealed that there was a significant difference between groups $t(77) = -14.241$, $p < 0.001$.

A one-tailed Student' *T*-test was conducted to compare the means of the Helpfulness of Feedback ratings between the Interactive condition ($m = 4.34$, $SD = 0.815$) and the Non-Interactive condition ($m = 3.85$, $SD = 1.136$). This was conducted in order to compare whether participants in the interactive feedback condition found their feedback to be helpful than the participants in the non-interactive feedback condition. The results revealed that there was a significant difference between groups $t(75) = 2.196$, $p = 0.016$. A one-tailed Student' *T*-test was conducted to compare the means of the Difficulty of Feedback ratings between the Interactive condition ($m = 4.47$, $SD = 0.893$) and the Non-Interactive condition ($m = 4.13$, $SD = 0.951$). This was conducted in order to determine whether participants in the interactive feedback condition found their feedback to be difficult to use compared to the participants in the non-interactive feedback condition. The results revealed that there was not a significant difference between groups $t(75) = 1.643$, $p = 0.052$.

Welch's test was conducted to compare the means of the Satisfaction of Post Test Performance ratings between the Interactive condition ($m = 3.90$, $SD = 0.863$) and the Non-Interactive condition ($m = 3.69$, $SD = 1.217$). The results revealed that there was not a significant difference between groups $F(1,68.6) = 0.843$, $p = 0.201$. A one-tailed Student' *T*-test was conducted to compare the means of the Satisfaction of Learning the Material ratings between the Interactive condition ($m = 4.16$, $SD = 0.945$) and the Non-Interactive condition ($m = 4.08$, $SD = 0.774$). The results revealed that there was not a significant difference between groups $t(75) = 0.412$, $p = 0.341$. These two tests were conducted to examine how participants felt about two facets of satisfaction in regards to their learning and mastery of content.

Table 12. Reception Descriptive Statistics by Condition

	Topic Difficulty				Prior Knowledge				Post Knowledge				Feedback Helpfulness			
	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis
Interactive	3.263	1.131	-0.081	-0.730	1.684	0.809	0.977	0.307	3.763	0.943	-0.107	-0.970	4.342	0.815	-1.357	1.825
Non-Interactive	3.077	1.201	-0.155	-0.997	1.795	0.923	1.280	2.167	3.410	0.993	-0.079	-0.279	3.846	1.136	-1.270	1.213
Overal	3.169	1.163	-0.133	-0.867	1.740	0.865	1.163	1.477	3.584	0.978	-0.113	-0.580	4.094	0.976	-1.425	1.957

	Feedback Difficulty				Satisfaction of Learned				Satisfaction of Performance			
	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis
Interactive	4.474	0.893	-1.843	2.723	4.158	0.945	-1.345	2.245	3.895	0.863	-1.119	2.435
Non-Interactive	4.128	0.951	-1.429	2.414	4.077	0.774	-0.495	-0.099	3.691	1.217	-0.475	-1.039
Overall	4.299	0.933	-1.536	2.130	4.117	0.858	-1.000	1.351	3.792	1.056	-0.741	-0.091

Note: n=77, all categories were rated from a scale of 0-5

Pearson correlation tests were conducted between the post-survey questionnaire measures, the Pre Test scores, Formative Feedback Test Scores, Post Test Scores, and Post-Pre Score Deltas. These correlations were conducted in order to determine if there were any trends that could inform further analyses. Topic Difficulty had a weak positive correlation with Prior Knowledge (r = .227, p = .047), a moderate positive correlation with Post-Survey Knowledge (r = .444, p < .001), a moderate positive correlation with Feedback Difficulty (r = .487, p < .001), a moderate positive correlation with Satisfaction of Post Test Performance (r = .608, p < .001), a weak positive correlation with Pre Test Scores (r = .263, p = .021), a moderate positive correlation with Formative Feedback Test Scores (r = .457, p < .001), and a moderate positive correlation with Post Test Scores (r = .485, p < .001).

Prior Knowledge had a weak positive correlation with Post-Survey Knowledge (r = .244, p = .032) and a moderate positive correlation with Pre Test Scores (r = .440, p < .001). Post-Survey Knowledge had a weak positive correlation with Feedback Difficulty (r = .397, p < .001), a moderate positive correlation with Satisfaction with Post Test Performance (r = .400, p < .001), a weak positive correlation with Formative Feedback Test Scores (*r* = .301, *p* = .008), and a weak positive correlation with Post Test Scores (*r* = .328, *p* = .004).

Helpfulness of Feedback had a moderate positive correlation with Feedback Difficulty (*r* = .402, *p* < .001), a weak positive correlation with Satisfaction of Post Test Performance (*r* = .251, *p* = .028), a moderate positive correlation with Satisfaction of Material Learned (*r* = .426, *p* < .001), and a weak positive correlation with Post Test Scores (*r* = .292, *p* = .010). Feedback Difficulty had a moderate positive correlation with Satisfaction of Post Test Performance (*r* = .505, *p* < .001), a weak positive correlation with Satisfaction of Material Learned (*r* = .235, *p* = .039), a moderate positive correlation with Formative Feedback Test Scores (*r* = .460, *p* < .001), and a moderate positive correlation with Post Test Scores (*r* = .626, *p* < .001).

Satisfaction of Post Test Performance had a weak positive correlation with Satisfaction of Material Learned (*r* = .303, *p* = .007), a weak positive correlation with Pre Test Scores (*r* = .227, *p* = .047), a moderate positive correlation with Formative Feedback Test Scores (*r* = .590, *p* < .001), and a moderate positive correlation with Post Test Scores (*r* = .623, *p* < .001). Satisfaction with Material Learned had a weak positive correlation with Post Test Scores (*r* = .398, *p* < .001).

A multiple regression was conducted to examine if Topic Difficulty, Post-Survey Knowledge, and Feedback Difficulty scores were predictors of Formative Feedback Test Scores. This was conducted in order to see if these factors could explain how participants performed in the feedback phase. The results indicated that the regression model was significant, F(3, 76) = 9.689, p < .001, and explained 25.5% of the variance in Formative Feedback Test , Adjusted R^2 = .225. The results also indicated that Topic Difficulty ratings significantly predicted Formative Feedback Test Scores, β = 0.288, t(73) = 2.412, p = .018. Additionally, Feedback Difficulty ratings significantly predicted Formative Feedback Test Scores $\beta = 0.298$, $t(73) = 2.556$, $p = .013$. Post-Survey Knowledge did not significantly predict Formative Feedback Test Scores $\beta = 0.055$, $t(73) = 0.484$, $p = .630$.

A multiple regression was conducted to examine if Topic Difficulty, Post-Survey Knowledge, Feedback Helpfulness, Feedback Difficulty, Satisfaction of Learned Material, and Satisfaction of Post Test Performance ratings were predictors of Post Test Scores. This was conducted in order to determine if these factors might explain how students performed in their final assessment. The results indicated that the regression model was significant, $F(6, 76) = 14.619$, $p < .001$, and explained 51.8% of the variance in Post Test , Adjusted $R^2 = .518$. The results also indicated that Satisfaction of Post Test Performance ratings significantly predicted Post Test Scores, $\beta = 0.345$, $t(70) = 3.194$, $p = .002$. Similarly, Satisfaction of Learned Material ratings significantly predicted Post Test Scores $\beta = 0.213$, $t(70) = 2.344$, $p = .022$. Additionally, Feedback Difficulty ratings significantly predicted Post Test Scores $\beta = 0.408$, $t(70) = 3.919$, $p < .001$. Topic Difficulty ratings did not significantly predict Post Test Scores $\beta = 0.048$, $t(70) = 0.438$, $p = .663$. Prior Knowledge did not significantly predict Post Test Scores $\beta = -0.026$, $t(70) = -0.277$, $p = .782$. Feedback Helpfulness did not significantly predict Post Test Scores $\beta = -0.048$, $t(70) = -0.494$, $p = .623$.

A simple linear regression was conducted to examine if Prior Knowledge is a predictor of Pre Test Scores. This was conducted in order to confirm if the baseline assessment for the participants could be explained by their prior experience with the topics. The results indicated that Prior Knowledge scores significantly predicted Pre Test Scores, $\beta = 0.440$, $t(77) = 8.191$, $p < .001$. The regression model was significant, $F(1, 77) = 18.006$, $p < .001$, and explained 19.4% of the variance in Pre Test scores, $R^2 = .194$.

In regards to preferences towards the type of feedback they would like to receive, 49.4% of the participants preferred non-interactive feedback, 19.5% of the participants preferred low-interactive feedback, and 31.2% participants preferred highly-interactive feedback. In the Non-Interactive condition, 61.5% of these participants preferred non-interactive feedback, 20.5% preferred low-interactive feedback, and 17.9% preferred highly-interactive feedback. In the Interactive condition, 36.8% of these participants preferred non-interactive feedback, 18.4% preferred low-interactive feedback, and 44.7% preferred highly-interactive feedback. A chi-square test of independence revealed a significant association between feedback condition and preferred feedback style, $\chi^2(2, N = 77) = 6.853$, $p = 0.032$.

Feedback Style Preferences were also examined by grouping participants into low scorers (below average) and high scorers (above average) for the Pre Test, Formative Feedback Test, and Post Test scores (see Figure 3). The Feedback Style responses were divided into Non-Interactive Feedback and Interactive Feedback (which was a combination of Low- and High- Interactive Feedback). In regards to Pre Test scores, 47.1% of the Low Scorers preferred non-interactive feedback and 52.9% preferred interactive feedback. For the High Scorers, 50.0% of these participants preferred non-interactive feedback and 50.0% preferred interactive feedback. A chi-square test of independence revealed that there was not a significant association between Pre Test Score group and preferred feedback style, $\chi^2(1, N = 78) = 0.066$, $p = 0.797$.

In regards to Formative Feedback Test scores, 33.3% of the Low Scorers preferred non-interactive feedback and 66.6% preferred interactive feedback. For the High Scorers, 56.9% of these participants preferred non-interactive feedback and 43.1% preferred interactive feedback. A chi-square test of independence revealed a significant association between Formative Feedback Test Score group and preferred feedback style, $\chi^2(1, N = 78) = 3.912$, $p = 0.048$.

In regards to Post Test scores, 34.5% of the Low Scorers preferred non-interactive feedback and 65.5% preferred interactive feedback. For the High Scorers, 57.1% of these participants preferred non-interactive feedback and 42.9% preferred interactive feedback. A chi-square test of independence revealed that there was not a significant association between Post Test Score group and preferred feedback style, $\chi^2(1, N = 78) = 3.744$, $p = 0.053$.

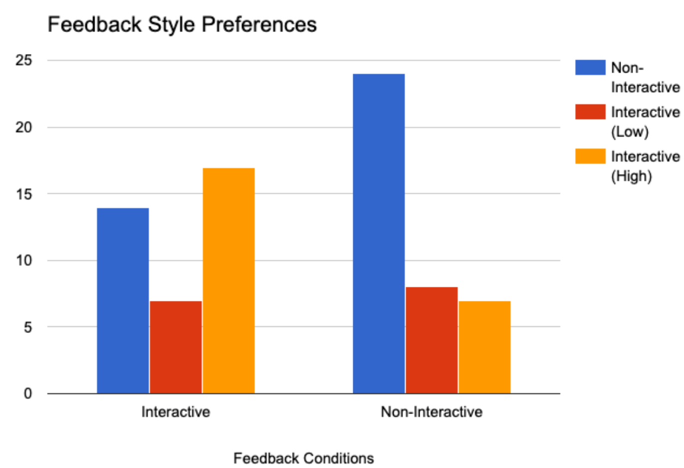


Figure 2: Feedback Interactivity Preferences by Condition

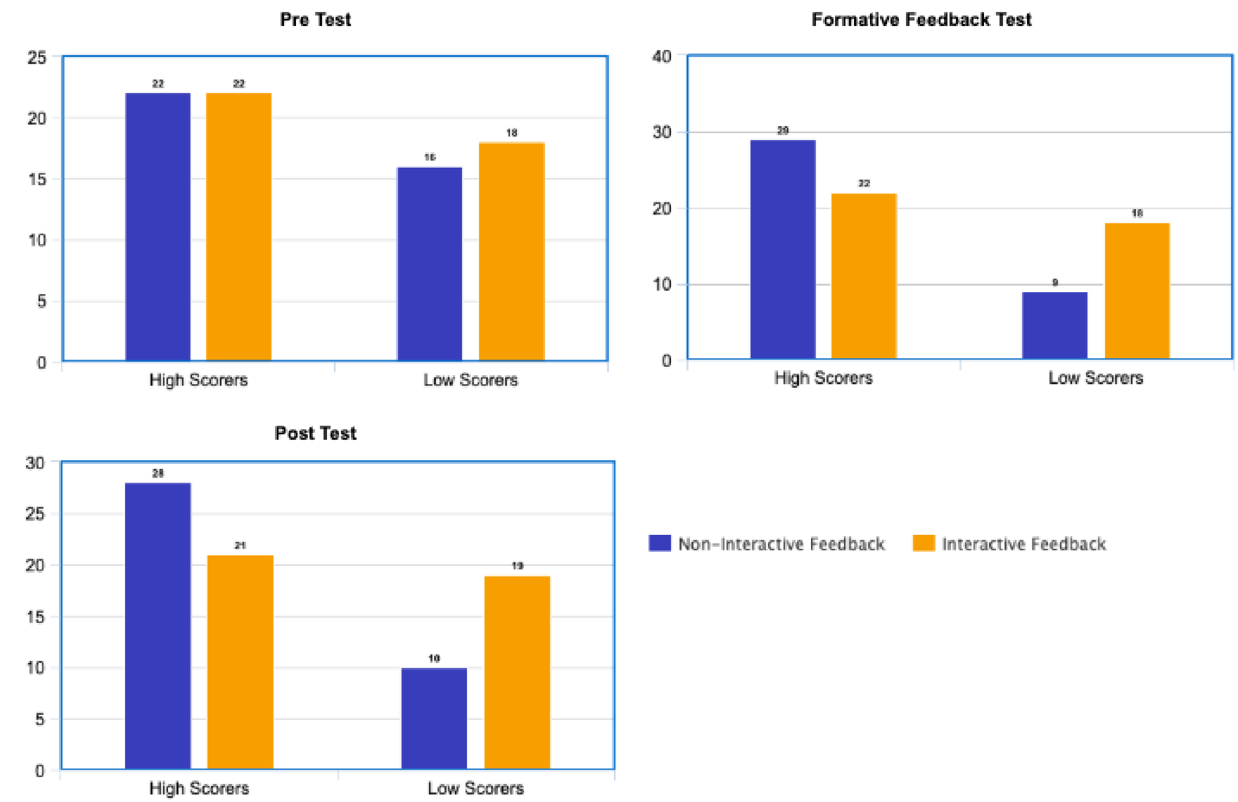


Figure 3: Feedback Interactivity Preferences by Assessment Scores

6.5 Cognitive Load

The composite scores for the participants' perceptions of their cognitive load were out of a scale of five. Cognitive Load data was collected three times throughout the study. The average cognitive load score at T1 is $m = 3.74$ ($SD = 0.975$), at T2 is $m = 2.52$ ($SD = 1.176$), and at T3 is $m = 2.27$ ($SD = 1.275$). The data meets the assumption of normality due to the central limit theorem, in that there are greater than 30 participants in each condition. The homogeneity of variances were assessed using Levene's test. The results revealed that there were equal variances between the groups at T1 [$F(1, 77) = 0.047$, $p = .829$], T2 [$F(1, 77) = 2.058$, $p = .155$], and T3 [$F(1, 75) = 2.175$, $p = .144$].

Table 13. Cognitive Load Descriptive Statistics by Condition

	Cognitive Load Time 1				Cognitive Load Time 2				Cognitive Load Time 3			
	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis	Mean	SD	Skew	Kurtosis
Interactive	3.821	0.933	-0.578	-0.340	2.581	1.083	0.650	-0.037	2.026	1.187	1.411	1.241
Non-Interactive	3.659	1.020	-0.949	0.997	2.450	1.271	0.506	-1.002	2.513	1.326	0.532	-0.880
Overall	3.738	0.975	-0.792	0.469	2.515	1.176	0.525	-0.660	2.273	1.275	0.893	-0.304

Note: Cogntive Load scores used the average of n=3 items on a scale from 1-5

A two-tailed Student' *T*-test was conducted to compare the means of the Cognitive Load Scores at T1 between the Interactive condition ($m = 3.82$, $SD = 0.933$) and the Non-Interactive condition ($m = 3.66$, $SD = 1.020$). This was conducted in order to confirm that participants did not differ in baseline self-ratings of their cognitive strain after the pre-test. The results revealed that there was a not significant difference between groups $t(77) = 0.736$, $p = 0.464$. A one-tailed Student' *T*-test was conducted to compare the means of the Cognitive Load Scores at T2 between the Interactive condition ($m = 2.58$, $SD = 1.083$) and the Non-Interactive condition ($m = 2.45$, $SD = 1.271$). The results revealed that there was a not significant difference between groups $t(77) = 0.492$, $p = 0.312$. A one-tailed Student' *T*-test was conducted to compare the means of the Cognitive Load Scores at T3 between the Interactive condition ($m = 2.03$, $SD = 1.187$) and the Non-Interactive condition ($m = 2.51$, $SD = 1.326$). This was conducted in order to determine if the participants in the interactive condition reported less strain on their cognitive load compared to participants in the non-interactive condition due to the feedback they engaged with. The results revealed that there was a significant difference between groups $t(75) = -1.697$, $p = 0.047$.

Pearson correlation tests were conducted between the Cognitive Load measures, the Pre Test scores, Formative Feedback Test Scores, Post Test Scores, and Post-Pre Score Deltas. These correlations were conducted in order to determine if there were any trends that would inform further analyses. Cognitive Load at T1 had a moderate negative correlation with Pre Test Scores (r = -.429, p < .001). Cognitive Load at T2 had a strong negative correlation with Formative Feedback Test Scores (r = -.674, p < .001), a strong negative correlation with Post Test Scores (r = -.583, p < .001), and a weak negative correlation with Post-Pre Delta Scores (r = -.256, p = .024). Cognitive Load at T3 had a strong negative correlation with Post Test Scores (r = -.662, p < .001), and a weak negative correlation with Post-Pre Delta Scores (r = -.347, p = .002).

A simple linear regression was conducted to examine if Cognitive Load at T1 is a predictor of Pre Test Scores. The results indicated that Cognitive Load at T1 significantly predicted Pre Test Scores, β = -0.429, t(77) = 10.962, p < .001. The regression model was significant, F(1, 77) = 17.351, p < .001, and explained 18.4% of the variance in Pre Test scores, R^2 = .184.

A simple linear regression was conducted to examine if Cognitive Load at T2 is a predictor of Formative Feedback Test. The results indicated that Cognitive Load at T2 significantly predicted Formative Feedback Test Scores, β = -0.674, t(77) = 26.676, p < .001. The regression model was significant, F(1, 77) = 63.955, p < .001, and explained 45.4% of the variance in Formative Feedback Test scores, R^2 = .454.

A multiple regression was conducted to examine if Cognitive Load scores at T2 and T3 were predictors of Post Test Scores. This was conducted in order to determine if the reported strain on cognitive load from the previous two phases could serve as an explanation for how participants performed in their final assessment. The results indicated that the regression model was significant, F(2, 76) = 29.877, p < .001, and explained 43.2% of the variance in Post Test , Adjusted R^2 = .432. The results also indicated that Cognitive Load scores at T3 significantly predicted Post Test Scores, $\beta = -0.531$, $t(74) = -3.572$, $p < .001$. Cognitive Load scores at T2 did not significantly predict PostTest Scores $\beta = -0.161$, $t(73) = -1.084$, $p = .282$.

A simple linear regression was conducted to examine if Cognitive Load at T3 is a predictor of Post-Pre Delta Scores. The results indicated that Cognitive Load at T3 significantly predicted Post-Pre Delta Scores, $\beta = -0.347$, $t(77) = 6.903$, $p < .001$. The regression model was significant, $F(1, 77) = 10.247$, $p = .002$, and explained 12.0% of the variance in Post-Pre Delta scores, $R^2 = .120$.

Chapter 7: Conclusion

The findings of this study reveal complex and multidirectional responses to the initial research questions, reflecting both the potential benefits and limitations of feedback mechanisms in educational assessment. While certain aspects of the results appear to be somewhat contradictory, these inconsistencies align with prior literature, underscoring the nuanced nature of feedback efficacy in learning contexts. Such findings highlight the need for further investigation into the specific conditions under which various forms of feedback yield optimal learning outcomes.

7.1 Proportion Correct

One of the most significant findings of this study pertains to the role of feedback in enhancing learning, as measured by assessment scores. Regardless of whether participants received interactive or non-interactive feedback, their overall scores improved markedly during the feedback phases. However, a notable stagnation was observed following the formative feedback test, with little additional progress in the Post Test phase. This plateau suggests a potential ceiling effect in the assessments, where participants had already maximized their performance within the scope of the test parameters. Specifically, by the formative feedback phase, 20% of the test questions had already reached a proportion of correct answers above 85%. By the Post Test phase, this proportion had doubled to 40%, indicating that a substantial subset of participants may have encountered a performance ceiling that constrained their ability to demonstrate further measurable growth. However it may be that much of the improvement in the scores may be expected to a degree due to the feedback provided. Even so, this limitation in test design underscores the importance of developing assessments that allow for continued differentiation of learning gains beyond a predefined threshold of mastery. It may be beneficial for follow-up studies to be conducted using more difficult questions in order to reduce the likelihood of concerns related to a ceiling effect. Adaptive testing strategies or more complex problem-solving tasks may provide a more sensitive measure of continued learning.

Another possible explanation for the observed performance trends is the influence of a testing or practice effect. Given that participants were repeatedly exposed to similar question formats and content — focused specifically on two similar mathematical formulas — improvements in scores may have been partially attributable to repeated exposure rather than genuine cognitive gains. This concern is supported by pilot study findings, wherein the control group that did not receive feedback also exhibited increases in scores between Pre Test and Post Test phases. However, the increases found in both the interactive and non-interactive feedback in this study were much greater than the increase found in the control group from the pilot study. Future research should consider incorporating alternative question formats and varied problem-solving contexts to minimize this confounding factor and provide a more robust measure of conceptual understanding. The presence of a practice effect highlights the need for diversifying assessment formats to differentiate genuine conceptual mastery from familiarity-based performance improvements.

A particularly unexpected finding was the absence of significant differences in test scores between participants in the interactive and non-interactive feedback conditions. At no stage of assessment did one condition yield superior results over the other, nor were there notable differences in the magnitude of score improvements between groups. These findings challenge existing literature suggesting the superiority of interactive feedback (Van der Kleij et al., 2015) and instead align with studies reporting mixed or inconclusive effects of interactivity in feedback delivery (Mandernach, 2005; Slowaik & Lakowske, 2017; Van der Kleij et al., 2012). The lack of differentiation between conditions may suggest that the effectiveness of feedback is not solely contingent upon its interactivity but rather on other factors such as the learners' engagement with the feedback process. However it is also important to note that although there were not significant differences in Post Test scores between the two conditions, the baseline scores of the interactive feedback participants were slightly lower than the non-interactive participants, which indicated that they had a slightly greater growth compared to the non-interactive participants in order to reach equal post test scores.

Despite the overall general equivalence between feedback conditions, additional analysis identified a key factor influencing participant performance: proportion of correct responses to feedback prompts. Participants in the interactive feedback condition, who were required to answer multiple-choice and fill-in-the-blank questions as part of their feedback process, exhibited a significant correlation between Feedback Score and subsequent test performance. Those who provided more accurate responses within the feedback phase also tended to achieve higher scores on both the formative feedback test and the Post Test. This finding implies that a deeper engagement with feedback—rather than the interactivity of the medium itself—may be the critical determinant of learning outcomes. It suggests that students who are able to correctly process and apply feedback-related content gain a more comprehensive understanding of the material, ultimately enhancing their performance. Interestingly, this may also conflict with findings related to feedback duration, which will be discussed as well.

7.2 Reception

The present study extends beyond the measurement of student growth through assessment performance by also considering students' self-perceptions of their learning and comprehension. The findings indicate that participants, on average, reported a notable improvement in their understanding of the material after completing the study. This self-reported gain in knowledge aligns with the observed increase in assessment scores between the preliminary and Post Tests, suggesting a cohesive relationship between self-awareness in learning and actual performance improvements. These findings contribute to the growing body of research highlighting the connection between cognitive development and metacognitive self-perception, further emphasizing the importance of incorporating student self-reflection as a metric of learning efficacy.

An additional dimension of this study involved examining differences in how participants in the interactive and non-interactive feedback conditions perceived the content and their experiences with the feedback mechanisms. Interestingly, there were minimal differences between the two conditions across most factors analyzed. Participants from both groups reported similar perceptions regarding the overall difficulty of the topic, suggesting that the type of feedback received did not significantly influence their perceived complexity of the subject matter. Furthermore, both groups provided comparable ratings regarding the difficulty of using their respective feedback systems. This finding is particularly noteworthy, as a key concern was that the interactive condition might be viewed as more challenging due to the additional cognitive effort required to engage with the feedback. However, the data suggest that students did not find interactive feedback more demanding despite its higher level of engagement.

Another notable finding was the similarity in self-assessed baseline knowledge across both groups. This result was anticipated, as participants had no prior exposure to the material before engaging in the study. More importantly, both groups also reported equivalent levels of perceived knowledge after completing their respective feedback conditions. This outcome aligns with the Post Test results, which showed no significant differences between the interactive and non-interactive feedback groups in terms of assessment performance. The consistency between self-reported understanding and actual test scores reinforces the notion that both feedback mechanisms were equally effective in fostering student learning, at least within the measured parameters. These findings suggest that students, regardless of the feedback type they receive, perceive their final level of comprehension at comparable levels, supporting the idea that engagement style alone may not necessarily influence perceived mastery.

In addition to performance-based assessments and knowledge ratings, this study also explored participants' satisfaction with their learning experience. Two distinct dimensions of satisfaction were examined: mastery-oriented satisfaction (contentment with overall learning) and performance-oriented satisfaction (contentment with test performance relative to peers). Results indicate no significant differences between the interactive and non-interactive feedback groups in either of these satisfaction measures. This suggests that students across both conditions were equally satisfied with their perceived mastery of the topic and their comparative performance. These findings imply that regardless of feedback interactivity, students experienced similar levels of confidence and contentment with their educational outcomes, reinforcing the comparable effectiveness of both feedback approaches.

However, a key point of divergence emerged in students' perceptions of feedback usefulness. The participants in the interactive feedback condition felt that the feedback was more helpful compared to the participants in the non-interactive feedback condition. This suggests that the greater engagement required by interactive feedback may positively influence students' perceptions of its utility. Interestingly, despite this higher perceived usefulness, the interactive feedback did not yield superior performance outcomes compared to non-interactive feedback. This discrepancy highlights a crucial psychological aspect of feedback implementation: even if different feedback modalities produce similar learning outcomes, students may still exhibit preferences for certain types of feedback based on perceived utility rather than objective effectiveness.

These findings have significant implications for educational practice and the design of feedback mechanisms. Prior research (Timmers & Veldkamp, 2011; Bailey & Garner, 2010; Duncan, 2007) has consistently reported that students often disregard feedback, regardless of its actual effectiveness. Given this tendency, understanding how to make feedback more engaging and approachable is essential for enhancing student uptake and application of feedback. The fact that participants in the interactive feedback condition perceived their feedback as more helpful, despite its lack of measurable advantage in performance, suggests that engagement strategies play a crucial role in how students interact with and value feedback. If students are more likely to utilize feedback they perceive as useful, even in the absence of significant learning gains, then increasing the perceived helpfulness of feedback could be an essential strategy for promoting feedback engagement.

Future studies should explore additional variables that influence the perceived value of feedback, such as personalization, adaptability, and multimodal delivery methods. Investigating how factors like motivation and prior learning experiences interact with feedback perceptions may provide deeper insights into optimizing feedback strategies. Moreover, long-term studies examining whether heightened perceptions of feedback usefulness translate into better knowledge retention over extended periods would be valuable.

This study highlights the importance of considering not only objective performance outcomes but also students' self-perceptions of their learning experiences. While interactive and non-interactive feedback methods resulted in comparable learning gains and self-reported knowledge levels, participants in the interactive condition perceived their feedback as more useful. This suggests that engagement strategies play a critical role in shaping students' attitudes toward feedback, which in turn may influence their willingness to engage with it. Given that prior research has demonstrated that students frequently ignore feedback, designing feedback systems that students perceive as valuable—even in the absence of superior performance outcomes—could be a key factor in improving educational effectiveness. Moving forward, further research should explore how feedback delivery can be optimized to maximize both objective learning gains and subjective engagement, ultimately fostering a more effective and student-centered learning environment.

Beyond the analysis of student performance and perceptions of feedback effectiveness, this study also uncovered significant relationships between participants' self-reported perceptions and their actual performance outcomes. While many subjective preferences and evaluations of the feedback mechanisms did not differ significantly between the interactive and non-interactive conditions, several notable trends emerged. These trends reveal important connections between self-perceived difficulty, baseline knowledge, performance outcomes, and overall satisfaction, offering deeper insights into how students interact with learning materials and feedback.

One of the most striking findings was the correlation between participants' perceived difficulty of the learning material and their perception of feedback difficulty. Participants who rated the subject matter as challenging were also more likely to report that the feedback itself was difficult to use. This pattern suggests that individuals who struggle with the core content may experience additional cognitive load when engaging with feedback, potentially limiting its effectiveness. This finding aligns with cognitive load theory (Sweller, 1988), which posits that learners with lower domain-specific knowledge may struggle more when processing additional instructional material. It also suggests that making feedback more adaptive and accessible for students with lower baseline knowledge may be a crucial step in optimizing its effectiveness.

Similarly, this trend was observed in students' self-assessments of prior and post-study knowledge. Those who perceived the topic as easier not only tended to rate their baseline knowledge higher but also reported greater confidence in their understanding after completing the study. This could indicate that individuals with a more intuitive grasp of the material are more adept at recognizing and internalizing learning gains. Conversely, students who struggled with the material may underestimate their knowledge due to lower self-efficacy, a phenomenon widely studied in the literature on metacognition and self-regulated learning (Dunning *et al.*, 2003). These insights highlight the importance of designing feedback mechanisms that not only present corrective information but also support students' self-efficacy and confidence in their learning progression.

Beyond self-perceptions, perceived difficulty also demonstrated clear associations with actual performance metrics. Participants who found the topic easier generally scored higher on the pretest, which is consistent with their higher self-reported baseline knowledge. This correlation suggests that prior knowledge plays a significant role in shaping perceptions of difficulty and ultimately in determining success on initial assessments. Furthermore, the predictive nature of difficulty ratings extended beyond the pretest phase, as students who perceived the material as more challenging also tended to perform worse on the formative feedback test and the Post Test. These findings reinforce the well-documented impact of prior knowledge on learning outcomes (Chi, Feltovich, & Glaser, 1981) and suggest that early interventions targeting struggling students may be beneficial in closing performance gaps.

Additionally, perceived difficulty was linked to students' satisfaction with their performance. Those who found the material to be more accessible expressed greater satisfaction with their assessment outcomes, indicating a potential psychological component to academic success. This could reflect a positive feedback loop in which students who feel confident in their abilities experience increased motivation and engagement, leading to better performance and greater satisfaction. Conversely, students who struggle with the material may experience frustration or disengagement, which could negatively impact both their learning trajectory and their perception of the feedback's usefulness. Future research should explore whether targeted interventions, such as scaffolding or personalized support, could mitigate these effects and enhance student satisfaction across different levels of prior knowledge.

The findings from this study highlight the complex relationship between self-perception, performance, and satisfaction in an educational setting. The fact that perceived difficulty serves as a predictor of both test scores and subjective learning experiences suggests that instructional designers should consider strategies to better support students who initially struggle with content. One potential avenue for improvement is the implementation of adaptive feedback systems that adjust to students' competency levels, providing more structured guidance to those who find the material challenging while offering more autonomy to those who grasp it with ease. Furthermore, the observed relationships between prior knowledge, difficulty ratings, and satisfaction suggest that increasing student awareness of their own learning progress could be a valuable instructional strategy. Techniques such as formative self-assessment, metacognitive training, and structured reflection activities may help students develop a more accurate understanding of their learning trajectory and reduce the likelihood of self-doubt undermining performance.

Given that students who perceived the feedback as more difficult also tended to struggle more with the material itself, future studies should examine whether simplifying feedback presentation—such as through multimodal formats (e.g., video explanations, step-by-step interactive guides)—could enhance accessibility for lower-performing students. Additionally, research could investigate whether increasing student autonomy in selecting feedback formats tailored to their learning preferences might improve engagement and outcomes.

Another key insight from this study is the observed link between students' perceptions of feedback usability and their perceptions of its helpfulness. Specifically, participants who found the feedback easier to use also rated it as more helpful. This finding underscores a persistent challenge in educational feedback design: the potential disconnect between the intended benefits of feedback and students' ability to effectively utilize it. If learners struggle to comprehend or engage with feedback, they may not perceive its value, leading them to disregard it entirely. This aligns with prior research on feedback literacy, which highlights the necessity of equipping students with the skills needed to interpret and apply feedback effectively (Carless & Boud, 2018). Future interventions should consider incorporating scaffolding techniques or explicit instructional guidance to ensure that students are able to maximize the benefits of feedback mechanisms, especially those who may initially find them challenging to use.

Additionally, the findings indicate that students who rated the feedback as more helpful also expressed higher levels of satisfaction with their Post Test performance and overall learning experience. This suggests that beyond its role in facilitating objective learning gains, feedback also influences students' motivation and perceived competence. The implications of this are particularly relevant for fostering a positive learning experience, as prior research has shown that learners who feel more satisfied with their educational progress are more likely to remain engaged and motivated (Ryan & Deci, 2000). Even if two feedback conditions yield similar Post Test scores, differences in how learners experience and interpret feedback can have long-term consequences on their willingness to seek and apply feedback in future learning contexts. As such, ensuring that feedback is both accessible and perceived as beneficial may be a crucial factor in enhancing student engagement with educational content.

Beyond influencing subjective learning experiences, feedback usability also had a measurable impact on objective performance outcomes. Participants who found the feedback to be more helpful demonstrated higher Post Test scores, suggesting that positive engagement with feedback mechanisms translates into improved academic performance. This finding reinforces the notion that students' attitudes toward feedback are not merely reflections of preference but may also serve as indicators of their ability to effectively integrate the provided guidance into their learning process. It raises important questions about how instructional designers can better structure feedback to maximize its effectiveness, particularly for students who may struggle with initial engagement.

Similarly, perceptions of feedback difficulty were also associated with both satisfaction and performance. Participants who found the feedback easier to use reported greater satisfaction with their Post Test performance and mastery of the material. Furthermore, these usability ratings served as predictors of both formative feedback test scores and Post Test scores. This once again highlights the dual role of feedback in shaping not only students' affective experiences but also their objective learning trajectories. Given these results, it may be beneficial to explore strategies for enhancing feedback accessibility, such as incorporating multimodal feedback options (e.g., visual explanations, step-by-step breakdowns) to accommodate diverse learning preferences.

This study highlights the need for future research to explore the long-term effects of feedback perceptions on student learning. While immediate Post Test scores provide valuable insights, investigating how students' engagement with feedback influences their learning habits over time could yield deeper insights into best practices for instructional design. For instance, do students who initially struggle with feedback usability develop better feedback literacy skills with continued exposure? Does perceived difficulty in feedback discourage engagement with future learning opportunities? Addressing these questions could provide actionable recommendations for refining feedback mechanisms to better support diverse learner needs.

One limitation of this study is that participants were only exposed to a single type of feedback, either interactive or non-interactive. Consequently, their perceptions and ratings of the feedback were formed exclusively through their experience within one condition, preventing them from making direct comparisons between the two modalities. Without first hand exposure to both feedback conditions, their evaluations might not fully capture potential differences in effectiveness, usability, or engagement across conditions. This limitation should be considered when interpreting findings regarding participant preferences and performance outcomes. Despite this constraint, participants were provided with illustrative examples of all feedback types at the end of the study. This included not only the interactive and non-interactive feedback they had encountered but also a third, intermediary form of mid-interactive feedback that incorporated elements of both conditions. Following this exposure, participants were asked which type of feedback they would prefer to receive in future learning scenarios, yielding notable findings on student preferences and engagement.

A majority of participants (49.4%) expressed a preference for entirely non-interactive feedback, while 31.2% preferred fully interactive feedback, and a smaller subset (19.5%) favored the intermediary level of mid-interactive feedback. These results indicate a strong inclination toward non-interactive feedback, a preference that aligns with prior research suggesting that students often gravitate toward learning methods that require less effort. The interactive feedback condition required students to actively respond to multiple-choice and fill-in-the-blank questions, which inherently demands more cognitive engagement. In contrast, non-interactive feedback allowed for passive consumption, potentially making it more appealing to learners seeking efficiency or convenience over active engagement. However, it is also important to consider that combining the frequencies for both the mid-interactive and high-interactive feedback preferences reaches a total of 50.6%, which is almost equal to the amount who preferred the non-interactive feedback. This seems to demonstrate that although in general participants still prefer the non-interactive feedback, there may be a large number of students who would be open to some level of interactivity over a completely non-interactive option.

This finding is particularly intriguing given the earlier results indicating that participants exposed to interactive feedback perceived it as more helpful than those in the non-interactive condition. This discrepancy highlights a common challenge in educational research: students may recognize the benefits of an instructional strategy but still prefer an alternative that requires less effort. The cognitive load associated with interactive feedback, while potentially beneficial for deep learning and retention, may discourage students from voluntarily choosing it as their preferred mode of instruction. This underscores the importance of balancing effectiveness with student motivation when designing feedback mechanisms in educational settings.

A more nuanced analysis of participant preferences reveals an interesting divide between those who experienced different feedback conditions. Among participants in the non-interactive feedback condition, 61.5% indicated a preference for continuing with non-interactive feedback, while only 36.8% of those in the interactive feedback condition preferred to switch to non-interactive feedback. This suggests that students who had firsthand experience with interactive feedback may have better understood its benefits and were less inclined to revert to a purely passive learning method.

Additionally, only 17.9% of participants in the non-interactive feedback condition expressed a desire to switch to interactive feedback, whereas 44.7% of those in the interactive feedback condition preferred to continue using interactive feedback. This indicates that once students had adapted to a feedback format that required greater cognitive effort, they were more willing to persist with it, likely due to their recognition of its perceived usefulness. Conversely, students initially exposed to the less demanding non-interactive feedback format were less inclined to voluntarily adopt a more effort-intensive approach. This pattern suggests that initial exposure and adaptation play a crucial role in shaping student perceptions of learning tools.

A further dimension of our analysis involved investigating the relationship between students' assessment performance and their expressed preferences for varying feedback styles. Specifically, this inquiry sought to determine whether students' choices regarding the degree of interactivity in feedback—classified by combining both the mid- and high-interactivity response options into a single category representing a preference for some level of interactivity—correlated meaningfully with their performance outcomes. This approach provided a clearer lens through which to assess the impact of assessment success on feedback preferences, particularly when contrasted against a baseline preference for non-interactive feedback mechanisms.

At the outset, when considering participants' baseline performance as measured by pre-test scores, the distribution of feedback preferences appeared evenly divided. Approximately half of the participants expressed a preference for feedback that included some level of interactivity, while the other half opted for non-interactive feedback. This parity suggests that initial competence or familiarity with the subject matter did not significantly predispose students to favor one feedback style over another at the beginning of the instructional sequence.

However, when we extended the analysis to include performance data from both the formative feedback assessment and the subsequent post-test, a more nuanced pattern emerged. Among participants who achieved high scores on either the formative or the post-instruction assessment, preferences remained relatively balanced between those favoring interactive feedback and those preferring no interaction. This finding implies that students who had successfully assimilated the instructional content, irrespective of the feedback style provided, did not exhibit a pronounced inclination toward investing additional cognitive effort in engaging with more elaborative or interactive forms of feedback.

In contrast, a discernible trend was observed among participants who performed poorly on these later assessments. This subgroup demonstrated a stronger preference for feedback incorporating some degree of interactivity. Their selection suggests that students encountering difficulty with the material perceived greater utility in feedback formats that encouraged active engagement and elaboration. This observation aligns with prior findings from Smits et al. (2008), who reported that learners with higher prior knowledge on a given topic tend to prefer less elaborative, and often more concise, feedback. Conversely, individuals with lower prior knowledge typically exhibit a preference for feedback that is more detailed and interactive, as it offers additional scaffolding to support comprehension and knowledge construction.

What distinguishes the present findings, however, is that the differentiation in feedback preferences appears to be influenced not by prior knowledge as assessed by initial pre-test performance, but rather by students' subsequent performance outcomes during the course of the learning experience. This distinction suggests that students may recalibrate their feedback preferences in response to their perceived competence or success with the material, rather than their initial familiarity with it. Specifically, participants who demonstrated a firm grasp of the subject matter appeared disinclined to devote additional cognitive resources to engaging with interactive feedback when it was no longer perceived as necessary for their learning progress. In contrast, those who struggled to achieve satisfactory assessment outcomes increasingly valued the opportunity for more interactive and elaborative feedback, likely viewing it as a critical resource for addressing learning gaps and enhancing their understanding of the content.

This pattern highlights the dynamic nature of learner preferences and underscores the importance of adopting flexible feedback systems that can be responsive to students' evolving needs over time. It also suggests avenues for future research exploring how students' feedback preferences might shift in response to their experiences within a learning environment, and how instructional designs can accommodate these changing preferences to optimize learning outcomes.

Overall, these findings reveal a complex relationship between perceived usefulness, ratings of cognitive effort, and student preference. While interactive feedback was regarded as more helpful, students' reluctance to engage with it voluntarily highlights the need for strategies that promote engagement without overwhelming learners. Future research should explore ways to scaffold interactive feedback more effectively, perhaps by gradually increasing interactivity or integrating motivational incentives to encourage sustained engagement. Understanding these dynamics is crucial for optimizing feedback interventions in educational settings, ensuring they are both effective and appealing to students.

7.3 Duration

The role of duration in student performance and engagement with feedback presents a crucial component of the learning process that warrants further investigation. In this study, we observed a consistent decrease in the amount of time participants required to complete each successive assessment, a trend that was evident across both feedback conditions. Specifically, participants required less time to complete the formative feedback test compared to the pretest, and subsequently, even less time was needed to complete the Post Test. This pattern suggests that as participants progressed through the study, they developed a more comprehensive understanding of the material and became more familiar with the cognitive processes required to navigate the questions efficiently. The reduction in completion time may reflect improvements in problem-solving strategies, an increased level of comfort with the test format, or a deeper internalization of the mathematical concepts being assessed.

Since this decreasing duration was observed in both the interactive and non-interactive feedback conditions, it implies that the effect is not necessarily dependent on the type of feedback received. Rather, it may be attributed to repeated exposure to the test material, reinforcing the notion that practice plays a fundamental role in learning outcomes. This aligns with existing research on the testing effect, which posits that repeated retrieval practice enhances learning and retention (Roediger & Butler, 2011). However, while decreased duration may indicate improved efficiency in problem-solving, it is also possible that participants were responding more quickly due to increased test familiarity rather than a genuine deepening of conceptual understanding. This distinction is crucial, as shorter completion times do not always equate to better comprehension or mastery of the material.

Another possible explanation for the reduction in test duration is the development of procedural fluency, wherein participants become adept at executing the necessary steps to arrive at correct answers without excessive cognitive strain. This could suggest that while participants may not have significantly improved in conceptual understanding — especially given the absence of major differences in Post Test scores between conditions — they may have become more proficient at navigating the assessments through pattern recognition and repetition. If this is the case, it raises important pedagogical considerations regarding the design of learning interventions that prioritize not just efficiency but also deeper cognitive engagement with the content.

Additionally, this study raises questions about the role of cognitive load in influencing test duration and performance. Cognitive load theory (Sweller, 1988) suggests that when students first engage with new material, they experience a higher level of cognitive strain, which may initially lead to longer response times. As learning progresses and schema development occurs, cognitive load is reduced, and students can process information more efficiently. The observed decrease in test duration could therefore be an indicator of reduced extraneous cognitive load, allowing participants to focus more on applying knowledge rather than struggling with the test format itself.

Despite these promising insights, it is essential to consider individual differences in how students allocate their time during assessments. Some students may have become overconfident in their understanding, leading them to complete later assessments more quickly at the expense of their score. Others may have strategically optimized their test-taking strategies, using previous feedback to refine their approach. Future studies should incorporate qualitative measures, such as participant reflections or think-aloud protocols, to better understand how students perceive their time management and whether shorter durations reflect actual improvements in learning or mere test familiarity.

Ultimately, while the findings highlight a clear trend of decreasing test duration across assessments, the implications of this trend remain multifaceted. On one hand, it suggests that participants may have improved in efficiency, reflecting gains in procedural knowledge and test-taking fluency. On the other hand, it also raises concerns about the extent to which learning gains are genuine rather than a byproduct of repeated exposure. Further research is needed to distinguish between improvements in genuine understanding and mere familiarity with test content, as well as to explore how different feedback mechanisms may interact with time allocation and learning outcomes. By examining these factors, future studies can provide more nuanced recommendations for optimizing feedback strategies to enhance student engagement, comprehension, and long-term retention of educational material.

Similar to the findings regarding assessment scores between the interactive and non-interactive feedback conditions, there were no significant differences in the total time participants spent completing each section of the study. The total duration of participation was nearly identical across both groups, with only minor variations in pretest duration, formative feedback test duration, and Post Test duration. The interactive condition resulted in slightly shorter durations for assessments, a factor that was counterbalanced by the additional time participants spent engaging with the interactive feedback. Given that the interactive feedback required students to actively respond to embedded multiple-choice and fill-in-the-blank questions, this additional time commitment was expected. Conversely, in the non-interactive condition, participants had the flexibility to skim or entirely bypass the feedback, resulting in a more streamlined feedback phase.

While an increased duration in engagement with feedback could be justified if it led to significant improvements in Post Test performance, the present findings indicate that this was not necessarily the case. The additional time spent on interactive feedback did not yield higher assessment scores, suggesting that the increased cognitive engagement required for this form of feedback did not translate into measurable academic benefits. This challenges the assumption that greater exposure to feedback inherently leads to enhanced learning outcomes. In fact, there are indications that prolonged interaction with feedback might not only be inefficient but could also be counterproductive in certain scenarios.

One of the most striking findings was the inverse relationship between time spent on feedback and assessment performance. Students who devoted more time to engaging with feedback, irrespective of condition, tended to score lower on both the formative feedback test and the Post Test. Since this trend was not seen with the Pre Test scores, it highlights the relation between the engagement with feedback to the assessments during and after exposure. This suggests that prolonged engagement with feedback may not always equate to deeper comprehension or improved retention of the material. Instead, it may indicate that students struggling with the content were spending more time attempting to process the feedback but were still unable to fully grasp the concepts necessary for improved performance. This aligns with previous research demonstrating that excessive time spent on instructional material can sometimes signal confusion or cognitive overload rather than mastery (Sweller, 1988; Kirschner, Sweller, & Clark, 2006).

This trend was particularly evident within the interactive feedback condition, where increased time spent engaging with the feedback correlated with lower proportion of correct responses to the embedded multiple-choice and fill-in-the-blank questions. This finding raises concerns about the efficacy of interactive feedback mechanisms when students are unable to effectively utilize them to enhance understanding. Rather than acting as a tool for reinforcement and correction, interactive feedback may become an additional cognitive burden for students who are already struggling with the material. If learners are investing more time in interactive feedback yet still failing to improve their scores, this suggests that the design and delivery of feedback require further refinement to ensure accessibility and efficacy.

The implications of these findings extend beyond this particular study and align with broader research on feedback effectiveness. As mentioned before, prior studies have shown that feedback interventions can sometimes have unintended negative consequences, leading to decreased performance rather than improvement (Kluger & DeNisi, 1996). This occurs when feedback is either too complex, too difficult to interpret, or does not align with the student's current level of understanding. If feedback does not effectively scaffold learning or provide clear corrective guidance, students may struggle to extract meaningful insights, resulting in wasted effort and diminished motivation.

Given these considerations, it is imperative that instructional designers and educators critically evaluate not only the type of feedback provided but also the manner in which students engage with it. Simply increasing interactivity or requiring longer engagement with feedback is not necessarily beneficial if it does not facilitate improved comprehension and application of the material. Future research should explore strategies to optimize feedback mechanisms, ensuring that students can efficiently process and apply feedback without experiencing cognitive overload. Additionally, investigating individualized approaches, such as adaptive feedback that adjusts in complexity based on student performance, may provide a more effective way to support diverse learning needs.

Ultimately, the findings of this study highlight the complexity of feedback duration and its impact on learning outcomes. While interactive feedback offers the potential for increased engagement and deeper cognitive processing, its effectiveness is contingent upon students' ability to comprehend and utilize the provided information. The fact that greater time spent on feedback was associated with lower assessment performance suggests that the structure, clarity, and usability of feedback require careful consideration. If interactive feedback is to be a viable tool for enhancing student learning, it must be designed in a way that facilitates comprehension, minimizes unnecessary cognitive load, and ensures that additional time investment translates into meaningful academic gains.

7.4 Learner Characteristics

The characteristics of participants play a crucial role in shaping their engagement with assessments and, consequently, their overall learning outcomes. One of the most prominent factors influencing performance in this study was self-efficacy. The data revealed that self-efficacy served as a significant predictor of both Pre Test and Post Test scores, suggesting that students who exhibited greater confidence in their ability to succeed were more likely to perform well both at the baseline and final stages of assessment. This finding aligns with prior research indicating that self-efficacy is closely linked to academic achievement, as students with higher confidence tend to adopt more effective learning strategies, persist longer through challenges, and demonstrate greater motivation to engage with the material. Additionally, a strong sense of self-efficacy may facilitate a more open and receptive approach to learning, allowing students to absorb and apply new concepts more effectively. The consistency of this effect across both the initial and concluding phases of the study suggests that confidence not only influences students' readiness to engage with new material but also enhances their ability to process and retain information throughout the learning process.

Conversely, test anxiety was found to have a negative correlation with assessment performance across all testing phases, including the Pre Test, formative feedback test, and Post Test. This inverse relationship is consistent with previous findings in educational psychology, which suggest that heightened anxiety can impede cognitive functioning by diverting mental resources away from problem-solving and comprehension. When students experience excessive stress in a testing environment, they may struggle to recall relevant information, misinterpret questions, or second-guess their responses, all of which can contribute to lower scores. Since all three assessments in this study involved newly introduced mathematical material, the pressure of learning and demonstrating proficiency in a novel domain likely exacerbated the effects of test anxiety. This underscores the importance of implementing instructional strategies that mitigate anxiety, such as incorporating low-stakes formative assessments, providing structured guidance within feedback mechanisms, and fostering a supportive learning environment where students feel encouraged rather than pressured to succeed. Addressing test anxiety in the context of feedback design could be a critical consideration for future research and pedagogical applications.

Another key factor that emerged as a predictor of assessment performance was participants' intrinsic value in mathematics. Those who expressed a greater interest and appreciation for math demonstrated higher scores on both the Pre Test and Post Test. This suggests that an inherent motivation to engage with mathematical concepts may provide students with a foundational advantage, enabling them to approach new material with greater enthusiasm and attentiveness. Students who value mathematics intrinsically may be more adept at identifying patterns, leveraging prior knowledge, and employing problem-solving strategies that enhance their ability to grasp new concepts efficiently. The observed relationship between intrinsic math value and assessment scores is particularly noteworthy when considering its implications for instructional design. Encouraging students to develop a deeper appreciation for mathematical learning through real-world applications, interactive problem-solving, and personalized feedback could be instrumental in fostering more meaningful and sustained engagement with the subject.

Furthermore, within the interactive feedback condition, participants with higher intrinsic math value also demonstrated greater proportion of correct responses to embedded feedback questions. This suggests that students who already possess a strong affinity for mathematics may be more inclined to engage thoroughly with interactive learning tools, thereby maximizing the benefits of such interventions. Given this, it is essential to consider ways to enhance the appeal of interactive feedback mechanisms for students who may not inherently value mathematics as highly. Future research could explore how different framing techniques, gamified learning elements, or contextualized problem-solving approaches might make interactive feedback more engaging and accessible to a broader range of learners.

Taken together, these findings highlight the multifaceted role that individual learner characteristics play in shaping engagement with feedback and assessments. While self-efficacy fosters a proactive and confident approach to learning, test anxiety can act as a barrier to performance, necessitating the integration of supportive interventions to alleviate cognitive strain. Similarly, intrinsic interest in mathematics serves as a key driver of success, reinforcing the importance of cultivating motivation and curiosity within instructional practices. These insights underscore the need for personalized feedback mechanisms that accommodate diverse learner profiles, ensuring that feedback is not only informative but also psychologically supportive and motivationally enriching. As educators and researchers continue to refine feedback methodologies, it will be critical to consider these learner attributes to optimize the effectiveness of assessment-driven learning experiences.

7.5 Cognitive Load

The role of cognitive load in student learning and assessment performance presents a critical factor in the implementation and effectiveness of interactive feedback mechanisms. Cognitive load theory suggests that the way information is presented and the amount of mental effort required to process it can significantly impact learning outcomes. One of the primary concerns surrounding interactive feedback is its dual potential: it can reduce cognitive load by fostering deeper engagement with the material and reinforcing learning processes, yet it can also increase cognitive burden due to the additional effort required to interact with the feedback. The findings of this study provide valuable insights into how these opposing forces manifest in real learning scenarios. Data from this study indicate that there was no significant difference in the ratings of cognitive load reported by participants immediately after completing the pretest, which is to be expected given that all participants, regardless of condition, engaged in the same initial assessment. This consistency in cognitive load suggests that any subsequent differences in ratings of cognitive strain can be more directly attributed to the nature of the feedback intervention rather than initial disparities in participants' baseline cognitive capacities.

More intriguingly, participants also did not report significant differences in cognitive load immediately following the formative feedback assessment. Given the inherent differences between the interactive and non-interactive feedback conditions, this lack of disparity challenges the assumption that interactive feedback necessarily imposes a greater cognitive burden. Despite requiring participants to actively engage with multiple-choice and fill-in-the-blank questions, the interactive feedback did not appear to be perceived as more mentally taxing than the passive review of non-interactive feedback. This finding suggests that interactive feedback may be structured in a way that maintains a manageable cognitive load, thereby mitigating concerns about excessive mental strain during the feedback process.

However, a significant difference emerged at the final measurement point, following the Post Test. Participants in the interactive feedback condition reported experiencing lower cognitive load compared to those in the non-interactive condition. This finding suggests that while the immediate cognitive demands of interactive feedback may not be dramatically different from non-interactive feedback, its long-term benefits become apparent over time. The structured engagement required by interactive feedback may have contributed to improved processing efficiency and a greater overall mastery of the material, ultimately reducing ratings of strain during subsequent assessments. The delayed effect observed here emphasizes the potential of interactive feedback to serve as a cognitive scaffold, initially guiding students through more effortful learning experiences that ultimately translate into more efficient problem-solving and information retrieval.

Further analysis of the relationship between ratings of cognitive load and test performance provides additional support for the crucial role of cognitive burden in learning outcomes. Higher reported cognitive load at the first measurement point was a significant predictor of lower pretest scores. This correlation suggests that students who struggled with the pretest also experienced greater cognitive strain, reinforcing the well-established notion that higher cognitive demands can interfere with effective information processing and retrieval. Similarly, ratings of cognitive load at the second measurement point negatively correlated with both formative feedback test scores and Post Test scores, further demonstrating that excessive cognitive strain may hinder performance even after exposure to instructional feedback. Students who found the feedback stage mentally demanding may have struggled to effectively integrate and apply the information, resulting in diminished performance on subsequent assessments.

A particularly noteworthy finding was the negative correlation between cognitive load ratings at the final measurement point and overall learning gains. Participants who reported higher cognitive strain by the end of the study not only performed worse on the Post Test but also demonstrated lower overall development, as indicated by the difference between their pretest and Post Test scores. This pattern highlights the long-term implications of cognitive load: excessive mental strain is not merely an immediate obstacle but can have lasting consequences on learning trajectories. When students experience heightened cognitive load throughout the learning process, their ability to assimilate and retain new information is compromised, leading to suboptimal academic progress.

Given these findings, it is evident that minimizing cognitive load is crucial for optimizing student performance and learning outcomes. Feedback interventions should be designed with careful consideration of their cognitive demands, ensuring that they support learning without imposing undue mental strain. The interactive feedback condition in this study provides an illustrative example of how feedback can be structured to balance engagement and cognitive load effectively. By requiring active participation while avoiding excessive complexity, interactive feedback facilitated deeper engagement without overwhelming students, ultimately resulting in lower cognitive strain and improved processing efficiency by the final assessment.

These insights have important implications for instructional design and pedagogical strategies. Educators and curriculum developers should be mindful of cognitive load principles when implementing feedback systems, aiming to create interventions that promote meaningful engagement while maintaining manageable mental effort. Future research should further explore the mechanisms through which interactive feedback influences cognitive load over time, investigating factors such as the optimal level of interactivity, the role of individual differences in cognitive capacity, and the potential for adaptive feedback models that tailor cognitive demands to students' proficiency levels. By advancing our understanding of the interplay between cognitive load and feedback mechanisms, we can work toward more effective educational interventions that support student learning and long-term academic success.

7.5 Limitations and Future Considerations

The findings from this study provide valuable insights into the role of feedback interactivity in student engagement and performance. However, several key considerations and limitations must be acknowledged when interpreting these results. One primary limitation of this study is its between-subjects design, which restricted participants to experiencing only one form of feedback—either interactive or non-interactive. Although they were shown a brief example of the alternative feedback condition at the conclusion of the study, this limited exposure does not allow for a fully informed comparison. Without experiencing both feedback conditions in a meaningful way, participants' preferences and perceptions remain constrained by their singular experience. Future research should consider employing a within-subjects design, where participants engage with both feedback types in a controlled manner. This would provide a more comprehensive understanding of how students interact with each feedback mechanism and enable a more direct comparison of engagement, effectiveness, and perception.

A second important limitation pertains to the study's experimental design, which included only two treatment conditions and did not feature a formal control group in the final experiment. While a control group was included in the pilot study, confirming the expected trend of lower scores relative to the treatment groups, the revised procedure in the final experiment may have interacted differently with the presence of a control group. Thus, the absence of a direct control group in the final study prevents a full assessment of the unique contributions of interactive and non-interactive feedback relative to a baseline condition. Additionally, the study examined only two levels of interactivity, whereas a broader range of interactive feedback conditions might yield more nuanced insights. Future studies should explore varying degrees of interactivity to determine whether there exists an optimal level that balances engagement, cognitive load, and learning effectiveness. By incorporating multiple feedback conditions, researchers could pinpoint an approach that maximizes student receptiveness and performance while minimizing unnecessary cognitive strain.

Another noteworthy limitation concerns the potential ceiling effect observed in the formative feedback test and Post Test scores. Despite efforts to increase question difficulty following the pilot study, it remains possible that the assessments were not sensitive enough to detect meaningful growth among higher-achieving students. This suggests that some participants may have already reached a level of proficiency that limited their potential for improvement, thereby obscuring differences between feedback conditions. To address this issue, future iterations of this study should implement more challenging assessments or adaptive testing mechanisms that scale question difficulty based on student performance. By doing so, researchers could better capture the full range of student learning and identify whether feedback interactivity exerts a differential impact on students at various skill levels.

Closely related to this concern is the potential influence of practice effects. Because participants were repeatedly exposed to assessment questions throughout the study, improvements in performance may have resulted from increased familiarity with the test items rather than genuine learning gains facilitated by the feedback. This repeated exposure could diminish the ability to distinguish the true effects of interactive versus non-interactive feedback. Future studies may benefit from alternative assessment methodologies, such as using parallel test forms with different but conceptually equivalent questions to mitigate practice effects. Additionally, incorporating alternative instructional approaches, such as embedded real-world problem-solving tasks, may help differentiate between the effects of feedback engagement and simple repetition-based learning.

Another critical limitation of this study is the reliance on self-reported measures for evaluating cognitive load, learner characteristics, and perceptions of feedback. Self-report measures are inherently subject to biases, including social desirability bias and inaccurate self-assessment. Participants may overestimate or underestimate their cognitive load or engagement, leading to potential distortions in the data. To supplement self-reported cognitive load, future research should explore more objective measurement techniques, such as electroencephalography (EEG) or physiological markers like pupil dilation, heart rate variability, or galvanic skin response. These methods would provide more precise insights into cognitive strain and help validate the subjective reports provided by participants.

Similarly, engagement with feedback was measured using only duration and score metrics, which, while informative, may not fully capture the depth of student interaction with the feedback. For example, a participant who spends an extended period on a feedback page may not necessarily be engaging with the material meaningfully; they may be distracted or disengaged despite the recorded duration. Conversely, a participant who answers questions quickly may have actively engaged but required less time due to prior knowledge. Future studies should incorporate additional engagement metrics, such as eye-tracking technology, to determine whether participants are actively reading and processing the feedback content. This would allow for a more nuanced understanding of how students interact with different feedback formats and highlight the elements that contribute most effectively to learning.

Furthermore, the generalizability of the findings is constrained by the study's sample and content focus. The study was conducted with undergraduate students and centered on mathematical content, which may not fully represent how different student populations engage with interactive feedback. Learning styles, prior knowledge, and subject matter complexity may all influence the effectiveness and reception of feedback. Future research should examine how these feedback mechanisms function across diverse academic levels, including elementary, secondary, and graduate education. Additionally, expanding the study to include subjects beyond mathematics, such as humanities, social sciences, and natural sciences, would provide insight into whether feedback interactivity influences learning differently across disciplines. The cognitive demands of different subjects may necessitate distinct feedback approaches, and understanding these variations could inform more tailored instructional designs.

Overall, this study underscores the multifaceted nature of feedback in educational settings. Effective feedback must not only enhance learning outcomes but also be engaging, well-received, and efficient in its cognitive demands. While interactive feedback presents potential benefits, such as increased engagement and reduced cognitive load ratings over time, its implementation must be carefully designed to avoid unnecessary complexity or student resistance. Future research should continue refining feedback strategies to strike a balance between interactivity and accessibility, ensuring that students can effectively engage with and benefit from the feedback provided. By addressing the limitations and considerations outlined above, future studies can further illuminate the most effective approaches for integrating feedback into educational practice, ultimately leading to more efficient and engaging learning resources for students across many contexts.

References

Ammons, R. B. (1956). Effects of knowledge of performance: A survey and tentative theoretical formulation. *Journal of General Psychology, 54*, 279-299.

Ashford, S. J., & Cummings, L. L. (1983). Feedback as an individual resource: Personal strategies of creating information. *Organizational Behavior and Human Performance, 32*, 370–398.

Attali, Y., & van der Kleij, F. (2017). Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving. *Computers and Education, 110*, 154–169.

Bandura, M. M., & Dweck, C. S. (1985). Self-conceptions and motivation: Conceptions of intelligence, choice of achievement goals, and patterns of cognitions, affect, and behavior. Manuscript submitted for publication.

Bempechat, J., London, P., & Dweck, C. (1991). Children's conceptions of ability in major domains: an interview and experimental study. *Child Study Journal, 21*(1), 11–36.

Brandl, K. K. (1995). Strong and weak students' preferences for error feedback options and responses. *The Modern Language Journal , 79*(2), 194-211.

Butler, R. (1993). Effects of task- and ego-achievement goals on information-seeking during task engagement. *Journal of Personality and Social Psychology, 65*, 18–31.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315-1325.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive science, 5*(2), 121-152.

Cohen, A. D. (1987). Student processing of feedback on their compositions. *Learner Strategies in Language Learning*, 57-89.

Collins, M., Carnine, D., & Gersten, R. (1987). Elaborated corrective feedback and the acquisition of reasoning skills: A study of computer-assisted instruction. *Exceptional Children, 54*,254-262.

Corbalan, G., Kester, L., & Van Merriënboer, J. J. G. (2009). Dynamic task selection: Effects of feedback and learner control on efficiency and motivation. *Learning and Instruction, 19*(455), 4-65.

Corbalan, G., Paas, F., & Cuypers, H. (2010). Computer-based feedback in linear algebra: Effects on transfer performance and motivation. *Computers and Education, 55*, 692-703.

Crisp, V., & Ward, C. (2008). The development of a formative scenario-based computer assisted assessment tool in psychology for teachers: The PePCAA project. *Computers and Education, 50*(4), 1509–1526.

DeLucenay, A., Conn, K., & Corigliano, A. (2017). An evaluation of the impact of immediate compared to delayed feedback on the development of counseling skills in pharmacy students. *Pharmacy Education*, *17*(1), 322.

Dweck, C. S., & Bempechat, J. (1983). Children's theories of intelligence: Consequences for learning. *Learning and Motivation in the Classroom.*

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current directions in psychological science, 12*(3), 83-87.

Epstein, J. I. (1997). The effects of different types of feedback on learning verbal reasoning in computer-based instruction (Doctoral dissertation). Available from ProQuest Dissertations and Theses database.

Finn, B., & Metcalfe, J. (2010). Scaffolding feedback to maximize long-term error correction. *Memory & cognition*, *38*(7), 951-961.

Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction, 39*, 72-87.

Hathila, P., Baria, D., Damor, C., & Mahajan, S. (2023). Students' reflection on immediate feedback during formative assessment. *National Journal of Physiology, Pharmacy and Pharmacology, 13*(3), 1–647.

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & cognition*, *40*, 514-527.

Jaehnig, W., & Miller, M. L. (2007). Feedback types in programmed instruction: A systematic review. *Psychological Record, 57*,219-232.

Kim, J. Y. L., & Phillips, T. L. (1991). The effectiveness of two forms of corrective feedback in diabetes education. *Journal of Computer-Based Instruction, 18*(1), 14-18.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist, 41*(2), 75-86.

Kornell, N., & Metcalfe, J. (2014). The effects of memory retrieval, errors and feedback on learning. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying science of* *learning in education: Infusing psychological science into the curriculum* (pp. 225–251). Society for the Teaching of Psychology.

Kulhavy, R. W., & Stock, W. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*(4), 279–308.

Mandernach, J. B. (2005). Relative effectiveness of computer-based and human feedback for enhancing student learning. *The Journal of Educators Online, 2*(1).

Máñez, I., Vidal-Abarca, E., Kendeou, P., & Martínez, T. (2019). How do students process complex formative feedback in question-answering tasks? A think-aloud study. *Metacognition and Learning, 14*(1), 65–87.

Manson, E., & Ayres, P. (2021). Investigating how errors should be flagged and worked examples structured when providing feedback to novice learners of mathematics. *Educational Psychology, 41*(2), 153-171

Metcalfe, J., Kornell, N., & Finn, B. (2009). Delayed versus immediate feedback in children's and adults' vocabulary learning. *Memory & cognition*, *37*(8), 1077-1087.

Mory, E. H. (2004). Feedback research revisited. Handbook of Research on Educational *Communications and Technology* (pp. 745–783).

Petrović, J., Pale, P., & Jeren, B. (2017). Online formative assessments in a digital signal processing course: Effects of feedback type and content difficulty on students learning achievements. *Education and Information Technologies, 22*(6), 3047-3061.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in cognitive sciences, 15*(1), 20-27.

Rowe, A. D., & Wood, L. N. (2008). Student perceptions and preferences for feedback.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology, 25*(1), 54-67.

Schute, V. J. (2008). Focus on formative feedback. *Review of educational research, 78*(1), 153-189.

Sinha, N., & Glass, A. L. (2015). Delayed, but not immediate, feedback after multiple-choice questions increases performance on a subsequent short-answer, but not multiple-choice, exam: Evidence for the dual-process theory of memory. *The Journal of General Psychology, 142*(2), 118-134.

Slowiak, J. M., & Lakowske, A. M. (2017). The influence of feedback statement sequence and goals on task performance. *Behavior Analysis: Research and Practice, 17*(4), 357–380.

Smits, M. H. S. B., Boon, J., Sluijsmans, D. M. A., & Van Gog, T. (2008). Content and timing of feedback in a web-based learning environment: effects on learning as a function of prior knowledge. *Interactive Learning Environments, 16*(2), 183–193.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science, 12*(2), 257-285.

Van der Kleij, F., Eggen, T. J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers and Education, 58*(1), 263–272.

Van der Kleij, F., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A Meta-Analysis. *Review of Educational Research, 85*(4), 475–511.

VandeWalle, D. (2003). A goal orientation model of feedback-seeking behavior. Human Resource Management Review, 13(4), 581–604.

VandeWalle, D., Ganesan, S., Challagalla, G. N., & Brown, S. P. (2000). An integrated model of feedback seeking behavior: Disposition, context, and cognition. Journal of Applied Psychology, 85, 996–1003.

Williams, J. W. (2012). Formative assessment: Comparing immediate and delayed feedback.

Appendix A

Learner Characteristics Questionnaire

	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
I like learning math topics	○	○	○	○	○
You can always greatly change how intelligent you are	○	○	○	○	○
I want to make sure I get a good score on a math task	○	○	○	○	○
When I take a test I think about how poorly I am doing	○	○	○	○	○
I feel uneasy to do something if I am not sure of succeeding	○	○	○	○	○

	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
Even when I do poorly on a test I try to learn from my mistakes	○	○	○	○	○
I think my math skills are not as good as other students	○	○	○	○	○
Your intelligence is something about you that you can't change very much	○	○	○	○	○
I worry a great deal about tests	○	○	○	○	○
I prefer work that is challenging so I can learn new things	○	○	○	○	○

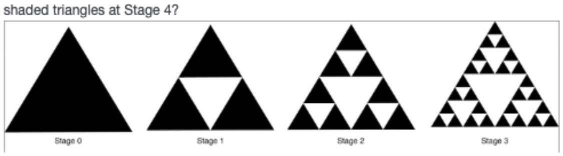
	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
Compared to others, I expect to do well on math tasks	○	○	○	○	○
I like situations in which I can find out how capable I am	○	○	○	○	○
You have a certain amount of intelligence, and you really can't do much to change it	○	○	○	○	○
It is NOT important for me to learn math topics	○	○	○	○	○
Compared to others, I think I know a great deal about math	○	○	○	○	○

	Strongly Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Strongly Agree
I am sure I can do an excellent job on math problems and tasks	○	○	○	○	○
It is NOT important for me to do very well on a math task	○	○	○	○	○
I do not have any anxiety when I take tests	○	○	○	○	○
No matter who you are, you can change your intelligence a lot	○	○	○	○	○
I am so nervous during a test that I cannot remember facts I have learned	○	○	○	○	○

Appendix B

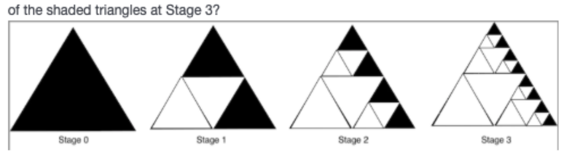
Sample of Pre Test, Formative Feedback Test, and Post Test Questions

The area of this Sierpinski Triangle at Stage 0 is 256 square inches. What is the area of the shaded triangles at Stage 4?



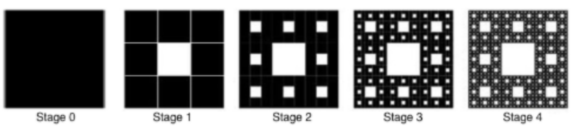
- 27 Square Inches
- 81 Square Inches
- 192 Square Inches
- 256 Square Inches
- 64 Square Inches

The area of this **altered** Sierpinski Triangle at Stage 0 is 48 square inches. What is the area of the shaded triangles at Stage 3?



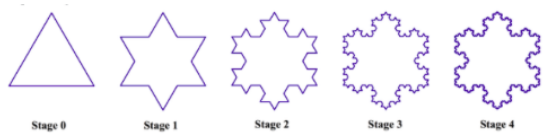
- 12 square inches
- 3 square inches
- 48 square inches
- 24 square inches
- 6 square inches

The area of this shape (Sierpinski Carpet) at Stage 0 is 729 square inches. What is the area of the shaded squares at Stage 1?



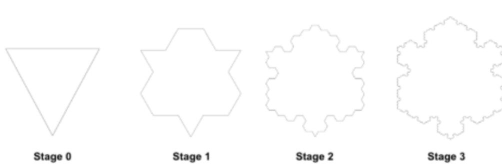
- 486 square inches
- 729 square inches
- 81 square inches
- 648 square inches
- 432 square inches

The perimeter of this Koch's Snowflake at Stage 0 is 54 inches. What is the perimeter of the shape at Stage 3?



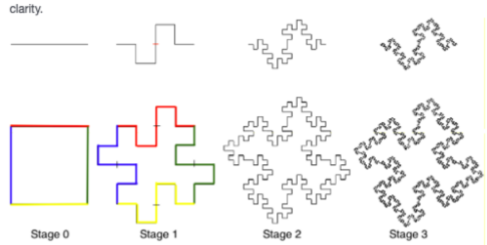
- 128 inches
- 54 inches
- 72 inches
- 96 inches
- 81 inches

The perimeter of this **altered** Koch's Snowflake at Stage 0 is 36 inches. What is the perimeter of the shape at Stage 2?



- 36 inches
- 500 inches
- 2500 inches
- 500/3 inches
- 100 inches

The perimeter of this shape (Minkowski's Island) at Stage 0 is 12 inches. What is the perimeter of the shape at Stage 3? A diagram of one side at each stage is provided for clarity.

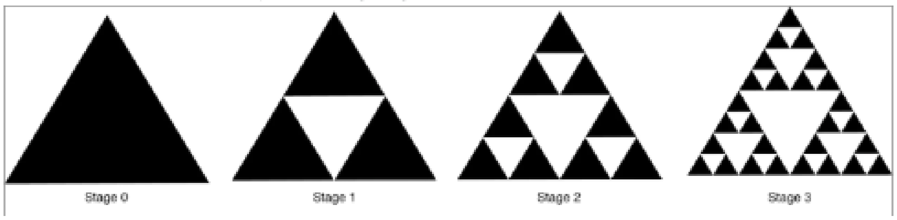


- 12 inches
- 3/2 inches
- 96 inches
- 384 inches
- 48 inches

Appendix C

Sample Acquisition Phase Mini-Lesson

Here we will examine the Sierpinski Triangle again.



As you can see, the shaded area of the initial triangle decreases with each stage as more triangles are removed. It is possible to determine the shaded area remaining at each stage using the following formula:

$$A = A_i \left(\frac{S}{T}\right)^n$$

Where

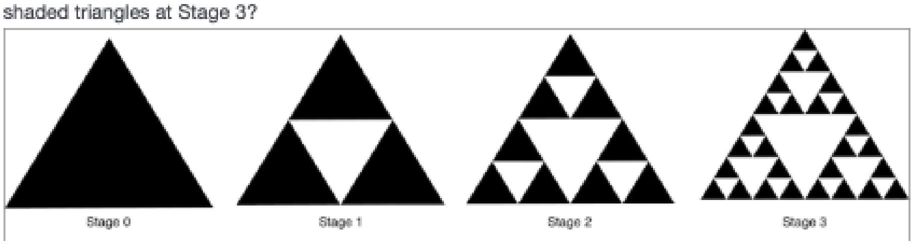
A_i represents the initial area of the shape at Stage 0

S represents the number of shaded triangles at Stage 1

T represents the total number of triangles at Stage 1 (shaded and unshaded)

n represents the stage for which you are determining the area of shaded region

For example, the area of this Sierpinski Triangle at Stage 0 is 64 square inches. What is the area of the shaded triangles at Stage 3?



In this case

A_i = 64 (initial Area at Stage 0)

S = 3 (shaded triangles at Stage 1)

T = 4 (total triangles at Stage 1

n = 3 (for stage 3)

When we plug these values into our formula we get

$$A = 64 \left(\frac{3}{4}\right)^3$$

When we solve this we have

$$A = 64 \left(\frac{3}{4} \times \frac{3}{4} \times \frac{3}{4}\right)$$

$$A = 64 \left(\frac{27}{64}\right)$$

$$A = 27 \text{ in}^2$$

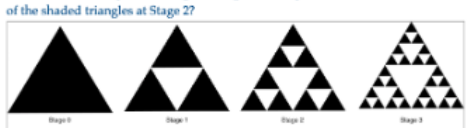
So the area of the shaded triangles at Stage 3 for this example will be **27 square inches**

Appendix D

Example of Interactive Feedback

Which formula do we use for this problem?

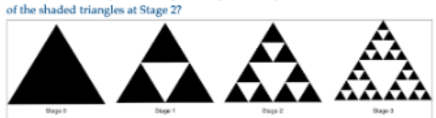
The area of this Sierpinski Triangle at Stage 0 is 32 square inches. What is the area of the shaded triangles at Stage 2?



$A = A_i \left(\frac{S}{T}\right)^n$

$P = \frac{(P_i)(S_1/S_0)^n}{S_0^n}$

The area of this Sierpinski Triangle at Stage 0 is 32 square inches. What is the area of the shaded triangles at Stage 2?



A_i = ___ (initial area of the shape at Stage 0)

S = ___ (# of shaded triangles at Stage 1)

T = ___ (total # of triangles at Stage 1)

n = ___ (stage)

When we plug in the values into the formula we get

$A = 4\left(\frac{3}{32}\right)^2$

$A = 32\left(\frac{3}{4}\right)^2$

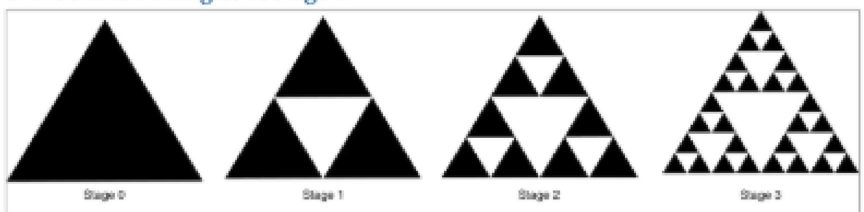
$A = 3\left(\frac{2}{32}\right)^4$

$A = 32\left(\frac{2}{4}\right)^3$

Appendix E

Example of Non-Interactive Feedback

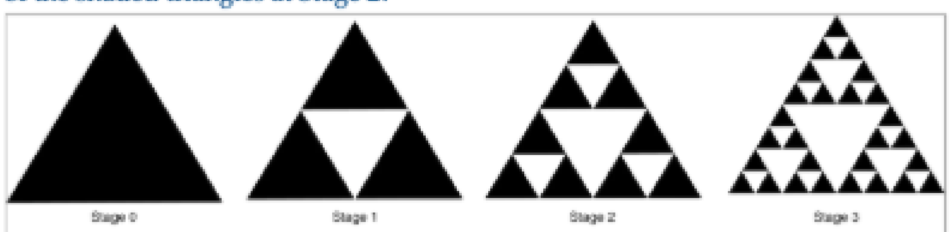
The area of this Sierpinski Triangle at Stage 0 is 32 square inches. What is the area of the shaded triangles at Stage 2?



We will use the area formula for this question

$$A = A_i \left(\frac{S}{T}\right)^n$$

The area of this Sierpinski Triangle at Stage 0 is 32 square inches. What is the area of the shaded triangles at Stage 2?



Ai = 32 (initial area of the triangle at Stage 0)

S = 3 (# of shaded triangles at Stage 1)

T = 4 (total # of triangles at Stage 1)

n = 2 (stage)

When we plug in the values into the formula we get

$$A = 32 \left(\frac{3}{4}\right)^2$$

Appendix F

Sample of Post Test Questionnaire

Read the following question below and then choose the option that best fits how you feel.

	Very Difficult	Somewhat Difficult	Neither Easy nor Difficult	Easy	Very Easy
How difficult was this topic for you?	○	○	○	○	○

Indicate how much you understood fractals before this study and now

	None at all	A little	A moderate amount	A lot	A great deal
Before	○	○	○	○	○
Now	○	○	○	○	○

Read the following question below and then choose the option that best fits how you feel.

	Very Unhelpful	Unhelpful	Neither Helpful nor Unhelpful	Helpful	Very Helpful
How helpful to your learning did you find the explanation guides?	○	○	○	○	○

Read the following question below and then choose the option that best fits how you feel.

	Very Difficult	Somewhat Difficult	Neither Easy nor Difficult	Easy	Very Easy
How difficult was it for you to use the explanation guide?	○	○	○	○	○

Read the following question below and then choose the option that best fits how you feel.

	Very Unsatisfied	Somewhat Unsatisfied	Neither Satisfied nor Unsatisfied	Somewhat Satisfied	Very Satisfied
How satisfied are you with how you think you did on the post test?	○	○	○	○	○
How satisfied are you with how much you learned?	○	○	○	○	○