THE ROLE OF DIGITAL NUDGES IN ENGINEERING STUDENTS' ENGAGEMENT WITH AN EDUCATIONAL MOBILE APPLICATION

by

Ahmed Ashraf Butt

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Engineering Education

West Lafayette, Indiana

August 2023

THE PURDUE UNIVERSITY GRADUATE SCHOOL

STATEMENT OF COMMITTEE APPROVAL

Dr. Muhsin Menekse, Chair

School of Engineering Education

Dr. Mathew Ohland

School of Engineering Education

Dr. Sean Brophy

School of Engineering Education

Dr. Saira Anwar

Department of Multidisciplinary Engineering
Texas A&M University

Approved by:

Dr. Senay Purzer

I dedicate this dissertation to my family, teachers, and friends for their guidance, love, and support.

ACKNOWLEDGMENTS

I am grateful to God, the most benevolent, and the most merciful. Without God's blessings, it would not have been possible for me to embark on and successfully complete my doctoral journey. It is with God's grace that I have always found myself surrounded by a strong support network comprising mentors, friends, and family members. This support network has played an important role in all my accomplishments and in making me the person that I am today.

I want to express my gratitude to my mother. Her love, dedication, and commitment to raising her children has had a profound influence on my personal and academic growth. My father who left us long ago always wished that we get a college education. My mother fulfilled his wish and enabled us to attend college. I now stand proud as the first generation to get a college degree. All thanks to my parents, their constant love, and endless sacrifices. I am also extremely grateful to my wife for her unwavering support and love that kept me going through the most difficult of times. My love, I will forever be grateful for the sacrifices you made and the belief you had in me. Thank you for being my rock and helping me achieve this milestone. I am also thankful for all my siblings for always being there for me whenever I needed advice and for staying patient with me during this journey.

I would also like to express my gratitude to my committee members, Drs. Muhsin Menekse, Saira Anwar, Mathew Ohland, and Sean Brophy, for their invaluable support, expertise, and guidance throughout my doctoral research.

As my advisor and head of my dissertation committee, Dr. Menekse has played an important role in my professional and personal growth. I have learned a lot from him in the areas of mentorship, teaching, and conducting research. The one thing that stands out about him is his empathetic nature. Whenever I struggled on this journey, his kind and gentle support was always there to guide me. The endless discussion sessions with him about my research proved to be vital not only for the conceptualization of the research idea that guided my dissertation but also transformed my world view about engineering education, academia, and life as a researcher. Dr. Saira Anwar has been my mentor for more than a decade now. She has always been my go-to person whenever I needed advice, personally or professionally. In my dissertation, apart from being patient with several questions asked, her expertise in quantitative research has helped me conduct and analyze the results of my experiments in the study. I am appreciative of Dr. Ohland, who has provided me with detailed feedback on my dissertation. I am very grateful to Dr. Brophy for his expertise in educational application that really helped me in designing the nudge interventions in my study. Additionally, he inspired me with his energy and thoughtfulness in different aspects of academic life and the evaluation of research.

A very special thanks to Dr. Ruth Streveler for her mentorship and guidance during my academic journey. Dr. Saira Anwar, Dr. Zahra Atiq, and Loretta McKinniss for believing in me, and without them, I would have left the ENE in my first semester. Their constant encouragement helped me to achieve this goal.

I am also grateful to all my lab members who worked on CourseMirror both at Purdue within Dr. Menekse’s lab and at University of Pittsburg at Dr. Littman’s lab. The close collaboration and constant feedback on research from my lab mates improved my research and in doing so strongly shaped my identity as a researcher. I am also grateful to all my ENE community at Purdue whose presence and commitment to excellence helped me in one way or another in this journey.

A very special thanks to my friend's cohort, Talha Zubair Janjua, Waleed Sheikh, Syed Ali Kamal, Muhammad Junaid, and Usman Bashir, for their constant advice, guidance, and inspiration towards achieving this goal.

I also want to thank my teachers from school, and college, specially, Mr. Khalid, Mr. Azam, Mr. Amjad, Mr. Numan, Ms. Tayyaba, Dr. Saira, and Dr. Qurat-ul-Aien, for laying the foundation that helped me achieve this goal.

I acknowledge the time and effort the instructors put into implementing the CourseMIRROR application in their classrooms. Also, this study could not have been possible without the participating students. Thanks a lot.

Lastly, I acknowledge the support of CourseMIRROR (www.CourseMIRROR.com) research teams that has helped me in the smooth implementation of these experiment. Also, a big thanks to IES for providing the resources that made this research possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The proliferation of digital educational applications (apps) has revolutionized the pedagogical landscape for students and instructors, both within and beyond the confines of traditional classrooms. Educational apps offer a variety of features that can help students learn more effectively, including personalized instruction and real-time feedback. However, some studies have found that students may not be engaging with the apps regularly or for extended periods of time. This lack of engagement can limit the apps' potential to improve student learning. Consequently, researchers have investigated methods to enhance students' app engagement, including the use of digital nudges. Digital nudging is a strategy that proposes utilizing small, non-intrusive cues that capitalize on individuals' cognitive biases to influence their behavior.

This dissertation makes a significant contribution to ongoing efforts by examining the effectiveness of nudge-based digital interventions in improving students' engagement with the CourseMIRROR educational app. CourseMIRROR is an educational mobile app that prompts students to reflect on the interesting and confusing aspects of lectures throughout a semester. The CourseMIRROR app uses Natural Language Processing (NLP) algorithms to 1) scaffold the students while generating reflections and 2) summarize the students' submitted reflections. This study focuses on designing digital nudges to improve students' cognitive and behavioral engagement with specific features of the app that are crucial to achieving its primary purposes. These primary purposes include 1) facilitating students to submit reflections, 2) enabling students to view the reflection summary interface, and 3) scaffolding students to write in-depth and comprehensive reflections. The study consists of three experiments investigating the effectiveness of these digital nudges for improving student engagement with the CourseMIRROR app.

For this dissertation, I conducted three experiments by implementing the CourseMIRROR app in multiple sections of a first-year engineering course at Purdue University over a semester. ***Experiment 1*** investigated the impact of social comparison nudge and neutral reminder nudge to increase students' reflection submissions by using the app. Students were randomly assigned to one of three conditions: social comparison nudge, neutral reminder nudge, or baseline (no nudge). The social comparison nudge involved reminding and showing peers' behavior through their reflection submissions, and the neutral reminder nudge involved sending automated reminders to students to submit their reflections. The results indicated that social comparison and neutral reminder nudges were effective in increasing reflection submissions compared to the baseline condition. However, the social comparison nudge was slightly more effective in improving the number of reflection submissions than the neutral reminder nudge. Also, the nudge interventions became effective in increasing the reflection submissions by refocusing the students' attention as time progressed in the semester.

Experiment 2 explored the impact of summary reminder nudges and interface nudges to increase students' visits to the reflection summary interface in the app. Students were randomly assigned to summary reminder nudge, interface nudge, or baseline conditions. The summary reminder nudge involved reminding students to visit the reflection summary interface in the app. The interface nudge involved making the summary available lecture more prominent to draw students' attention to the reflection summary interface. The result revealed that summary reminder and interface nudges did not significantly improve the number of students' visits to the reflection summary interface. Also, for all conditions, students' visits to reflection summary interface decreased over time as time progressed.

Experiment 3 examined the impact of scaffolding and throttling mindless nudges on promoting more comprehensive and lengthier reflection submissions. Students were randomly assigned to one of three conditions: scaffolding nudge, throttling mindless nudge, or baseline. The scaffolding nudge involved providing students with real-time feedback to guide their reflection writing, while the throttling mindless nudge involved giving a pause to re-think if they want to move forward to the next question or revise their reflection in the application. Overall, the results showed that scaffolding and throttling mindless nudges effectively promoted more comprehensive and lengthier reflection submissions over the semester and within each time. However, students' reflections in all conditions remained either consistent or decreased in reflection text length and specificity score over time in a semester.

The study's results indicate that digital nudges can effectively enhance students' engagement with educational applications, especially in reflection activities using CourseMIRROR. These findings provide valuable insights into designing and implementing digital nudges in educational apps and evaluating their impact on student engagement. Future research should build on these results to develop a more comprehensive understanding of the potential of digital nudges to support student engagement in educational technology settings.

INTRODUCTION

Over the last two decades, the advancement in the digital technology has revolutionized the teaching and learning experiences (Bilyalova et al., 2020). With this digital transformation, students and instructors can access digital tools, such as mobile devices and applications that can improve their teaching and learning experiences (Menon, 2022). In this regard, a notable digital advancement is an educational application (app) defined as software designed for instructional purposes that can be used on different devices, including smartphones, and tablets (Geissinger, 1997). Prior studies have discussed that such apps help reduce the students' cognitive load by providing access to interactive and easy-to-learn content beyond time and space constraints (Camilleri & Camilleri, 2017; Zydney & Warner, 2016). Consequently, educational apps have become an essential teaching and learning tool for both instructors and students.

Education literature has discussed a range of benefits that educational apps provide to both students and instructors in enhancing the learning experience (Zhang & Liao, 2015). For students, educational apps provide easy access to high-quality learning materials and resources, irrespective of their location, or socioeconomic status (Falloon, 2013). A key benefit of educational apps is their ability to increase student engagement with learning activities (Wu et al., 2013). This has been confirmed by several studies that incorporated game-like elements (e.g., Bartel & Hagel, 2014), providing real-time feedback (e.g., Aljohani & Davis, 2013), and improving collaboration (e.g., Bouta & Retalis, 2013) to make learning interactive and fun. Similarly, educational apps have helped instructors to inform their pedagogical practices by monitoring student progress and identifying areas where additional support may be needed using different data analytic techniques (Fan et al., 2015). Additionally, educational apps can save teachers' time by automating certain tasks, such as grading and assessment, freeing up more time for lesson planning and instruction (Lim & Yunus, 2021; Zhao, 2019).

The widespread acceptance of educational apps as a valuable learning tool is also evident from the increase in application downloads globally. The educational app downloads have increased from 522 million in 2017 to 936 million in 2020 at both Google Play store and Apple App store (Statista, 2020). Seeing the increasing importance of the educational app, the literature has appeared focusing on the design and development of educational apps (Falloon, 2013; Papadakis et al., 2018). However, a limited literature is exploring the students' engagement with the educational app and have raised concerns related to the educational app's ineffectiveness to engage students (Melcher et al., 2022; Pechenkina et al., 2017; Pham & Chen, 2018).

In the literature, app engagement is defined as the user's meaningful and improved interactions with technology (Doherty & Doherty, 2018). This engagement can have three dimensions: behavioral, cognitive, and emotional. Behavioral engagement pertains to physical participation or willingness to interact with technology (Bouta & Retalis, 2013; Islas Sedano et al., 2013), cognitive engagement involves putting in the effort to learn or master a particular task through the use of technology (Greene, 2015), and emotional engagement encompasses the emotional response that arises during the use of technology (Doherty & Doherty, 2018). In this study, I adopted the same conceptualization of a multidimensional perspective of app engagement, with the distinction that the users were college students in my case. Furthermore, I only explored the behavioral and cognitive dimensions of students' engagement with the application.

In the context of app engagement, prior studies have argued that students are unable to fully utilize the educational apps as they get disengaged and eventually abandon the app (Melcher et al., 2022). This is evident by the educational apps' retention rate (i.e., continuous usage) of 27%

after 90 days of use, which is lower than the retention rate for other applications at 35% (Sefferman, 2021). Moreover, the annual retention rate for educational apps is only 4% (Ben-Joseph, 2021). This low retention rate is often attributed to the inability of these apps to engage their users with the app, including instructors and learners (Pachler et al., 2009). The lack of engagement can result in a decreased motivation to use the app, which ultimately reduces its effectiveness in promoting learning. Thus, it could improve the educational app's retention rate by improving the students' engagement.

Limited studies have explored the reasons related to the educational apps' inability to engage students with them regularly. One reason is that it is challenging for learners to prioritize and remain engaged with any educational app, especially if they have many other apps and digital tools competing for their attention (Loveless, n.d.). Additionally, integrating an educational app into their daily routine can be difficult for learners, especially if it does not fit seamlessly into their existing habits. These challenges can lead to low levels of engagement and utilization of educational apps.

The survey of educational literature indicates that studies have typically focused on the content and effectiveness of various educational apps on the students' learning experience. However, limited attention is given to the understand the students' interaction with the app. In other words, there is awareness about the effective content or structure of the app, but understanding of how students interact or engaged with the educational apps is limited (Anwar et al., 2022). Therefore, there is a need to understand the students' engagement with the educational app to improve their experience for effective learning outcomes. In this regard, limited studies have discussed the strategies used by the educational app designers to improve the students' app engagement. These strategies include: 1) regularly updating the application with fresh content

(Chiong & Shuler, 2010), 2) ensuring that the application has a well-designed user interface (UI) and user experience (UX; Fard, n.d.), 3) using push notifications (Pham et al., 2016), and 4) creating interactive content that promotes engagement and active learning (Oh et al., 2015). Although the implementation of these strategies has shown an increase in app engagement, the literature argued the need to explore other ideas for improving the learners' app engagement.

Among the ongoing efforts to improve learners' app engagement, the idea of nudging has shown some promising results as a cost-effective approach to improve app engagement (Eslambolchilar et al., 2011; Fritz, 2017). "Nudging," as introduced by Thaler & Sunstein (2008), involves using subtle, indirect, and low-cost interventions to influence people's decision-making and behavior in a positive way while still preserving their freedom of choice. When this approach is applied in the digital context, it is referred to as "digital nudge," and the process of implementing such interventions is known as "digital nudging."

In the education literature, limited studies have explored the impact of nudging interventions to improve the students app engagement (Damgaard & Nielsen, 2018; Harley et al., 2007; Sherr et al., 2019). These studies have mostly employed reminder nudges (bringing the students' attention to a particular task or decision) to keep the students engaged with the educational app (e.g., Simmons et al., 2018). Therefore, this study has explored the literature on human-computer interaction (HCI) to explore approaches used to design nudge interventions in this study. Furthermore, prior studies have mostly relied on app analytics (e.g., session length, total time spent) to explore the impact of nudges on the student's app engagement (Fancsali et al., 2021; Pham et al., 2016; Pham & Chen, 2019). This study also relied on the app analytics (e.g., number of reflection submission) to inform the research question. Furthermore, the aim of the study is to enhance the literature by designing different nudging interventions and studying their impact on the students' app engagement within the classroom settings.

In this study, I used the idea of nudging to design digital nudge interventions and explored their impact on students' engagement in an educational app, i.e., CourseMIRROR. CourseMIRROR is a mobile educational app that prompts students to reflect on their learning experiences (Fan et al., 2015; Menekse et al., 2018). After each class, students are asked to reflect on the confusing or interesting aspect of the lecture. Furthermore, a set of NLP algorithms is used to summarize students' reflections submitted for each lecture, and scaffold students during reflection writing in the application. It also has an associated instructor website that provides reflection summaries and individual reflections through different data analytics (e.g., reflection submission rate). However, this study is focused only on the mobile application of the CourseMIRROR system.

Previous research has shown that CourseMIRROR has a positive impact on various aspects of students' learning (e.g., Menekse, 2020; Menekse et al., 2018). However, studies have also indicated a lack of students' app engagement with the CourseMIRROR app. For instance, Fan et al. (2015) conducted a pilot study by implementing the CourseMIRROR application in a STEM classroom. The study found the students after using the app for some time submitting shallow or irrelevant reflections to the prompt or lecture content such as "N/A" or "all good." Additionally, the students' reflection length decreased over time. These results suggested that CourseMIRROR was unable to engage students throughout the semester. Moreover, prior studies have argued that students needs to remain highly engaged with the educational app to realized its' full potential (Hirsh-Pasek et al., 2015; Kim & Baek, 2018; Pham & Chen, 2019). In this regsrd, no previous studies have designed strategies to improve the students' app engagement with the

CourseMIRROR mobile application. Therefore, this study serves as the initial effort to investigate the impact of nudging interventions on the student's engagement with the CourseMIRROR app.

Specifically, I conducted three different experiments to understand the impact of nudging interventions on students' engagement with features of the CourseMIRROR apps, essential to achieve its primary goals. The primary goal of the application is to 1) encourage students to submit more reflections, 2) enable the students to visit the reflection summary interface to read and learn from their classmates' reflection summary, and 3) scaffold students to write comprehensive and detailed written reflections (Fan et al., 2015, 2017; Luo et al., 2015). This study designed five nudge interventions to improve the students' engagement with different features of the CourseMIRROR app. The nudge interventions include reminder nudge (directs attention), social comparison nudge (uses social norms to refocus students' attention), interface nudge (highlights choices through an interface design), scaffolding nudge (offers in-time feedback), and throttling mindless nudge (prompts informed decisions by introducing a pause). Furthermore, this study used educational app analytics as an engagement measure to understand the engagement differences among students in different conditions (control vs. treatment).

Research questions

The following research question guided this study:

What is the effectiveness of nudging interventions on students' engagement with the CourseMIRROR application?

Specifically, I have following research questions guiding each experiment.

Experiment 1: Facilitating students' reflection submissions.

1. Do students receiving neutral reminder nudges submit more reflections compared to the students receiving no nudge?
2. Do students receiving social comparison nudges submit more reflections compared to the students receiving no nudge?
3. What is the relative effectiveness of both nudge interventions on the students' reflection submissions?
4. How do the students' reflection submissions change over time in each condition (i.e., neutral reminder nudge, social comparison nudge, and baseline)?

Experiment 2: Supporting students' reflection summary views.

1. Do students receiving summary reminder nudges visit the reflection summary interface more often than those who do not receive nudges?
2. Do students receiving interface nudge visits the reflection summary interface more often than those who do not receive nudges?
3. What is the relative effectiveness of both nudge interventions on the students' number of visits to the reflection summary interface?
4. How do the students' reflection summary interface views change over time in each condition (i.e., summary reminder nudge, interface nudge, and baseline)?

Experiment 3: Scaffolding students to generate specific reflections.

1. Do students receiving scaffolding nudges show improvement in the specificity and length of their reflections compared to those receiving no nudge?

2. Do students receiving throttling mindless nudges show improvement in the specificity and length of their reflections compared to those receiving no nudge?
3. What is the relative effectiveness of both nudge interventions on the specificity and length of students' reflections?
4. How do the students' reflection specificity and text length change over time points in each condition (i.e., scaffolding nudges, throttling mindless nudges, and baseline)?

Significance of the study

This study contributes to the literature in several ways. First, the study aims to contribute to the limited nudge literature in the STEM domain by designing nudge interventions in real classroom settings. Second, this study employed an experimental research design and interdisciplinary approach to explore the impact of the nudges on the students' app engagement. Even though studies have used the idea of nudging, limited studies have provided a theoretical framework to ground their studies in previous research (e.g., Sherr et al., 2019). In this study, I have conceptualized a theoretical framework to guide different stages of my study, such as the selection of engagement measures, design of nudge intervention, and interpretation of the results.

Furthermore, prior studies mostly implemented single nudge intervention to achieve desired goals (e.g., Harley et al., 2007). However, this study explores the relative efficacy of the different nudge interventions toward a particular student's behavior. Moreover, this study provided insights into the design and guidelines for the discussed nudging interventions in similar contexts for future research.

THEORETICAL FRAMEWORK

I am using the multidimensional engagement theory and the nudge theory as the basis of the theoretical framework in this study. This theoretical framework has played a pivotal role in informing all aspects of this research, including the design of interventions, data collection, analysis, and interpretation of results. The following section provides a detailed overview of each theory and its contribution to the study's theoretical framework.

Multidimensional perspectives of engagement

The concept of user engagement with digital technology is multidisciplinary in nature (O'Brien, 2016), with conceptualizations based on context, technology, and dimensionality (Doherty and Doherty, 2018). One of the earliest conceptualizations was presented by Chapman (1997), who defined engagement as "something that 'engages' us is something that draws us in, that attracts and holds our attention" (*p*. 3). Engagement is also considered a multi-stage process where the user goes through various phases (O'Brien & Toms, 2008). These phases include adoption (where users first adopt the application), engagement (where users begin interacting with the application), disengagement (where users stop interacting with the application), and re-engagement (where users re-engage with the application). Users can be seen as highly engaged with the technology once they successfully complete these steps.

Owing to the complexity and diverse conceptualization of the topic, various disciplines (e.g., Education, Human-Computer Interaction, and Behavioral Sciences) have often adopted a multidimensional perspective of engagement as a theoretical foundation for investigating user engagement with digital technology (Doherty & Doherty, 2018; O'Brien, 2016). This multidimensional perspective of engagement is defined as the user's (or student's) meaningful and improved interaction with digital technology across three dimensions: behavioral, cognitive, and emotional (Holdener et al., 2020; Kim et al., 2017; O'Brien, 2016; O'Brien & Toms, 2008). The ***Behavioral*** dimension of engagement emphasizes users' actions, willingness to participate, and continuous interaction with technology, making it objectively measurable (Bouta & Retalis, 2013; Islas Sedano et al., 2013). Previous studies have used a variety of measures to assess behavioral engagement in educational apps. The commonly used quantitative measures of engagement are the app usage (e.g., Pham & Chen, 2018), number of completed tasks (e.g., Anwar et al., 2022), and number of accesses to the app feature (e.g., Jayasekaran et al., 2022). These measures guided the studies by understanding the user's behavior while interacting with the app.

The cognitive dimension of engagement refers to the conscious effort of the user to understand or master a learning task associated with technology. This conscious interaction includes the user's attention (Pham et al., 2016), effort, or awareness (Islas Sedano et al., 2013) while interacting with the technology. Cognitive engagement with educational apps has been assessed through various measures, including self-reported measures (i.e., personal opinions, perceptions, and experiences) and objective measures (i.e., observable and measurable phenomena). One commonly used self-reported measure is questionnaires asking learners to rate their engagement with an educational app. For example, Song et al. (2022) used a questionnaire to measure cognitive engagement to enhance primary students' vocabulary learning engagement, which included items such as "I was totally absorbed in what I was doing."

On the other hand, objective measures of cognitive engagement are typically educational apps data analytics that revelated the learners' interaction with the learning content or their performance with the learning tasks, such as quiz attempts (e.g., Kizilcec & Chen, 2020) or performance scores (e.g., Pham & Chen, 2018). Additionally, some studies have started using the physiological measures (e.g., electroencephalography (EEG) and eye tracking) to explore cognitive engagement while learners interact with educational apps (e.g., Apicella et al., 2022). For instance, Halderman et al. (2021) used EEG measures to understand brain activity to examine students' cognitive engagement while they attempted an online simulated GRE.

Lastly, the emotional dimension of engagement refers to the emotional reaction developed by the users while interacting with the technology (Doherty & Doherty, 2018). Furthermore, emotions are also interrelated with user behavior and cognitive development (Ruth et al. 2002). In other words, emotional engagement can also be achieved by improving the other two aspects of the engagement. Prior studies into the emotional aspect of app engagement mostly relied on the self-report measure, which requires participants to rate their emotional experiences while using the app. For example, Ding & Chai (2015) used a survey to assess emotional engagement with the usage of the mobile learning app. Their results revealed that emotions have an impact on the continuous usage of mobile applications. Some studies have started to use advanced objective measures to understand emotional engagement. These measures include physiological measures such as heart rate variability, electrodermal activity, and facial expressions (de Vreede et al., 2019). These measures provide an objective assessment of emotional engagement that is less susceptible to self-reported biases.

In this study, I primarily focused on the two dimensions of engagement, i.e., cognitive and behavioral, to fully understand the role of student engagement with the CourseMIRROR mobile app. These dimensions have guided my conceptualization of student engagement, engagement measures selection, and data interpretation in this study. Through the investigation of the students' interaction with the application, I was able to identify the measures of their app engagement that are relevant to my study. The student's cognitive engagement in the CourseMIRROR app is indicated by their deep and critical thinking while writing reflections on their learning experiences. To measure their cognitive engagement, I used the reflection specificity score as a cognitive engagement measure which is essentially a numeric value showing the relevance of their reflection with the reflection prompts and lecture content. For the behavioral engagement, I relied on the students' behavior analytics, i.e., count of reflection submissions, reflection text length, and summarization views. Furthermore, I analyze these engagement measures to understand the extent to which students engage with the application. This multidimensional perspective of students' app engagement has allowed me to explore the impact of introduced interventions on the students' interaction with the CourseMIRROR app.

A growing body of educational technology literature suggests that nudging can be used as an effective and cost-effective way to improve students' app engagement in the digital environment (Brown et al., 2019; Fancsali et al., 2021; Pham et al., 2016). In this study, I have used the idea of nudging to improve the students' app engagement using the CourssMIRROR educational application. The use of nudges in educational applications seems like a natural choice to improve students' app engagement, as Weinmann et al. (2016) have observed that user interface designers frequently employ nudging interventions, whether intentionally or not, to influence user behavior during app interactions.

Nudge theory

The other major component of the theoretical framework of this study is based on the Nudge theory, which suggests that despite being aware of their best interests, people sometimes behave in irrational ways and fail to make the best choices (Thaler & Sunstein, 2008). In addition, people's actions can be predictable. Therefore, behavioral interventions can be designed to help encourage individuals to make better decisions (Ariely, 2009). The Nudge theory proposes that positive reinforcement or indirect suggestions (i.e., small behavioral interventions) can influence people’s behavior and decision-making to achieve a desirable outcome (Weijers et al., 2020). Moreover, different mindsets (i.e., set of beliefs) have been investigated by Dweck (2016), such as fixed mindset and growth mindset. It is believed that people with a fixed mindset give up easily, whereas those with a growth mindset persist and try to expand their learning. Dweck (2016) further believed that small behavioral interventions such as nudges could effectively develop a growth mindset.

Thaler and Sunstein (2008) defined the term “nudge” as “any aspect of the choice architecture that alters people’s behavior predictably without forbidding any options or significantly changing their economic incentives” (p. 6). Here, choice architecture refers to the environment in which an individual is exposed to various choices. The choice architecture is designed to influence peoples’ behavior, where people have choices, but their options are not limited, which allows for autonomous decision-making. Any intervention that alters human behavior can be considered a nudge if it is easy, simple, non-commanding, cost-effective, and supports an autonomous choice of options. For instance, introducing a norm nudge such as a descriptive social norm (informing people about your norm of paying taxes) in tax letters has increased the repayment rate by 15 percentage points (Office Behavioural Insights Team, 2012). Similarly, the nudge through chocolate placement beside champagne instead of meat or any other place would positively impact the sale of chocolates (Thaler and Sunstein 2008).

Nudge theory is broadly based on two theories: 1) Prospect theory and 2) dual-process framework. According to the prospect theory, humans make irrational decisions not because of mental overload, lack of calculating capacity, or limited information (Neuhaus, 2020). Instead, human beings naturally use heuristics or shortcuts to ease or accelerate their decision-making process. Moreover, Thaler and Sunstein (2008) linked prospect theory with the dual process theory. The dual process theory states that human beings process information based on two systems, namely System 1 and System 2. In System 1, they deal with automatic, instinctive, and uncontrolled thinking. To facilitate this thinking, System 1 takes quick actions by using heuristics and shortcuts. For example, spontaneous responses, like snacking on the food placed in front of us or getting startled by sudden movements or loud noises. On the other hand, System 2 is concerned with "reflective," meaning controlled, deliberate, thought-out, slow, rational, and self-aware thinking. Therefore, this system takes more information and in-depth analysis into the decision-making process. Some examples of behavior characterized by System 2 include parking a car in a narrow space, comparing two laptops for the best value, or filling out a tax form.

Both the prospect theory and dual process theory discuss nudging as a way to alter people's behavior using heuristics and shortcuts to make rational choices. Furthermore, dual process theory suggests that human behavior is often guided by the system 1, as little effort is involved in making decisions (Weijers et al., 2020). This reliance on System 1 can cause behavioral inconsistencies with a person's goals. For instance, if a person's goal is to lose weight, they may still indulge in snacking due to overreliance on System 1. The suggested absence of rationality in System 1 generates some insignificant environmental cues that can strongly influence behavior, whereas in the case of System 2 these cues are insignificant.

The nudge theory assumes that instead of resisting or countering the lack of rationality of System 1, it should be accepted and brought to good use. Thaler and Sunstein (2008) believe that people's behavior should be predictably altered through nudges in the environment instead of restricting their options or changing the incentives. The nudges use the lack of rationality of System 1 and can help people make better decisions. Based on this argument, research has shown that well-designed nudges can influence students' behavior to improve their interaction with the app and significantly impact their app engagement (e.g., Castleman et al., 2014; Pham et al., 2016).

The rationale behind using nudge theory as a construct in my theoretical framework is that it helps us in understanding that how nudge interventions can be designed and used to leverage students' cognitive biases and their tendency to use heuristics to influence their behavior and encourage their engagement with the app.

In this study, I designed the nudge interventions within the CourseMIRROR application and explored their impact on the students' app engagement using the theoretical lens of Nudge theory and multi-dimensional perspective of engagement. Nudge theory, on the one hand, informs this study by explaining that nudging interventions can work as positive reinforcement or indirect cues for influencing students' decision-making. On the other hand, the multi-dimensional perspective of student engagement informs us about students' meaningful and continuous interaction with the CourseMIRROR application. The collective understanding offered by these two theoretical lenses in this theoretical framework has enabled a profound understanding of app engagement and its dimensions in designing digital nudges to target behavioral and cognitive dimensions of engagement.

LITERATURE REVIEW

Digital nudging

The concept of nudging has been extended to the digital space, which is referred to as digital nudges and the process of introducing them as digital nudging (Barev, 2020). Digital nudging is defined as “a subtle form of using design, information, and interactive elements to guide user behavior in digital environments, without restricting the individual’s freedom of choice” (Meske & Potthoff, 2017, *p*. 2589). The term digital environment refers to the collection of user interfaces and interactive tools that enable individuals to navigate and make informed decisions within the digital realm. These interfaces include simple menu-driven interfaces, recommendation engines, search algorithms, and content management systems. While there has been extensive research on the impact of nudging in physical environments (e.g., framing of options in a form to opt for a service; (Pichert & Katsikopoulos, 2008), there is limited but growing interest in the potential of nudging to influence human decision-making in the digital realm (Teuber et al., 2022).

The concept of nudging has a long history in technology literature, with user interface designers acting knowingly or unknowingly as choice architects influencing people’s decisions through their designs (Weinmann et al., 2016). Although it has not always been referred to as a “digital nudge”, the application of nudging intervention has a natural extension in the digital realm. For instance, Carr (2013) discussed the mobile payment system “Square,” which utilizes digital nudges to increase the tip amount for clients. The nudge strategy in this system involves setting a default behavior for tipping, which requires clients to actively opt out if they do not wish to leave a tip. By using this simple default nudge strategy, the payment system can increase the tipping amount. Moreover, any intervention that satisfies the general principles of nudging (simple, cost-effective, non-commanding, and suggestive) and is presented in the digital environment can be considered a digital nudge.

Digital nudging is a relatively new idea, but a considerable amount of literature has started discussing the topic in recent years. Various studies have explored the use of digital nudges and their effectiveness in different contexts. For instance, the Fitbit activity monitor was used with a combination of nudges (e.g., reminder nudge) to help people increase their physical activity (Mele et al., 2021). Additionally, warnings have been employed as digital nudges to discourage unnecessary online purchases (Esposito et al., 2017). Furthermore, several studies have been conducted on the design of effective digital nudges, which propose different models for designing digital nudges. For instance, Meske & Potthoff (2017) suggested a 'digital nudging process model' that outlines a design pattern for digital nudges. This model consisted of three stages: analyzing the desired behavior and goals of a digital nudge, designing the appropriate nudge aligned with the goal, and evaluating whether the nudge achieved its goal.

Similarly, Schneider et al. (2018) discussed a cyclical model for designing and assessing the effectiveness of the digital nudge. This model includes steps such as defining a goal, understanding the users, designing the nudge, and evaluating the effectiveness of the nudge. Broadly, all previous nudge design models aimed to create an effective nudge aligned with targeted people's behavior (i.e., identified goals) and then evaluate whether the nudge achieves the desired behavior. This study followed the same design pattern where nudging interventions were designed to improve the students' app engagement and then evaluate the effectiveness of these nudge interventions in achieving the desired behavior. Additionally, researchers have also explored digital nudges and their application in various studies, including systematic literature reviews (Bergram et al., 2022; Mirsch et al., 2017), policy papers (e.g., Einfeld, 2019; John et al., 2009), work-in-progress papers (e.g., Hummel et al., 2017; Wilkinson et al., 2017), or even full explanatory studies (e.g., Brown et al., 2019; Schneider & Graham, 2017).

Ethical considerations for designing a digital nudge

Researchers have emphasized considering the ethical implications while designing the nudge (Lembcke et al., 2019; Paunov et al., 2019). As digital nudge is supposed to influence people's behavior directly, they must be designed and implemented ethically. While introducing the idea of nudging, Thaler & Sunstein (2008) discussed several ethical principles for nudging, with the most important being the principle of non-paternalism. The principle emphasized that nudges should not be designed to promote the preferences of choice architects but rather to serve the interests of the people being nudged. Therefore, designing digital nudges with ethical considerations is crucial to ensure that the nudges do not infringe the people's rights or manipulate them in unethical ways (Lembcke et al., 2019; Schmidt & Engelen, 2020).

Prior studies have also argued that nudges can have unintended consequences resulting in a backlash if ethics are not considered. For instance, the Netherlands passed a bill in 2016 to increase organ donation rates using the default nudge, presuming citizens as donors unless they opted out. However, the bill backfired as the number of citizens refusing to donate broke records. The Dutch rebelled against it, feeling their autonomy violated (*Disappointing Donor Week*, n.d.). The incident highlights the importance of ethical considerations when designing nudges, even with the intention of the greater good, as not considering ethics could lead to unintended consequences.

There has been much debate surrounding the ethical implications of designing digital nudges. To address these concerns, Lembcke et al. (2019) have identified a set of widely accepted ethical considerations that should guide the design of nudges. These include ensuring freedom of choice, transparency in the implementation of nudges, and the use of goal-oriented justifications.

Freedom of choice means that individuals should have complete autonomy over their decisions and actions. Digital nudges should not limit or prohibit specific choices (Barton & Grüne-Yanoff, 2015) but rather provide information and encourage individuals to make their own informed decisions. The potential harm caused by nudges that undermine autonomy is significant. Nudges that restrict choices and remove agency can lead to resentment and pushback from those who feel that their freedom has been infringed upon. An example of a digital nudge that violates freedom of choice would be a pop-up advertisement to automatically redirects users to a specific website without their consent (Soe et al., 2020). This undermines the users' autonomy to decide where whether they want to visit the website. Such a digital nudge would also be manipulative, as it forces the users to view the website without giving them a choice. This violates the ethical principles of digital nudging, which should always prioritize informing and empowering users to make their own decisions.

In contrast to the previous example, a digital nudge described by van der Laan & Orcholska (2022) utilizes a self-scanning function within an app to suggest healthier alternatives when a user scans an unhealthy food item. Their research demonstrated that the app design, with prompts and suggestions for healthier options, could encourage users to make better food choices. In this case, the app designer didn't use the nudge to restrict users' food choices or enforce a diet plan. Instead, the nudge provided users with information that helps them to make an informed decision about their food choice while nudging them to develop healthy habits.

Transparency is another crucial ethical consideration that should be considered when designing nudges. When using digital nudges, it is important to ensure transparency and openness so that individuals can easily recognize when they are being nudged. This can be achieved through various means, such as highlighting the digital nudge with borders, textual hints, or other recognizable features (Lembcke et al., 2019). Transparency is important to build trust and credibility with individuals and to ensure that nudges are not perceived as manipulative or coercive. For example, an online shopping website may use digital nudges to encourage customers to purchase more items by highlighting specific products as a limited-time deal compared to other deals (*30 Best Examples of Nudge Marketing in ECommerce*, 2022). However, this nudge violates the transparency rule because it tries to exploit the customer's fear of missing out without informing them that these recommendations are based on their previous search history and purchase behavior.

In contrast to the previous example, a digital nudge through email may be used by a teacher to remind their students about an upcoming project deadline (Williams, 2021). The email could mention that past students had taken a week of hard work to complete similar projects and suggest that the students start working on their projects. The goal of the email is to help students refocus their attention on their projects and suggest them either complete the work or ignore the email. This transparent approach helps to ensure that students are aware of being nudged and allows them to make their own decisions.

Goal-oriented justifications are also essential ethical considerations when designing nudges. Nudges should be designed with a clear and justifiable goal in mind, such as improving learning outcomes, app engagement, health outcomes, or promoting environmental sustainability. For instance, Teuber et al. (2022) used a nudge sent through email where students were prompted to take breaks during their study sessions. In the email, students had 5-7 minutes videos with guided physical exercises and health-promoting explanations without a clear or justifiable goal in mind. The intent was to improve the students' mental well-being. However, it violates the goal-oriented justifications rule as the students are unable to understand its goal. The result of this nudge also shows no significant impact.

In contrast, a language learning app (Pham & Chen, 2018) uses reminder digital nudges to encourage users to practice their language skills by sending reminders to complete daily lessons and quizzes. The app clearly communicates the goal of improving language proficiency and provides feedback and progress tracking to help users achieve their language learning goals. This aligns with the goal-oriented justifications rule, as the nudges are designed to promote a clear and justifiable goal that is aligned with the interests of the intended population. Overall, ethical considerations are critical when designing nudges. They help to ensure that nudges are implemented in a way that respects individual autonomy, promotes transparency, and is goal-oriented toward their interest.

Furthermore, digital nudging has been vastly adopted and is constantly evolving in the digital design space, especially in the user interface (UI) field. However, there is still a need for research studies that could provide a better understanding of the theoretical mechanisms explaining the digital nudge. In this way, future researchers will be able to ground their investigation on a sound theoretical understanding of nudging and make informed designs, especially for the development of persuasive technology (e.g., behavior-change support systems; Oinas-Kukkonen, 2010).

Building upon the previous discussion, this study aims to expand the literature on digital nudging by carefully designing and incorporating digital nudges to influence students' engagement with educational applications. The study will make two significant contributions to the digital nudging literature. Firstly, it explored the effects of digital nudges on students' engagement with educational applications. This investigation will provide valuable insights into how digital nudges can be designed to promote better students' engagement. Secondly, the study discussed the design of digital nudges that can be used in the mobile educational application, addressing a gap in the current literature.

Nudge design approaches in Human-Computer Interaction (HCI)

Human-computer interaction (HCI) is a multidisciplinary field that is defined as "a discipline that is concerned with the design, evaluation, and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them" (Hewett et al., 1992, p.5). Therefore, the majority of HCI focuses on designing user interfaces for engaging and efficient interaction between the user and the Information technology (IT) artifacts (e.g., websites and software applications). The aim of HCI research is to improve the user experience (Grudin, 1992) by making it more natural, efficient, and enjoyable. In this context, the concept of nudging has emerged as a valuable and natural choice tool for interface designers seeking to guide human-computer interaction (Jesse & Jannach, 2021). By using nudges, designers can encourage users to make optimal choices and achieve their goals more effectively. As a result, nudging has become an increasingly important area of study in HCI research, offering valuable insights into how to design the digital environment for better user engagement and satisfaction.

Seeing the relevance of nudging in the HCI domain, it has been eagerly adopted and applied in several contexts, such as promoting healthy behavior (Lee et al., 2011), encouraging attention toward privacy settings while using applications (Harbach et al., 2014), protecting the unintended disclosure on the mobile application (Wang et al., 2014), and many more (see Caraban et al., 2019). HCI researchers believe that digital nudges can effectively facilitate the users' decision-making in the digital space as they often have to make multiple decisions and may lack the ability to make informed choices (Mirsch et al., 2017). An example of nudging in HCI is the study by Turland et al. (2015), who developed an application that used color coding and presentation order to nudge users towards choosing a safer Wi-Fi network. Given the potential impact of nudges on user decision-making, the nudging can be considered an important skill for HCI researchers when designing a new interface.

With the increasing use of digital technologies and the constant need to make decisions, the nudging can help users make better choices without feeling overwhelmed. The use of nudging in the domain of HCI has gained significant attention, leading to a myriad of literature. For instance, Harbach et al. (2014) used nudge to appraise the users about the risks associated with giving app permissions. They did so by redesigning the dialogue of the Google Play Store. Their results revealed that it improves the user's attention when giving permission to the application. Lee et al. (2011), on the other hand, made use of three cognitive biases to develop a robot that nudged users toward healthy eating. Their result showed that the nudging interventions were effective but directly depended upon the user's awareness of the healthy choices.

Furthermore, the literature suggests that the majority of recommendation systems make use of nudging because their purpose is to present the users with their desired choices (Jesse & Jannach, 2021). For instance, Jung et al. (2018) discussed an adaptive financial recommendation system that can detect the situation when the user is prone to decision inertia (i.e., the tendency to repeat previous choices). By detecting the situation, the system adapts the interface elements in a way to nudge them to make informed financial decisions. Additionally, Caraban et al. (2019) conducted systematic literature and discussed the most commonly used nudging interventions in the HCI. They identified 23 different mechanisms of nudging and classified them into six categories based on the purpose of nudging. These six categories include facilitating (reducing cognitive effect to motivate the users; e.g., Egebark & Ekström, 2016), confront (pause to reflect on their choice; e.g., Agapie et al., 2013), deceive (provide a deception to promote desirable behavior; e.g., Cockburn et al., 2015), social influence (take advantage of user tendency to social norm adherence; e.g., Cheng et al., 2013), reinforce (increase the presence of desirable behavior; e.g., Ferreira et al., 2014), and fear (evoke the feeling of fear, loss or uncertainty; e.g., Kaptein et al., 2015).

In HCI literature, the digital nudge has been generally employed by designing an interface, providing information, or modification of interactive interface elements in a way to leverage users' cognitive biases to achieve desired user behavior, ultimately improving the interaction between users and IT artifacts. Following are some commonly used approaches in the HCI research to design the nudges for achieving the desired behavior.

Interface presentation

Interface presentation is an approach used to design the nudges in a way that modifies the user interface elements (e.g., informational components and input controls) to encourage a particular choice. The surveyed literature shows that the HCI research has abundantly used this approach to design the nudges (Brewer & Jones, 2015; Chittaro, 2016; Gouveia et al., 2016). For instance, Cai & Xu (2008) investigated the impact of the nudge on the users' consideration set and then purchasing the high-quality product. For this nudge, authors presented products using different sorting (i.e., descending, ascending, and random) based on the quality attributes and observed the user behavior. Their result found that the descending sorting of the products based on their quality attribute encourages the users to consider and then buy high-quality products. Similarly, another study (Kammerer & Gerjets, 2014) designed and investigated the facilitator nudge (i.e., presenting the search results in grid view) to improve the users' selection of trustworthy material on a search engine (e.g., Google). The result of the study indicated that users are more likely to select trustworthy pages if the search results are in the grid view rather than the traditional list view in the search engine.

Social comparisons information

In this approach, the HCI researchers design nudges by providing peer or social behavior information as a reference point (Brown et al., 2019) to guide user's decision making. In addition, the researchers have argued that this approach needs to be carefully used to design an effective nudge because it could backfire as we are enabling the users by providing a reference point in the form of their actions or people's behavior. For instance, showing below-average user performance can mislead users to adjust their behavior accordingly. Moreover, studies have shown that enabling the users' choice by showing the performance of similar users has proven to be effective. For instance, Eckles et al. (2009) used this approach to nudge the users into answering a particular question. They showed the percentage of the participants who answered a certain question, which led to the increased submission of the answer. Similarly, another study (Caraban et al., 2015) discussed the development of an electric plugin for a teeth brushing device, where the device tracked the user (child) brushing behavior and provided information to the parent through light on the device.

Digital alerts/feedback

Another commonly used approach for nudging in the HCI domain is to remind the user of a particular choice/behavior (Nekmat, 2020). These approaches have broadly used digital prompts (e.g., dialog box, push notification) and real-time feedback. For instance, Hirano et al. (2013) used 'WalkMinder,' a mobile phone application, where the users were nudged with an alert (i.e., buzz sound) whenever they did not work for a long time. Their study showed that the nudge was effective in making the users aware of their activity patterns.

Another study (Gouveia et al., 2016) incorporated nudging using this approach by showing only last-hour physical activity reports in a smartwatch. Thus, making the feedback scarce and encouraging the users to be more physically active. Their result revealed that this nudge has effectively increased the user’s physical activity. Even though these approaches have proven to be effective in influencing the users’ behavior, researchers have suggested that the studies using this approach should follow certain design considerations; otherwise, they may not be effective (Caraban et al., 2019). These design considerations are: 1) Timing: it is critical to consider the time when it will be most effective. 2) Frequency: The nudge designers need to be careful about the frequency of this approach as it could frustrate the users. 3) Personalization: this approach requires that the nudges be personalized for a given situation and target a particular behavior.

Default options

In HCI literature, another approach used to nudge the users is to pre-select the desired option as default (Paunov et al., 2019). This approach has been commonly used in studies to nudge users to make sustainable choices, such as their willingness to donate, food selection, or opt for green energy services (Lemken, 2021). For example, one of the exemplary studies using this approach is (Egebark & Ekström, 2016), which modified the default printer option to a “double-sided printer.” Their study revealed that this has effectively reduced printing paper consumption. Another study by Al-Ameen et al. (2015) used a CueR, a novel cued-recognition authentication scheme that provides users with random and memorable passwords based on their selected clues. Their result revealed that all users was able to recall their password after one week of registration.

Although this approach is a powerful way to nudge the users, the researcher has echoed caution while using this approach, as this could have ethical repercussions (Paunov et al., 2019). Therefore, the nudge designer must use them ethically and transparently. By making the default choice align with the user's interests and values, designers can nudge users toward making decisions that benefit themselves and society.

Throttling mindless activity

Throttling mindless activity is another approach used to nudge the users in the HCI literature. Mindless activity is the act of engaging or performing a task with little conscious thought or effort. In this context, throttling refers to restricting or regulating the user's engagement with any mindless activity. Therefore, the idea behind this approach is to pause the activity for some time and give users a chance to re-evaluate their choice (Bergram et al., 2020). Literature has shown that nudging the user through this approach has improved the user's ability to focus and engage with the activity(Caraban et al., 2019; Konstantinou et al., 2019). For instance, Wang et al., (2014) designed a plugin for the Chrome browser that withheld the publication of a Facebook post for 10 seconds, inciting the re-examination of the post's content. Although the countdown could have been avoided, their study revealed that several participants reformulated the content and even abandoned the publication during the time interval.

In addition to the approaches discussed, the HCI researchers have also explored different frameworks to study the cognitive biases of user behaviors that can be mitigated using optimal nudging interventions (Caraban et al., 2019). For example, Hansen & Jespersen (2013) put nudges into four distinct classifications on the basis of the thinking mode they target (i.e., automatic vs. reflective) and whether the user is aware of the purpose and means to bring about the behavioral change. Dolan et al. (2012) suggested a framework called "Mindspace" which encapsulates nine essential behavioral influences. The mnemonic "Mindspace" stands for nine important aspects that influence human decision-making in this framework. These aspects include messenger, incentives, norms, defaults, salience, priming, affect, commitment, and ego. Nudge designers can utilize these behavioral influences to develop effective nudges aligned with the desired goals. Additionally, Schneider et al. (2018) linked cognitive biases with the choice, i.e., binary, discrete, or continuous, and the various interface elements that can be used to influence behavior through interface elements such as checkboxes, radio buttons, and drop-down menus.

Despite the widespread adoption of nudging in HCI and promising results in this particular area, there are a few concerns regarding the use of nudging. First, the HCI researchers are not sure about the ability to nudge to bring a sustainable behavioral change. For instance, Egebark & Ekström (2016) found that although the impact of the “double-sided print” option remained for a few months, it began to fade with the use of new printers that had the option of single-sided print as the default. On the other hand, Rogers et al. (2010) found that the impact of twinkling lights (a nudge) that showed the closest path to the staircase remained for more than eight weeks and beyond, even after the removal of the nudge.

Another concern is that nudge execution may yield unexpected results since people often compensate for their actions (e.g., people might prefer to print more in case of the double side so they could carry less weight) and unexpected interpretations (e.g., once households were shown that they were below the average electricity consumption they started to consume more electricity). For instance, Kankane et al. (2018) found out that the users did not create their own password once they were made aware that they would receive an automatically generated password. Additionally, Munson et al. (2015) found out that people made lesser commitments once they were asked to make their commitments public because most of them feared criticism. Gouveia et al. (2016) found that user motivation increased because of enabling social comparisons in cases where the user performance was similar in comparison to others. This also implies that nudges can have an adverse impact in cases where this condition cannot be met. To that extent, the majority of studies did not inquire into possible backfires and unexpected effects of nudges. Therefore, a stronger emphasis on understanding the underlying behavioral mechanism is required to better design and utilize effective nudging.

Educational application, engagement, and nudging

Since the Internet revolution, there has been enormous growth in the design and development of educational apps (Papadakis & Kalogiannakis, 2017). Educational apps are software designed for educational purposes that can be used in smart devices such as smartphones and tablets (Geissinger, 1997). The educational app aims to make education more effective compared to the traditional means of education, leveraging the latest technology advancements to enhance the overall learning experience (Rocha & Coutinho, 2015). Also, these apps have helped educational researchers to redefine the learning experiences by providing a platform to promote self-study (e.g., He, 2018), teaching support (e.g., Littlemore & Farmer, 2014), and making education accessible to the remotest areas of the world (Sruthi & Mukherjee, 2020).

Seeing the importance of educational apps, a vast literature on their design and development has emerged (e.g., Muslimin et al., 2017; Singler et al., 2016). Educational apps have now been readily integrated into both traditional and non-traditional educational settings to serve varied learners (Kayalar, 2016; Tularam, 2018). For instance, these apps have been employed to teach people in diverse domains such as education (e.g., Cavus, 2011), health (e.g., Lin et al., 2018), and even moving towards standalone personalized learning platforms such as Coursera (Korableva et al., 2019), Byju (Sruthi & Mukherjee, 2020), and SoloLearn (Quinn, 2018).

Educational apps can take many forms, with the most popular ones being games, simulations, mobile apps, and virtual reality (Ivan Stojšić et al., 2017). The most popular type of educational app is games based on the concept of Gamification, which involves incorporating the game elements such as points, badges, and leaderboards to make learning more engaging and enjoyable (Mayer, 2019). Another type is the simulation applications, which allow learners to visualize the problem, practice skills, and apply knowledge in emulated real-time scenarios (Kincaid et al., 2003). These apps are popular in STEM education fields such as healthcare, engineering, and aviation (D'Angelo et al., 2014). Another commonly used educational app is the mobile application, which has made learning portable by providing simple access to the study material on students' cell phones or tablets (Bustillo et al., 2017). Lastly, Virtual reality(VR) applications allow students to experience environments and situations similar to those in the real world (Kavanagh et al., 2017), whether exploring historical sites, human body, or conducting experiments in a lab.

Educational app is also becoming relevant in the educational domain as they are being increasingly used to deliver/manage content, gamify educational topics, or facilitate course management for both teachers and students (Hirsh-Pasek et al., 2015). For instance, most educational institutes have at least one learning management system, such as the Blackboard learning management system (Ashok, 2011; Y. Hwang & Vrongistinos, 2012) and Brightspace (Francom et al., 2021; Jiang et al., 2020). Furthermore, the recent pandemic has also seen massive influx of educational apps because of the remote teaching. Among these apps, the most important emergence was the video conferencing applications such as Microsoft Teams and Zoom (Cavus & Sekyere-Asiedu, 2021). These app became virtual classrooms and primarily used to deliver lectures in the pandemic (Alameri et al., 2020).

To this end, the researchers have the consensus that educational apps are an effective tool to facilitate learning due to their potential to provide real-life experiences, provide scaffolding in complex cognitive tasks (e.g., solving math equations; Kim, 2016), personalized feedback (Gielen et al., 2018), and building collaboration among peers (Loughry et al., 2013). However, studies have pointed out the issue of learners' engagement with the educational application (Pham & Chen, 2018), an essential component to realize its full potential and improve students' learning outcomes. The majority of the literature on the educational app is focused on the design and development of educational applications and their impact on students' learning outcomes (e.g., Kim, 2016; Muslimin et al., 2017). However, limited studies have explored the learner's engagement with the educational application (Pham & Chen, 2018).

In literature, app engagement is defined as an augmented user experience that combines the behavioral, cognitive, and emotional aspects of an application's usability (Doherty & Doherty, 2018). Researcher has argued that when learners are fully engaged with an educational app, they are more likely to acquire knowledge and skills effectively, resulting in better learning outcomes (Papadakis et al., 2018). Conversely, if the students are disengaged with the application, it may lead to decreased motivation and even abandonment of the app.

Furthermore, prior studies have also demonstrated that there is a remarkable attribution rate (i.e., the proportion of app installations and usage) due to a lack of students' engagement with educational apps in different educational domains such as linguistics (Godwin-Jones, 2011), science (Zydney & Warner, 2016), and children's education (Mkpojiogu et al., 2018). Shuler (2012) analyzed the educational category in the Google and Apple store, two of the most famous popular mobile application distribution platforms. The result revealed that almost 9 to 23 % of applications were used only once and then deleted forever. Additionally, users opened only 39%

of the educational apps more than 11 times (Haggerty, 2019). It is interesting to note here that educational apps form the group of applications with the lowest retention rate (*User Retention Rate for Mobile Apps and Websites*, 2017). Student engagement that leads to application retention has therefore become an area of interest among researchers.

Despite the growing interest in this area, the literature on learners' engagement with educational apps is still in its early stages (Kim et al., 2017; Pham et al., 2016). Studies and experts have pointed out that educational apps are currently not working up to their full potential because of their lack of engagement (Pachler et al., 2009; *User Retention Rate for Mobile Apps and Websites*, 2017). Haslam (2019) also stated that various useful educational apps get deleted before the learner explores all their features. Therefore, future studies on engagement are of paramount importance in improving the effectiveness of educational apps.

Seeing the increasing interest in improving the learners' app engagement, studies have explored different mechanisms to improve the learners' app engagement, such as effective UI/UX design (*What Makes a Good Retention Rate?*, 2019), interactive designs (Pham & Chen, 2018), and continuous updates with fresh content (Grguric, 2023; Jay, 2015). Among the ongoing effort, the idea of nudging has been used in the past and is getting increased attention in the educational application literature to improve students' app engagement. For instance, Weinmann et al. (2016) stated that the user interface designers had been acting as choice architects knowingly or unknowingly since they influenced people's decisions. Through their design, they nudge people to navigate the online decision environment.

Nudging is a powerful tool that can be leveraged to enhance the behavioral, cognitive, and emotional aspects of app engagement in educational apps. Regarding the behavioral aspect, nudges can encourage users to engage more consistently and productively with the app. For instance, digital nudges can be used to remind the students to complete tasks (Motz et al., 2021), participate in discussions (Nguyen & Vo, n.d.), or scaffold the students (van Oldenbeek et al., 2019) that can help students to interact continuously with the app. As a result, nudges can help develop desired student habits that can raise the user's overall engagement with the apps and improve their learning outcomes.

Nudging can also impact the cognitive aspect of app engagement by promoting learning and skill-building. Educational apps can use nudges to prompt users to review material they have previously studied, practice specific skills, or solve problems (Dawood et al., 2013). These nudges can help to reinforce learning, improve retention, and build cognitive abilities over time. Additionally, nudges can be personalized based on users' learning progress and preferences, which can make them more effective and engaging (Brown et al., 2019). Finally, the nudging can impact the emotional aspect of app engagement by providing positive reinforcement and feedback. Educational apps can use nudges to acknowledge when a user has achieved a certain milestone or completed a challenging task (Edwards & Li, 2020), which can create a sense of accomplishment and pride. This positive reinforcement can help to build motivation and confidence, which can be crucial for maintaining engagement over time.

In educational application literature, researchers have used nudging to increase the student's engagement with the application without realizing it. The closer aliases used instead of nudging in previous studies includes push notification (Simmons et al., 2018), text messages (Harley et al., 2007), and alerts (Nekmat, 2020). Following are a few nudging interventions that have been used to improve the learners' app engagement.

Reminder nudge

In the literature, the most commonly used nudge intervention is the use of reminders to keep the student engaged with the application. Studies using the reminder nudge strategy targeted students' behavior, such as students' refocusing issues (Simmons et al., 2018) or their limited attention span issues (Castleman et al., 2014; Castleman & Page, 2015). By targeting these behaviors, the intention was to keep the students aware of the educational application and hence increase their engagement with the educational apps. Weston et al. (2015) investigated the engagement of patients in health-based educational apps. In the application, the patients were educated about their health, which helped them make informed decisions. To improve the learners' app engagement, studies employed reminder nudges, where patients were nudged three times to play a quiz. Their result revealed that the initial nudge was highly successful, resulting in increased participation of the users. However, the next two nudges were not as successful.

Furthermore, in the education technology literature, reminder nudges through push notifications have produced some promising results on students' app engagement (e.g., Pham et al., 2016). It seems like a natural choice in education apps because it maps to the principle of any intervention to qualify as the nudge. These principles are cost-effectiveness, autonomy, and suggestiveness. For instance, Manser (2016) discussed the University of San Diego's (USD) two nudge interventions for student engagement and communication centers. These interventions included sending messages through push notifications and messages. Their result revealed that the push notifications were three times more likely to be seen by the students as compared to other interventions. Overall, reminder nudge has produced some promising results in keeping the student engaged with the educational application. However, it has also been discussed as a source of disruption (Cutrell et al., 2001) for the users. Therefore, it could be one place for future research to start the inquiry and discuss reminder nudging through push notifications and students' app engagement.

Social comparison nudge

Another nudging intervention used by the studies was the use of a social comparison nudge (information about the behavior of relevant peers). Such nudges exploit students' tendency to adhere to the social norm. For instance, Brown et al. (2019) used a web-enabled coaching system to implement online personalized social comparison nudge. In their study, they alerted the students about the behavior of their peers that have outperformed them. This way, the intention was to increase the students' interaction with the application. Even though their study did not find any significant relationship between employed nudges and the student behavior, it provided a number of suggestions that can be used to enhance the nudge research. One of the suggestions is that future studies implementing the social comparison nudge need to be careful about their design because it may impact students' behavior negatively. For instance, if you share such nudges with well-performing students, they may become overconfident and reduce their work effort.

Even though the social comparison nudge has proven to be effective in other contexts (Szaszi et al., 2018), there are limited studies exploring the effect of this nudge strategy in the education application domain. Also, limited studies have explored the type of information about the peers in the social comparison nudge that can be more effective for students' behavior change (e.g., Brown et al., 2019).

Reinforce nudge

Another intervention discussed in the literature is the use of reinforcement nudge (increase the presence of desired behavior options) to promote student engagement in the educational application. For instance, Bruehlman-Senecal et al. (2020) designed an application to tackle the learners' loneliness during the transition from college to university. In this study, the students enrolled in the application were nudged to complete a social challenge, a reflection exercise, and present the student testimonial. All these activities were introduced to encourage students to develop social connectedness with their peers. Their studies collected the number of app page access, social challenges, and a number of task submissions to measure the learners' app engagement. Their study found improved app engagement where the students were nudged from the start of the study rather than delayed nudges.

Informational nudge

Few studies have used the informational nudge intervention to keep students engaged with educational apps. This nudge provides information to individuals to help guide their decision-making. This information nudge has been specifically employed to enable learners to self-assess their ability and behavior by providing them with personalized feedback. This way, the learners' interaction with the application is improved. For instance, Arnold & Pistilli (2012) designed a nudge intervention, where students were provided with an alert to refocus their engagement with the application. Their study hypothesized that the learners' increased app engagement would improve their academic performance. Hence, their study collected the learning analytics (e.g., previous performance, interaction with the LMS, and demographics). Their results were quite promising as this nudge strategy increased students' app engagement, as well as their performance. Also, literature referred (e.g., Karlsen & Andersen, 2019) to the above-mentioned nudge as the smart nudge (tailored according to the user's behavior). However, they discussed that ethical considerations should be kept in mind while designing such nudges to avoid misuse. Therefore, future research is needed to design and evaluate the nudges not only on the basis of their outcome but through ethical considerations as well.

Overall, the surveyed literature showed that most studies exploring nudging and engagement in educational apps have relied on educational application analytics (the measures thought to affect engagement) and established their relationship with students' engagement. Also, few studies have used a large sample size of students' user analytics to understand their engagement with educational apps. For instance, Pham and Chen (2019) used a large sample size of 95,430 learners. Broadly, the engagement measures used in the surveyed literature can be divided into two categories, i.e., intra-session, or inter-session engagement measures, as shown in table 1.

Table 1: Different measures to assess students' engagement

Measures	**Examples**
Inter-session engagement measures	Session length, session count, average duration per day, attempt per quiz, initial accuracy, session count, session length
Intra-session engagement measures	App retention, total time consumption, uninstall rate, mastery of the task, average student time spent, courses assessed, unique quiz attempted, total time consumption

Furthermore, these above-mentioned engagement measures can be divided into three categories representing their targeted purpose, i.e., popularity (how much the application is being used), activity (the number of user interaction events), and loyalty (the retention of the user within the app) as shown in table 2.

Table 2: Mapping of the categories with a measure

Categories	**Engagement measures**
Popularity	Session length, session count, total time consumption
Activity	Mastery of the task, courses assessed, unique quiz attempted, attempt per quiz, initial accuracy.
Loyalty	Average duration per day, app retention, uninstall rate

While exploring effectiveness of digital nudges on the learners' app engagement, the studies explored the students' increased application interaction or retention with the app. Hence, they used the measurements in popular and loyalty categories to study students' engagement. However, limited studies explored the students' cognitive engagement with educational apps. Therefore, measurements in the activity category were underutilized in the engagement. For instance, Pham and Chen's (2018) used popularity and loyalty measures (session count, session length, app retention, uninstall rate, and average duration per day) on student involvement and retention in the app, while they used loyalty (mastery of the vocabulary learning task) to measure the engagement's cognitive aspect.

Based on the above discussion, it is evident that digital nudges have shown promising results targeting students' behavior to improve the learners' engagement in educational applications. However, there is a further need for experimentation to explore the effectiveness of nudging interventions, specifically targeting the students' behavior while interacting with the educational application.

Nudging in STEM education

The field of STEM education has seen a growing interest in nudging in recent years, in part because it has a low implementation cost and has produced promising results to influence students' behavior (e.g., Bettinger & Baker, 2014; Hoxby et al., 2013). However, the use of nudging in STEM education is still at an early stage. For instance, Szaszi et al. (2018) conducted a recent systematic review of literature on nudging using 156 empirical studies, out of which a mere 4% were related to education, and even fewer were related to STEM education. Hence, there is a massive opportunity for researchers to study and implement effective nudging interventions in STEM education.

In STEM education literature, studies have employed various nudging interventions to help students make better choices and improve their learning outcomes (Damgaard & Nielsen, 2017). One such intervention is a *Reminder nudge* which involves helping students refocus their attention through focused messages/reminders. For instance, O'Neill (2019) used a reminder nudge to increase the participation in advisor activities between students and the advisor by constantly reminding them of the upcoming deadline and events. The purpose was to debase students' beliefs and help them to make informed decisions by taking into account their lack of self-control.

Another effective nudging intervention in STEM education is the *Framing nudge*. This involves a small intentional change in the presentation/orientation of the choice architecture to remove biases and encourage the students toward desired behavior. For example, McEvoy (2016) conducted an experiment in a statistics course where students in the control condition attempted the quiz starting with zero and earned points, while students in the treatment condition started with 560 points and then lost points. The study found that the fear of losing points improved the performance of the students in the treatment condition compared to the control condition.

Deadline nudge is another nudging intervention that works as a commitment device to help students plan for their class tasks and refocus their attention. For instance, Tuckman (1998) conducted an experiment where the students were divided into two conditions. The first condition of students was tested frequently (deadlines), and the other condition was given homework after the completion of each chapter. Furthermore, their study used a validated survey to profile the students into high, medium, and low levels of procrastination. Their analysis revealed that heavily procrastinating students with deadlines (frequent testing) outperformed the students in the other condition.

Lastly**,** *Scaffolding nudge* is another effective nudging intervention used increasingly in STEM education, specifically for teaching programming concepts (e.g., Sticklen et al., 2004) or improving students' writing (e.g., Wanchid, 2013). This approach intends to reduce the students' cognitive load to achieve the desired students' behavior. For instance, Zamprogno et al. (2020) conducted a study where they designed scaffolding nudges to improve the student's learning in the automatic programming assessment tool. By doing so, they were able to help students in re-focusing their effort on understanding their code failure and revisit the learning outcome of the course or assignment, demonstrating the potential for scaffolding nudges to improve student engagement and performance in STEM education.

Overall, the nudge literature has narrowly explored the cognitive biases to target students' behavior in the STEM literature. For instance, the studies have mostly utilized the students' cognitive biases, such as preference bias (i.e., inclination towards the short-term benefits) and limited attention bias, to influence the students' behaviors. These behaviors include students' self-control (e.g., Harley et al., 2007; O'Hara & Sparrow, 2019), attention limitations (e.g., Sherr et al., 2019; Tuckman, 1998), loss aversion (e.g., McEvoy, 2016), and limited will power (e.g., Patterson, 2018). For instance, Sherr et al. (2019) used reminder nudges to invoke students' attention toward course goals and deadlines that can impact their grades. The result revealed that students in the nudge condition performed better than the control condition.

Seeing the growing interest in the adoption of nudging in STEM, several institutions have developed systems that make use of analytics to nudge students (Fritz, 2017). For instance, the University of Michigan developed 'Expert Coach' (ECoach), which is a recommendation system that facilitates students to pass difficult courses by providing a scaffolding nudge using analytics (e.g., performance metrics and ECoach usage) and advice from their peers (McKay et al., 2012). Several studies have shown the positive effect of the ECoach system on students' academic performance in STEM gateway courses (e.g., . Similarly, Fritz (2010) developed Check my activity (CMA), a feedback system, at the University of Maryland to nudge the students with a simple feedback nudge on their usage of the Blackboard learning system (BbLMS). Also, the students can compare their BbLMS usage to an anonymous summary of peer usage. The intention with such a nudge is that the students will become aware of their peer usage of the system, which may help them to change their study behavior or encourage them to seek help.

The common trajectory found in the nudge literature in STEM education is that most studies have introduced nudging to target students' behavior and then measure its impact on students' learning performance. This process involves introducing nudge as an intervention, measuring students' behavior, and understanding its relationship with the outcome measure. For instance, Patterson (2018) introduced the deadline nudge using the reminder as a commitment device to refocus the student by showing their pre-set goals. In this case, the author used the indirect measure of student learning performance rather than the direct method, which is seeing their timely assignment submissions.

Even though the existing literature has employed a number of nudging interventions, it did not explore the relative impact of multiple nudging interventions in a single study. For instance, Sherr et al. (2019) introduced two different kinds of nudging reminders. The first type was used to remind the students to submit a self-monitoring log after each week. The second type was to remind them about course tasks such as quizzes, assignments, and exams. The study focused on the holistic view of the impact of nudging on student learning performance instead of the comparative analysis of interventions. Similarly, Brown et al. (2019) emphasized that future research should explore the relative efficacy of nudge interventions.

As the nudge literature in STEM is at a novice stage, studies have generally employed untargeted and unplanned nudging interventions. For instance, Sherr et al. (2019) used the reminder nudge to refocus students' attention on the course task, such as self-monitoring log or coursework. However, the prompts in the self-monitoring log were not designed to target students' metacognitive thinking, as discussed in the study. Similarly, another study (Kizilcec et al., 2014) also employed the reminder nudge intervention to increase the student's course participation. Still, its analysis didn't account for the information about tracing students' interaction with the intervention, i.e., they did not know whether the students receiving the reminder email were opening the information. Hence, there is a need for studies exploring the students' cognitive behaviors and biases to improve the process of designing effective nudging interventions. Furthermore, there is a need for studies that can appropriately evaluate the effect of nudging on the outcome measure.

Findings in the literature are generally heterogeneous when it comes to the effectiveness of nudging on students' behaviors. Therefore, there is a need to consolidate STEM nudge literature. For instance, Persistence Plus, a non-profit organization, conducted a longitudinal quasiexperimental study (Soricone & Endel, 2019) employing reminder nudge (reminding students of a particular task) and found a positive impact on the persistence of the students in the STEM field. In contrast, Kizilcec et al. (2014) explored the impact of reminders and persuasive reminders in online STEM courses using two separate experiments in a single study. The first experiment showed a short-term impact of reminder nudges on students' behavior (reducing procrastination) that can increase students' course participation. Also, the second experiment using a persuasive reminder nudge showed no short-term impact, and in the long term, the effect was negative on course participation.

In light of the above discussion, it is evident that there is a huge potential for STEM researchers to explore the effectiveness of nudging interventions in STEM education. In addition, the broader educational literature can be considered a guiding point for STEM researchers. For instance, Damgaard & Nielsen (2018) conducted systematic literature and identified 12 broadly used nudging interventions in education. Also, the authors acknowledge that these categories are by no means exhaustive due to constant increases in the literature. Therefore, STEM researchers can explore the broader range of nudging interventions discussed in the education literature.

RESEARCH METHOD

This study aimed to investigate the impact of nudging interventions on students' engagement with CourseMIRROR. In the educational literature, application (app) engagement is described as the improved level of interaction and involvement a learner has with an application (Anwar et al., 2022; Pham & Chen, 2018; Song et al., 2022). It is a multidimensional construct influenced by behavioral, cognitive, and emotional components of the user's experience (Kelders et al., 2020) with the app. This chapter begins with a description of CourseMIRROR, site of experiments, description of app engagement measures, and core principles used to design the nudge interventions. Finally, I provide a description of three experiments conducted to explore the impact of nudging interventions on students' app engagement. These interventions were specifically designed to improve the students' app engagement and effectively implement CourseMIRROR in a real classroom setting.

CourseMIRROR application

CourseMIRROR (i.e., mobile in-situ reflections and review with optimized rubrics) is a mobile educational application designed to engage students in reflective thinking and self-evaluation of their learning experiences (Fan et al., 2015; Menekse et al., 2018). CourseMIRROR prompts students to reflect on the confusing or interesting aspects of each lecture throughout the semester using their personal mobile devices (e.g., smartphones or tablets). Subsequently, a Natural Language Processing (NLP) algorithm is used to create students' reflection summaries for confusing and interesting questions by combining their reflection submissions based on common themes (Magooda et al., 2021). The summaries are made available to the students through the mobile application. These summaries help students to conceptualize the difficulties faced by their classmates. Also, the associated website is designed for the instructors, where they can see students' reflection summaries and individual reflections along with other analytics to inform their pedagogies for the class. Figure 1 shows interfaces of the CourseMIRROR mobile application.

a) b) c) d)

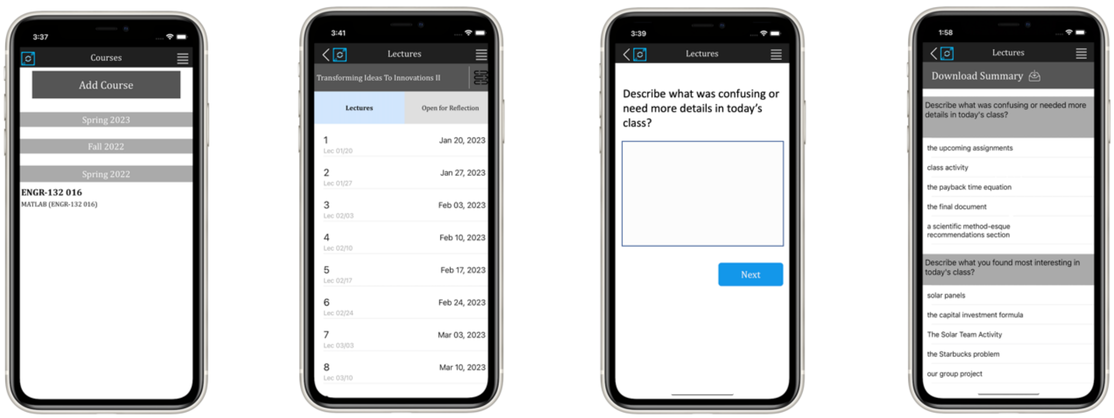

Figure 1: Primary interface of the CourseMIRROR a) Course screen, b) Lecture screen, c) Reflection writing screen, and d) Reflection summary screen

The CourseMIRROR mobile application is available to download for free on both Apple Appstore and Google Playstore. The students can log in to the application using their credentials (i.e., email and password) and get authenticated through the application's server. Once the students log into the application, they can see their enrolled courses and register in a new course by clicking on the "Add Course" button (figure 1a) and inputting passcode created by instructors on the associated site. Furthermore, the students can go to the corresponding lecture by clicking on the course item. The lecture screen is subdivided into two categories, i.e., Lecture tab (contains all lectures except those available to write reflections) and "Open for reflection" tab (contains only those lectures available to write reflections), as shown in figure 1b. The students can go to the reflection page (figure 1c) to submit and write a reflection by clicking on the lecture item under the “Open for reflection” tab. Once submitted, the reflection is stored in the application’s server database. After more than five students’ reflections are submitted to the database and the reflection writing time ends (lecture time + lecture duration + 36 hours), the NLP algorithm generates summaries of the reflections. Then, the summaries are shared with students and instructors on the CourseMIRROR app (figure 1d) and the instructor site respectively.

For the CourseMIRROR app to be an effective instructional tool, we need to support and scaffold students to generate detailed and comprehensive reflections, and encourage them to read reflection summaries and learn from them (Menekse et al., 2018). Achieving these goals can be challenging, especially in a technology-mediated learning context in a college classroom. To address this challenge, this study implemented nudge interventions within the CourseMIRROR application and explored their impact on students’ app engagement. The interventions were designed to promote the three core purposes of the application (Menekse et al., 2018). These purposes include 1) to enable and support students to submit reflections using the application, 2) to enable students to access the reflection summary interface and learn from peer reflection summaries, and 3) to enable students to provide in-depth and relevant (to the prompts and lecture content) reflections for the system to make sense and provide insight to the instructor/fellow students. Overall, this study aims to improve students’ app engagement by incorporating nudges that are tailored to the specific needs of students.

Site

Purdue University is a land grant research university located in the midwestern region of the United States. The first-year engineering program at Purdue enrolls around 3,347 undergraduate engineering students in the year 2021-2022, out of which women students represent

26.4%, while underrepresented minority students make up 9.7% of the total enrollment (*Purdue Engineering Degree Programs & Enrollment 2021*, n.d.). These students take courses in STEM and communication, including ENGR 131 and ENGR 132, which teach engineering design, data visualization, computer programming, communication, and teamwork. ENGR 132 focuses on introducing programming skills using MATLAB to first-year engineering students. It is offered year-round and is taken by the majority of the FYE students. ENGR 132 is a two-credit course that utilizes active, blended, and project-based learning methodologies, including lectures, pair programming in class, online modules, and team projects outside of class. Each section of the course enrolls up to 120 students. This study collected data from ENGR 132 classes in Spring 2022.

Engagement measures

In the literature, app engagement in educational apps has been defined as an improved user experience that encompasses behavioral, cognitive, and emotional aspects of its usability (Doherty & Doherty, 2018; O'Brien & Toms, 2008). Prior research has typically relied on learners' application analytics, such as session counts and quiz attempts, to inform their analyses (e.g., Brown et al., 2019; Bruehlman-Senecal et al., 2020; Pham & Chen, 2018). This study also relied on the student usage analytics provided by the CourseMIRROR app to assess the effectiveness of designed nudges on students' app engagement.

The app engagement measures (usage analytics) used in this study include reflection submissions (the number of times students submit reflections), reflection summarization views (the number of times students see the summary interface), reflection specificity score (relevancy score of reflection with the question prompts and lecture), and reflection text length. A detailed description of the app engagement measures is provided within each experiment in the subsequent sections. This study aims to use these engagement measures to understand the impact of digital nudges on students' engagement with the CourseMIRROR app.

Nudge interventions

In light of the literature (Damgaard & Nielsen, 2018), nudge intervention should adhere to the following three principles:

1. Simple: a nudge should use a language that is easily comprehensible for a large audience.
2. Cost-Effective: a nudge should not introduce additional costs to the design process.
3. Suggestive: the language of nudge should suggest the user is directed toward a desirable behavior.

The nudges used in this study are designed based on the above-mentioned core guiding principles. The study employed various nudge interventions, including reminder nudges, which remind students' of their intended behavior or action (Simmons et al., 2018), social comparison nudges that leverage social influence by providing peer information to influence behavior (Brown et al., 2019), interface nudges that use interface design to encourage or discourage specific behavior (Schneider et al., 2018), throttling mindless nudges, which interrupt mindless behavior to encourage thoughtful decision-making, and scaffolding nudges (Caraban et al., 2019), which provide additional support to the learners for making informed decisions. All nudge interventions used in the experiments were selected based on 1) the literature where these nudges were shown to be effective in achieving the desired behavior as intended in the experiments, and 2) thorough discussion with the research team to ensure whether the introduced nudge interventions were aligned with the experiments' goals. A detailed description of the nudges is provided in the subsequent sections, describing each experiment.

EXPERIMENTS

This study conducted three experiments that investigated the impact of nudging interventions on students' engagement with the CourseMIRROR. In this regard, each experiment explored nudge interventions designed to improve students' app engagement with the application's core purpose (Menekse et al., 2018). These core purposes are 1) facilitating student reflection submissions, 2) increasing their visit to peer reflection summary interface, and 3) scaffolding students to write in-depth and relevant reflections. This study was conducted in accordance and guidance set by the Institutional Review Board (IRB) at Purdue University. Each experiment was conducted using experimental research design and their data was then analyzed using quantitative approaches (e.g., mean comparisons). I used IBM SPSS statistics (v. 29.0) to conduct the data analysis. Moreover, assumption testing was performed before conducting the analysis. However, I discussed only those assumptions that were violated in the analysis. I also discuss the alternate analysis techniques used due to violation of certain assumptions.

Experiment 1: Facilitating students' reflection submissions

One of the primary purposes of the application is to encourage and facilitate the students' reflections submissions for a classroom. The literature suggests that students often tend to forget things, and reminders nudges have proven to be effective in reminding students to complete a task (Damgaard & Nielsen, 2018). Therefore, this study will introduce two different reminder nudge interventions and explore their impact on the students' submission of reflections using CourseMIRROR app. The study hypothesizes that the employed nudges would remind students to submit reflections and increase students' reflection submissions. More specifically, this experiment was guided by the following four research questions:

1. Do students receiving neutral reminder nudges submit more reflections compared to the students receiving no nudge?
2. Do students receiving social comparison nudges submit more reflections compared to the students receiving no nudge?
3. What is the relative effectiveness of both nudge interventions on the students' reflection submissions?
4. How do the students' reflection submissions change over time in each condition (i.e., neutral reminder nudge, social comparison nudge, and baseline)?

Participants

In this experiment, 181 undergraduate students from three sections of a first-year engineering programming course (ENGR_132) at Purdue university were recruited to participate. All of the students in the study used the CourseMIRROR application. Over the course of the semester, participants submitted a total of 1768 reflections over 24 lectures and on average, each student submitted around 10 reflections. I only used the students' reflection submission information from the first 24 lectures of each course section for the analysis. I excluded last week's lectures (i.e., final's week) from the analysis due to low reflection submissions.

Reflection submissions

In this experiment, I used the students' reflection submissions in the CourseMIRROR application to measure students' app engagement. The reflection submission is counted as the total number of times a student submits the reflections for a course. For instance, if the course has 24 lectures, then the reflection submissions for a student can range from 0 – 24, answering all four questions in the single reflection as shown in the figure 2. Also, I did not use the partial reflection submission by the student in the experiment. Therefore, each time a student submits a reflection after a lecture, it indicates that they are actively using and engaging with the application. In other words, the more frequently a student submits reflections, the higher their level of engagement with the app. This experiment aimed to measure the effectiveness of nudge interventions to improve students' reflection submissions, which is a behavioral aspect of their app engagement. Also, this experiment only used the students' data with more than 2 reflection submissions.

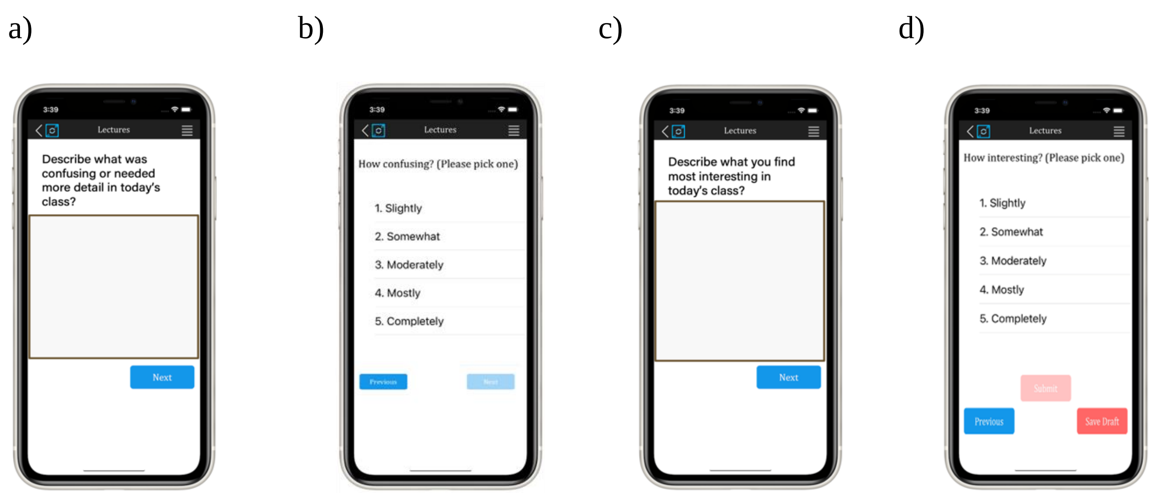

Figure 2: Reflection activity comprises of four questions in order from a - d

Nudge interventions

In this experiment, I introduced the following two reminder nudge interventions to improve the students' reflection submissions.

Neutral reminder nudge

This nudge is a subtle form of digital prompting that provides non-intrusive cues to users without making any explicit suggestions or recommendations (Simmons et al., 2018). I used this neutral reminder nudge in the CourseMIRROR app to refocus students' attention toward submitting their reflections after each lecture. This intervention used push notifications to nudge students, a pop-up message sent by a system to their mobile application. Two push notifications were sent to students with varying levels of information to nudge them to submit reflections. The message in both the notifications were thoroughly discussed by the research team to ensure their validity and reliability. Following is the description of both push notifications:

Notification 1

For the first nudge, push notifications were sent to the students at the end of each lecture, suggesting they submit their reflections. This push notification contained a message, i.e., "*Lecture (Number) is open to write a reflection for (class code)*". This nudge will be named Reminder 1.0 for future reference.

Notification 2

For the second notification, students were nudged after six hours of Reminder 1.0 with a push notification. The second notification contained the message i.e., "*Reminder! Lecture (Number) is still open to write a reflection for (class code)*". The second reminder is used only for those students who would not have submitted the reflection by then. This nudge will be named Reminder 2.0 for future reference in the study.

Social comparison nudge

This nudge uses social influence or comparison to encourage users to engage in certain behaviors or activities (Brown et al., 2019). In this experiment, this nudge intervention is similar to the neutral reminder nudge in reminding students through push notifications about the reflection submission. However, unlike the neutral reminder nudge, social comparison nudge includes information about peer behavior in the notification message, allowing students to make social comparisons with their classmates. Therefore, students will be able to draw a social comparison between their own behavior and that of their peers. Specifically, this nudge intervention provides students with information on how many of their peers have already submitted their reflections after a lecture, which can create a social norm that motivates students to engage with the app and submit their reflections. In this experiment, this nudge is introduced twice using push notifications, same as previous intervention. I discussed the message used in both notifications with the research team to ensure their validity and reliability. Following are the description of designed notifications:

Notification 1

For the first nudge, the students were sent a push notification at the end of each lecture containing a suggestive message, i.e., "*Lecture (Number) is open to write a reflection for (class code)*". This is the same as the previously discussed Reminder 1.0.

Notification 2

After six hours of Reminder 1.0, the students were nudged with a push notification to write the reflection. The notification carried the message: *"Reminder! (Percentage number) of your peers have already submitted their reflections. Lecture (Number) is still open to write a reflection for (class code)."* Furthermore, I used 10% as a default percentage of the peers' reflection submission if the reflection submissions are less than 10%. The default number of 10 were decided by discussing it with the research team. However, I used the actual peer reflection submission percentage in the message if the submission rate was more than 10%. This way, we tried to mitigate students' feelings that reflection submission is an unimportant activity. This reminder was only sent to the students who had not submitted the reflection by six hours of Reminder 1.0. To make it easy to reference the intervention, this nudge will be referred to as Reminder 2.1 throughout the study.

Prior studies suggest that a digital nudge containing personalized information can be referred to as a smart nudge (Karlsen & Andersen, 2019). In Reminder 2.1, students were provided information about their peers' behavior, thereby making this nudge a smart nudge. In addition to the three core nudge principles, this study followed Karlsen & Andersen's (2019) proposed eight steps to design the smart nudge. Table 3 shows the steps used to design the smart nudge (Reminder 2.1) in this study.

Table 3: Steps to design and nudge using a Reminder 2.1 (smart nudge)

Step 1	Define the goal	Encouraging student to involve in the reflection activity
Step 2	Understand the users	Identify the student, either they submitted the reflection or not.
Step 3	Understand the situation	Fixed after six hours of previous nudge
Step 4	Targeting an activity	Reminder for writing today's lecture reflection
Step 5	Calculating relevant information	Identify the peer students who have submitted the reflection
Step 6	Design the nudge	Based on the percentage, we will dynamically design the nudge message
Step 7	Present the nudge	Nudge the student using the push notification
Step 8	Nudges evaluation	Observing the success of the presented nudge

Procedure

In this experiment, I divided all sections of the class into three conditions by randomly assigning students to each condition. The first condition, referred to as the baseline, did not receive any intervention. The second condition referred to as neutral reminder nudge condition, received reminders (Reminder 1.0 and Reminder 2.0) after each lecture. The third condition referred to as social comparison nudge condition, received reminders (Reminder 1.0 and Reminder 2.1) after each lecture. The reminders (Reminder 2.0 and 2.1) were only sent to students who had not submitted reflections for the corresponding lecture. To achieve the random assignment of the students, I randomly assigned each student to one of the conditions as they enrolled in any of the course sections in the application. Also, the students could opt out of receiving the push notification at any time during the study. Table 4 shows the procedure to introduce the intervention on the lecture day and the sample size in each condition.

Table 4: Procedure for intervention on the lecture day for each class

Condition	**Nudging after lecture end**	**Nudging after six hours of lecture end**	**Sample size**
Baseline	None	None	57
Neutral reminder nudge	Reminder 1.0	Reminder 2.0	67
Social comparison nudge	Reminder 1.0	Reminder 2.1	57

Analysis and results for experiment 1

For research questions 1-3, I conducted a one-way ANOVA in which the within-subject factor was conditions (baseline, neutral reminder nudge, and social comparison nudge) and the dependent variable was the number of students' reflection submissions in 24 lectures as app engagement measure. The analysis aims to determine if the students' app engagement (i.e., number of reflection submissions) differed for conditions exposed to nudge interventions (neutral reminder nudge and social comparison nudge) and no nudge intervention as a baseline. The data violated the homogeneity of variances assumption, as assessed by Levene's test of homogeneity of variances (p = 0.036). Therefore, we used Welsh's one-way ANOVA analysis, which is robust to violations of this assumption. The analysis result showed significant differences in students' reflection submissions among conditions, as indicated by Welch's F (2,115.207) = 11.019, p < 0.001. As shown in figure 3, the mean of students' reflection submissions increased in the following order: from baseline to scaffolding nudge and then to social comparison nudge condition.

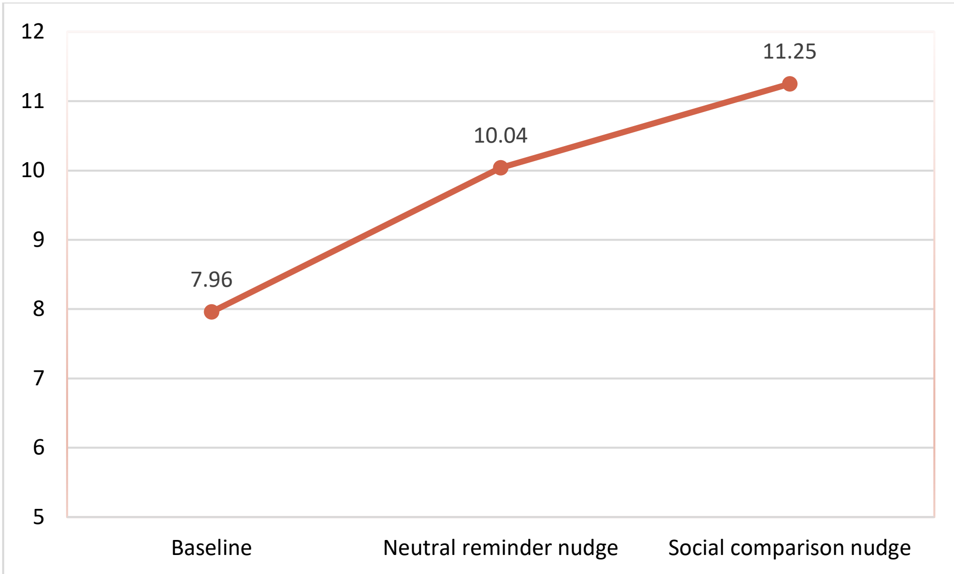

Figure 3: Mean of students' reflection submissions for each condition

Furthermore, I conducted pairwise comparisons using the Games-Howell method to see differences in the student' reflection submissions between conditions. Reflection submission of students in the social comparison and neutral reminder nudge conditions was significantly different from the baseline condition, and other comparisons were non-significant. Table 5 shows mean differences and *p* values for all comparisons.

Table 5: Result for pairwise comparison using Games Howell method

Comparisons	**Mean difference**	***p* value**
Neutral reminder nudge vs Baseline	2.07	0.006*
Social comparison nudge vs Baseline	3.28	< 0.001**
Social comparison nudge vs Neutral reminder nudge	1.20	0.30

*, ** indicate significant at $p < 0.05$ and $p < 0.001$, respectively.

To further understand the impact of nudging interventions on the students' reflection submissions, I explored the differences in students' reflection submissions both within (for research questions 1-3) and over time points (for research question 4). For the time points, I divided the lecture data into three equal time points, namely, lectures 1-8 (time point 1), 9-16 (time point 2), and 17-24 (time point 3).

Interventions impact within timepoints

Three separate one-way ANOVA were conducted for each time point to see the effectiveness of nudge interventions on the students' reflection submissions. The dependent variable in all analyses for each time point is the average of students' reflection submissions during that time. For ***time points 1*** and ***2***, the one-way ANOVA assumption of homogeneity of variance was violated as shown by Levene's test ($p < .05$). Therefore, a more robust alternative, the Welch ANOVA test, was used instead. The results of the Welch ANOVA test revealed no significant difference in the students' reflection submissions among conditions, $F(2,114.982) = 2.50$, $p = 0.086$ for the ***time point 1***, while a significant difference existed in the students' reflection submissions among conditions for the ***time point 2*** as indicated by, $F(2,116.339) = 4.72$, $p = 0.01$.

For ***time point 3***, the one-way ANOVA results revealed a significant difference in the students' reflection submissions among conditions, F (2,178) = 4.52, $p = 0.012$.

I conducted a follow-up comparison to investigate the differences between the conditions at ***time points 2** and **3***, as the students' reflections indicated significant variations between conditions during these time points. For ***time point 2***, pairwise comparisons were conducted using the Games-Howell test, and it was found that the mean of students' reflection submissions in the social comparison nudge condition was significantly higher as compared to the baseline condition. However, the mean students' reflection submissions in the neutral reminder nudge condition were similar to the other two conditions. Table 6 shows the mean differences and *p* values for all comparisons.

Table 6: Result for pairwise comparison using Games Howell method for time point 2

Comparisons	**Mean difference**	***p* value**
Neutral reminder nudge vs Baseline	0.51	0.323
Social comparison nudge vs Baseline	1.21	0.008*
Social comparison nudge vs Neutral reminder nudge	0.69	0.212

* indicate significant at $p < 0.05$.

For ***time point 3***, pairwise comparisons were conducted using the Bonferroni method and it was found that students' reflection submissions were significantly higher in the neutral reminder and social comparison nudge conditions as compared to the baseline condition. Table 7 shows the mean differences and *p*-values for all comparisons.

Table 7: Result for pairwise comparison using Bonferroni method for time point 3

Comparisons	**Mean difference**	***p* value**
Neutral reminder nudge vs Baseline	0.99	0.021*
Social comparison nudge vs Baseline	0.94	0.040*
Social comparison nudge vs Neutral reminder nudge	- 0.04	1.00

* indicate significant at $p < 0.05$.

Overall, the mean of students' reflection submissions at each time point for each condition is displayed in figure 4.

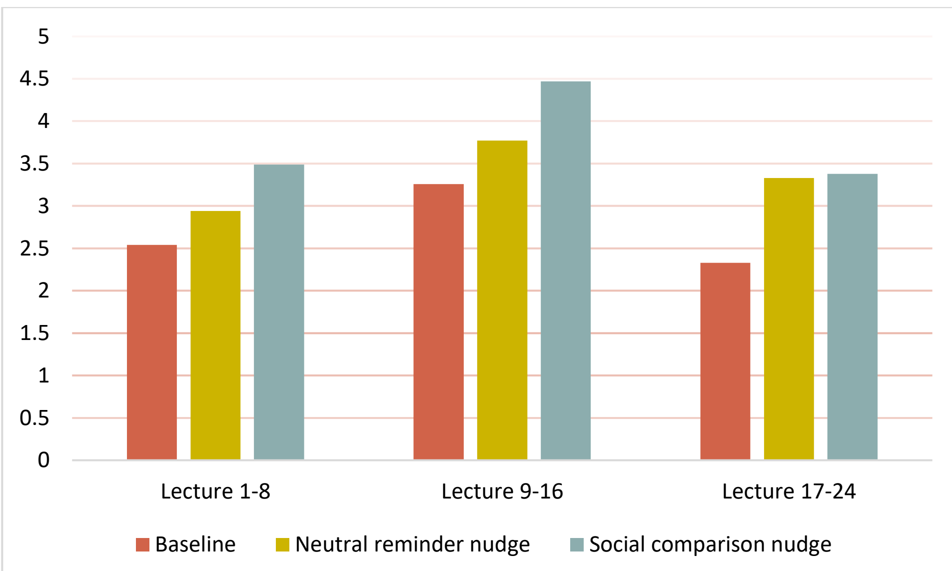

Figure 4: Mean of students' reflection submissions at each time point for conditions

Figure 4 clearly indicates that students in the social comparison nudge condition had the highest number of reflection submissions at each time point, followed by the reminder condition.

This suggests that the social comparison nudge was the most effective intervention in this experiment to improve student reflection submissions.

Interventions impact over time points

For research question 4, I conducted three separate one-way repeated measures ANOVA for each condition to see the change of students' reflection submissions over timepoints. For the ***baseline*** condition, the result revealed statistically significant differences in students' reflection submissions over three time points, F (2, 112) = 4.661, $p = 0.01$. For the students in the ***neutral reminder*** and ***social comparison nudge*** condition, the data violated the assumption of sphericity, as assessed by Mauchly's test of sphericity, $\chi 2(2) = 17.56$, $p < 0.001$, $\chi 2(2) = 14.092$, $p < 0.001$ respectively. I interpreted the result of one-way repeated measures ANOVA with Greenhouse and Geisser correction. The mean of reflection did not reveal statistically significant differences over time points in the reflection submission for the students in the ***neutral reminder nudge*** condition, indicated by F (1.617, 106.732) = 2.620, $p = 0.089$. For the ***social comparison nudge*** condition, the analysis revealed statistically significant differences in students' reflection submissions over the three time points, F (1.63, 91.35) = 4.708, $p = 0.017$.

Furthermore, I conducted the pairwise comparison with a Bonferroni adjustment for the conditions showing significant differences over time in the students' reflection submissions. For the ***baseline*** condition, the analysis revealed a statistically significant increase from first to second time point, decrease in the second to third time point, while all the other time point combinations showed a non-significant difference in the students' reflection submissions, as shown in table 8.

Table 8: Result for pairwise comparisons using Bonferroni method for baseline condition

Time points	**Mean difference**	***p* value**
2 vs 1	0.71	0.034*
3 vs 1	-0.21	1.00
3 vs 2	0.93	0.016*

* indicate significant at $p < 0.05$.

For the ***social comparison nudge*** condition, pairwise comparison with a Bonferroni adjustment revealed that there was a statistically significant increase in the reflection submissions from first to second time point, and second to third time point, but all the other time points combinations showed non-significant differences in the reflection submissions as shown in table 9.

Table 9: Result for pairwise comparisons using Bonferroni method for social comparison nudge condition

Time points	**Mean difference**	***p* value**
2 vs 1	0.98	0.018*
3 vs 1	0.21	1.00
3 vs 2	1.193	0.009*

* indicate significant at $p < 0.05$.

Overall, figure 5 shows the change of students' reflection submissions over time for each condition.

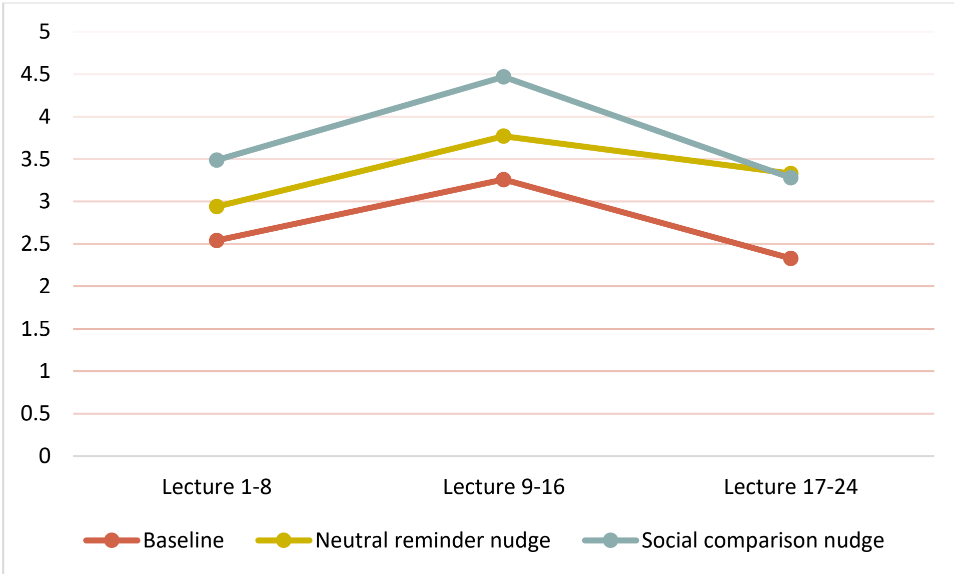

Figure 5: Mean of students' reflection submissions for conditions over time points

Figure 5 shows that the mean of students' reflection submissions in each condition varied over time. However, the students in the social comparison nudge condition at all time points submitted more reflection submissions, followed by the neutral reminder nudge condition, and then the baseline condition.

Discussion for experiment 1

The finding of this experiment showed that the social comparison and neutral reminder nudge interventions had a significant impact on the number of students' reflection submissions in the CourseMIRROR app over the semester. In addition, the pairwise comparisons showed that the students' reflection submissions in both nudge interventions conditions were significantly higher than the baseline condition. However, there was no significant difference in the mean of students' reflection submissions between the social comparison and neutral reminder nudge conditions. These findings suggest that both nudge interventions were effective in improving the students' reflection submissions. Figure 4 showed that the mean of students' reflection submissions increased in the following order: from baseline to neutral reminder nudge, and then to social comparison nudge condition. This finding provides evidence that nudging interventions can be effective at increasing students' engagement with educational apps, specifically the tasks needing a reminder. Moreover, social comparison nudges were particularly effective to engaging the students in refocusing their attention to reflection submissions in the app.

In addition to the between-condition analysis, I also explored the differences in students' reflection submissions within and across time points. I split the lecture data into three equal time points. ***Within timepoint***, the analysis revealed that there were no significant differences in reflection submissions among conditions at time point 1. However, at time point 2, students in the social comparison nudge condition submitted significantly more reflections compared to other conditions. Furthermore, at time point 3, students in both nudge intervention conditions submitted significantly more reflections compared to the baseline condition. In other words, the students at the start were submitting the reflections equally in each condition. However, as the semester progressed, nudging intervention became effective in reminding them to submit reflections. This finding is consistent with the literature that suggests nudges can have a long-lasting impact on students' behavior. For example, a study by Castleman and Page (2015) found that sending targeted text message reminders to college students significantly increased their likelihood of completing financial aid forms. Moreover, the effects of these nudges persisted for several months, even after the intervention ended.

Across time points, the mean of students' reflection submissions were significantly different across time points. Specifically, the mean of reflection submission increased from time point 1 to time point 2 and then significantly decreased further in time point 3 for students in the social comparison and baseline condition over time in a semester. However, the students' in the neutral reminder nudge condition were submitting reflections equally over time in the semester. Overall, these findings suggest that the students' engagement with the CourseMIRROR app increased from first to second time point and then decline in the later stages of the course. This trend means that the effectiveness of nudging interventions may be dependent on the engagement stage, with engagement increasing over time but plateauing in the later stages. One probable explanation is that students may undergo distinct phases of engagement throughout the duration of the course (O'Brien & Toms, 2008). The literature suggests that app engagement has different phases which includes a period of engagement, disengagement, or re-engagement (Doherty & Doherty, 2018). Therefore, it is possible that the students are in the stage of disengagement towards the end of the semester.

In this analysis, there was a trend of an increase in the first to second timepoint and then a decline in the mean of reflection submissions for students in all conditions. However, students in the social comparison nudge condition always submitted more reflections at all time points across conditions, as shown in figure 4. The reason could be that the neutral reminder nudge did not provide any additional information or feedback to the students other than reminding them and thus may not have been as effective at motivating engagement. These findings are consistent with the previous literature demonstrating the effectiveness of social comparison, and neutral reminder nudges in reminding students to complete a task (Castleman & Page, 2015, 2016, 2017; Unkovic et al., 2016). In the context of social comparison nudges, prior studies have investigated the use of social comparison nudges in various contexts, including education, health, and environmental sustainability, and have found that they can be effective in motivating individuals to change their behavior. For instance, Franklin Jr et al. (2022) conducted a study to test the impact of different digital nudges on the use of the Diagnostic Assessment and Achievement of College Skills

(DAACS) suite. The study found that a nudge providing information about the success of previous students who used DAACS previously resulted in increased engagement (i.e., completion of assessment) with the suite.

Another study found that social comparison nudges based on an interactive visualization of multiple behavioral indicators from past successful learning increased the completion rates of the Massive Open Online Course (MOOC; Davis et al., 2017). Similarly, the effectiveness is also aligned with social comparison theory (Van Lange et al., 2011), which suggests that individuals often use social comparisons to evaluate their own abilities and performance. The social comparison nudge in this study provided students with information about the number of reflections submitted by their peers, which may have motivated them to engage more with the app to keep up with or surpass their peers.

The findings of this experiment related to neutral reminder nudges are aligned with prior literature, showing the effectiveness of a neutral reminder nudge in improving students' engagement with educational apps. For example, Pham et al. (2016) found that sending reminder nudges to students increased their engagement with a language learning app. The nudges included notifications about upcoming tasks and encouraging messages to motivate students to continue using the app. Similarly, Motz et al. (2021) found that sending reminder nudges to students improved their engagement with an online Course. The nudges included information about the students' progress in the course, upcoming deadlines, and encouraging messages.

The effectiveness of reminder nudges is aligned with the Self-Determination Theory (Ryan & Deci, 2000), as it provides learners with a sense of autonomy by reminding them of the task they need to complete and giving them a choice to act on it. Reminder nudges may be perceived as providing learners with a sense of autonomy by reminding them of the task they need to complete and giving them a choice to act on it. The reminder nudge may also enhance their sense of competence by providing a cue to initiate the action and, thus, help them to achieve their goals. Additionally, the reminder nudge may help promote relatedness by reinforcing the connection between learners and the educational technology platform, making them feel more engaged with the platform and more likely to use it in the future.

Although the findings are consistent with the broad educational technology literature, there are some exceptions that are worth noting. One previous study by Brown et al. (2019) used a web-enabled coaching system to implement online personalized social comparison nudges. In their study, they alerted the students about the behavior of their peers that have outperformed them. This way, the intention was to increase the students' interaction with the application and, in turn, improve their academic performance. The study didn't find any significant difference on the students' assignment access or performance. In contrast, the present study found that the social comparison nudge was the most effective intervention in promoting app engagement. The ineffectiveness of social comparison nudges in the Brown et al. (2019) study could be attributed to the nature of the task, as the students were being graded on their attempted assignment. Conversely, in this experiment, the task was not graded and did not contribute to their final grade. Therefore, it can be argued that the nature of the task influences the effectiveness of the social comparison nudges in educational apps.

Similarly, the literature broadly discusses the effectiveness of reminder nudges (Pham & Chen, 2018; Sherr et al., 2019; Simmons et al., 2018). However, some studies have raised concerns about their ability to change behavior, as they may also serve as a source of disruption (Cutrell et al., 2001) for the user. The term "source of disruption" refers to the fact that everyday users receive many reminders in the form of push notifications (Pielot et al., 2014) from different app sources, which can lead to them ignoring the application or even deleting it altogether. For instance, Weston et al. (2015) examined the use of reminder nudges to improve patient engagement with health-based educational apps. The patients were educated on health to make informed decisions and were nudged three times to play a quiz during the game. While the initial nudge was successful in increasing user participation, subsequent nudges were not effective. Similarly, another study (Pham et al., 2016) found that overuse of reminder nudges through push notifications can have a negative impact on student engagement with the application. In this experiment, I was careful to avoid excessive use of reminder nudges. Specifically, I refrained from sending a second push notification to students who had already submitted their reflection if they have already submitted the reflection in case of both nudge interventions. This decision contributed to the effectiveness of both interventions, as they may not have caused disruption but served as reminders to submit reflections.

Overall, the current experiment showed that social comparison and neutral reminder nudges could effectively improve app engagement, in our case, it was students' reflection submissions. However, the effectiveness of these nudges may vary depending on factors such as the context and the target population. Specifically, this study focused on improving reflection submissions in the CourseMIRROR app within a classroom setting, while previous studies targeted tasks (e.g., completion of quiz, accessing the interface, or improving academic performance) in commercial apps, learning systems, or online MOOCs. The students in this experiment attended classes and observed the behavior of their peers and instructors, which could have influenced the effectiveness of the nudges. Therefore, it is possible that these nudges may not be effective for all students or in all educational settings. For example, some students may not respond well to social comparison or reminder nudge and may instead be more motivated by other types of nudges, such as feedback or goal-setting nudges. Therefore, it would be interesting to consolidate these findings by exploring the introduced nudge interventions' effectiveness in different educational settings.

Experiment 2: Supporting students' reflection summary views

Another purpose of the application is to encourage students' learning by showing peers' reflection summaries. The effectiveness of the CourseMIRROR application relies on the student viewing the reflection summaries as it enables them to gain new perspectives and insights about their learning while helping them understand their classmates' challenges (Luo et al., 2015). Therefore, this study has explored nudge interventions used in the broader nudge literature that can navigate the user toward particular content (Schneider et al., 2018) in the digital environment. The two nudge interventions that have been used for this purpose are interface nudge and reminder nudge interventions in the broader literature (e.g., Brewer & Jones, 2015; Caraban et al., 2019; Pham et al., 2016; Pham & Chen, 2018). Therefore, this study introduced these nudge interventions to explore their impact on the student reflection summary views, a behavioral aspect of app engagement. The study hypothesizes that the employed nudges would increase students' reflection summary views. More specifically, this experiment was guided by the following four research questions:

1. Do students receiving summary reminder nudges visit the reflection summary interface more often than those who do not receive nudges?
2. Do students receiving interface nudge visit the reflection summary interface more often than those who do not receive nudges?
3. What is the relative effectiveness of both nudge interventions on the students' number of visits to the reflection summary interface?

4. How do the students' reflection summary interface views change over time in each condition (i.e., summary reminder nudge, interface nudge, and baseline)?

Participants

In this experimental study, I recruited 266 students from four sections of an undergraduate first-year engineering programming course (ENGR 132) at Purdue University. All the study's participants used the CourseMIRROR application and submitted reflections throughout the semester. We focused on 24 lectures for three sections and 15 for one section. To merge the data, we used linear transformation to give weight to the data from the section with 15 lectures. I have discussed the linear transformation in the reflection summary view section of the experiment.

Reflection summary views

In this experiment, I used the students' reflection summary views in the CourseMIRROR application to measure their app engagement. The CourseMIRROR application is designed to engage students in reflection activity for a classroom. One feature of the application is to create a reflection summary after students have submitted their reflections using the application and make it available to the students (see an example of the reflection summary interface in figure 1d). Students are expected to see the reflection summary and understand their peer's difficulties and misunderstandings related to the lecture, a critical component for the effectiveness of the application on students' learning (Menekse et al., 2018). This experiment aimed to see the effectiveness of nudge interventions to encourage students' visits to the reflection summary interface for improving behavioral app engagement. Hence, the students' reflection summary views are used as an app engagement measure.

Reflection summary views provide a measure of the level of interaction between the students and the application. Therefore, if the students see the reflection summary interface frequently, it shows a higher level of app engagement. For this experiment, I implemented a feature in the application that kept track of whether the student visited the reflection summary interface at least once for a particular lecture. Reflection summary views are measured for the experiment, as the sum of student visits (yes/no) to the summary interface. For instance, if the course has 24 lectures, and the student visits the reflection summary of every lecture at least once, the summarization views will be 24.

Furthermore, I have one class section that has 15 lectures while all other class sections have 24. To combine the reflection summary views data, I performed a linear transformation to assign equal weightage to the students' reflection summary views in the section with 15 lectures as the other sections has 24 lectures, using the following formula:

$$\text{new\_value} = \frac{(\text{old\_value} - \text{old\_min}) * (\text{new\_max} - \text{new\_min})}{(\text{old\_max} - \text{old\_min}) + \text{new\_min}}$$

For instance, if a student accessed the reflection summary interface during six out of the 15 total lectures, we would set their old_min (minimum number of summary interface visits) value to 0 and old_max (maximum number of summary interface visits) value to 15. To adjust their range to 0 as the new_min value and 24 as the new_max value, we applied the linear transformation and got a new value of 9.6. This approach allowed to adjust the weightage of the reflection summary views for the classes with 15 lectures, whether considering the total reflection summary views in a semester or dividing the data into multiple time points.

Nudge interventions

In this experiment, I introduced the following two nudge interventions and explored their impact on the student reflection summary views.

Summary reminder nudge

This reminder nudge intervention helps students to refocus their attention on the availability of the lecture's reflection summary and encourages them to visit the reflection summary interface to review their peers' reflections, reinforcing their learning experience. By providing a timely and specific reminder, students are more likely to engage with the reflection summary interface, leading to increased interaction with the application and, ultimately, greater student app engagement.

In this experiment, the system nudged the students using push notifications once the reflection summary of the lecture became available in the mobile app. The reflection summary becomes available when the reflection writing time is over and more than five students' reflections are submitted for the lecture. The application nudged students using push notification with a general message, i.e., *"The reflection summary is available now for the lecture (Number) of (class code)."* The message used in the notifications was discussed with the research team to ensure their validity and reliability.

Interface nudge

In this nudge intervention, the application interface is designed to highlight the presence of the desired option and nudges the users for its selection (Brewer & Jones, 2015), usually referred to as reinforce nudge. However, I will refer this nudge as an interface nudge to make it more intuitive for the reader as it involves interface redesign. In this experiment, interface nudges were introduced by redesigning the application's current lecture screen (as shown in figure 6a) in a way that makes the availability of summarization lectures more prominent to the students. The modified lecture interface used for this experiment categorizes the lecture tabs into available lectures (contains lectures with available summaries) and remaining lectures (contains closed and upcoming lectures), as shown in figure 6b. By using the interface nudge, students are more likely to visit the reflection summary interface, resulting in increased app engagement.

a) b)

Figure 6: Default lecture interface (a) and Lecture interface nudge (b)

Procedure

In this experiment, I divided four sections of the class into three conditions and randomly assigned the students to one of three conditions: one condition received a summary reminder nudge, another condition received an interface nudge, and the third condition served as a baseline and received no nudge. For the random assignment, the students are assigned to one of the conditions upon enrolling in any section of the course. I will refer the condition by their received nudge intervention in this experiment. Table 10 shows the participant distribution in each condition.

Table 10: Distribution of students in each condition

Condition	Sample size
Baseline	97
Interface nudge	82
Summary reminder nudge	87

Analysis and results for experiment 2

For research questions 1-3, I conducted the one-way ANOVA analysis to see the impact of nudge interventions on the students' visits to the reflection summary interface of the CourseMIRROR application. I used a non-parametric alternative of One-way ANOVA, i.e., Kruskal-Wallis h test as the data didn't satisfied the normality assumption of the analysis. Distributions of students' reflection summary views were similar for all conditions, as assessed by boxplots. The analysis revealed that there is no significant difference in the students' reflection summary views among conditions, as indicated by $\chi2(2) = 1.316$, $p = .518$.

Furthermore, I was interested to see if the nudge intervention has had an impact on the student's reflection summary views within each time point (for research questions 1-3) and over time points (for research question 4). For the time points, I split the lecture data with 24 lectures into three equal time points, namely, lectures 1-8 (time point 1), 9-16 (time point 2), and 17-24 (time point 3). For the section with 15 lectures, I performed the linear transformation (using the formula explained in the reflection summarization view section) to give equal weight to students' reflection summary views in each time point by splitting the class into three time points namely, lecture 1-5 (time point 1), 6-10 (time point 2), and 11-15 (time point 3). Then, I merged both data from sections with 24 and 15 lectures.

Interventions impact within time points

To inform the study, I conducted three separate one-way ANOVA tests to see the difference in students' reflection summary views among conditions (baseline, interface nudge, and summary reminder nudge) within each time point in the semester. A non-parametric alternative, Kruskal-Wallis H test, was used as the one-way ANOVA assumption of normality was violated for each time point. Distributions of students' reflection summary views similar for all conditions at each time point, as assessed by boxplots. The analysis result revealed that students' reflection summary views were not significantly different among conditions, $\chi2(2) = 3.36$, $p = .186$, $\chi2(2) = 1.80$, $p = .407$ and $\chi2(2) = 3.12$, $p = .210$ for ***time point 1, 2, and 3***, respectively.

Interventions impact over time points

For research question 4, I used the same time points as above and conducted three separate Freidman tests for each condition to see the change of reflection summary views over time for each condition as data didn't meet the normality assumption of one-way repeated measure ANOVA. The results of the Freidman test indicated that there were statistically significant differences in students' summarization views at the three timepoints, $\chi2(2) = 49.42.039$, $p < 0.001$, $\chi2(2) = 99.10$, $p < 0.001$, and $\chi2(2) = 66.382$, $p < 0.001$ for students in the ***baseline***, ***interface nudge***, and ***summary reminder nudge*** conditions. Similarly, the post hoc analysis of all conditions shows that the reflection summary views significantly decrease over time as shown in table 11.

Table 11: Median of reflection summary views over time in each condition

	Time point 1	Time point 2	Time point 3	
Condition	Median values for reflection summary views			***p* values**
Baseline	1.00	0.00	0.00	< 0.001**
Interface nudge	1.00	0.00	0.00	<0.001**
Summary reminder nudge	1.00	0.00	0.00	<0.001**

** indicate significant at $p < 0.001$

Discussion for experiment 2

The findings of this experiment revealed that both nudge interventions, interface, and summary reminder nudge, did not significantly impact students' engagement with the reflection summary interface of the CourseMIRROR application. The analysis showed no statistically significant differences in the students' reflection summary views among conditions, indicating that the nudge interventions did not influence the students' engagement with the application. This finding contradicts our initial hypothesis that the nudge intervention would increase students' engagement with the reflection summary interface of the app.

I also explored the impact of interventions on the students' reflection summary views within and across three time points in the semester. ***Within time point***, the results indicated that there were no significant differences in reflection summary views among conditions at each time point. This analysis contradicts the finding of the experiment 1, where neutral reminder nudge was effective in increasing the students' reflection submissions. ***Across time points***, there was a statistically significant decrease in reflection summary views over time for all three conditions. This finding is consistent with previous research suggesting that students tend to lose interest in using educational apps over time (Pham & Chen, 2018).

The findings of this experiment on the interface nudge is not consistent with the existing broader literature on its' impact on students'(or general user) navigational behavior towards a particular content. Fitz-Walter et al., (2012) examined interface changes aimed at improving student engagement with an application designed to help university orientation. Their study revealed an increase in student engagement with the application. Similarly, another study (Gorissen et al., 2015) examined the impact of a of a tagging interface on student navigation of recorded lectures, finding that students who used the interface became more engaged with the videos and scored higher grades than those who used the regular interface. Also, the tagging interface helped students locate relevant parts of the lectures more efficiently. Despite the positive impact of interface nudges found in some studies, there are few studies that show non-significant results. Bowen et al. (2018) evaluated the effectiveness of interface nudges in two versions of a LibGuide. The study compared a longer, complex menu with more accessible course-related information to a shorter, simpler menu with less accessible information. The results showed that students needed equal time to reach for required course content.

In the context of the reminder nudge, most studies have shown the effectiveness of the reminder nudge in influencing the students' behavior (e.g., Dobronyi et al., 2019; Simmons et al., 2018). For instance, Pham & Chen, (2019) implemented the push notification to increase students' app engagement with the mobile application, i.e., PACARD (Personalized Adaptive CARD-based interface). Their result revealed that nudging the students through push notifications increases the students' session counts (i.e., number of app used at one time by a single user), duration of use, and app retention (i.e., continuous use of the app). However, few studies discussed how reminder nudges could be ineffective in improving students' engagement with the educational app (Cutrell et al., 2001).

There are several possible explanations for the non-significant results found in this experiment. One possible explanation is that the students did not find the nudge interventions used in the experiment compelling enough or relevant to their learning. This could have resulted in students ignoring the prompts and continuing with their usual study habits. Also, the engineering class used for this experiment has well-aligned course learning content, assessment and pedagogy and students generally have clear expectations, as a result, they generally do well in the course. This might lead them to believe that they will do fine in the class without engaging in the reflection activity implemented through CourseMIRROR app. Hence, a class that has a greater variety of performance might provide clearer results. Another possible explanation is that the reflection summary interface itself was not engaging or useful for the students, which could have led to a less frequent use than expected. For instance, I redesigned the lecture interface in such a way that the lectures with available reflection summaries were displayed separately, making them more prominent for students to approach as shown in figure 5b. The interface redesign for the nudge likely allowed students to ignore the lectures with available summaries as they may not have perceived them as useful for their learning. It is also possible that the sample size used in this study was not large enough to detect significant differences among conditions. A larger sample size might be needed to provide more statistical power and to detect smaller differences among conditions that could be meaningful.

These alternative explanations have several implications for future research. First, it is important to design nudge interventions that are relevant and compelling to students and that are tailored to their learning needs and preferences (Lembcke et al., 2019). Secondly, future research should investigate alternative ways of encouraging students to engage with reflection summaries, such as incorporating gamification elements or personalization features. Thirdly, future research should also consider using larger sample sizes to increase the statistical power of the study and to detect smaller differences among conditions. Finally, subsequent study from this experiment could be to qualitatively explore the reasons behind the nudge ineffectiveness to increase the reflection summary visits in the CourseMIRROR app. Similar approaches can be used by other studies exploring the impact of nudge interventions on the students' behavior to use multiple data sources for triangulating their findings.

Experiment 3: Scaffolding students to generate specific reflections

The CourseMIRROR application is specifically designed to facilitate the reflection writing process, and gather comprehensive reflections from students (Fan et al., 2015). Therefore, it is essential that students are able to provide relevant and in-depth reflections. This is a critical aspect of the application, as it relies heavily on the quality of reflections submitted by students. Therefore, it is imperative that students are able to write reflections thoughtfully and with careful consideration in the application. To achieve this, the experiment aims to help students in writing comprehensive reflection using the nudges, and in doing so, improve their app engagement.

In nudge literature, scaffolding and throttling mindless nudges have proven effective in providing real-time feedback to the users, which helps them to re-evaluate their thought processes and make informed decisions (Brown et al., 2019; Fritz, 2017). Hence, this experiment aimed to examine the impact of these two nudges to encourage students to write more detailed and comprehensive reflections in the CourseMIRROR application. The study hypothesized that the designed nudges would encourage students to rethink their reflection writing process and write more relevant and lengthy reflections, ultimately leading to improved cognitive and behavioral engagement with the app. This experiment is guided by the following four research questions:

1. Do students receiving a scaffolding nudge show improvement in the specificity and length of their reflections compared to those receiving no nudge?
2. Do students receiving a throttling mindless nudge show improvement in the specificity and length of their reflections compared to those receiving no nudge?
3. What is the relative effectiveness of the scaffolding and throttling nudges on the specificity and length of students' reflections?
4. How do the students' reflection specificity and text length change over time in each condition (i.e., scaffolding nudges, throttling mindless nudges, and baseline)?

Participants

In this experiment, 317 students from six sections of an undergraduate introductory first-year engineering programming course (ENGR_132) at Purdue university voluntarily participated in the study. The participating students used the CourseMIRROR app and submitted the reflection throughout the semester. For the analysis, I only included students who had submitted at least two reflections throughout the semester in the class. All participating students submitted a total of 3891 reflections. I focused on the reflection specificity of 24 lectures for each section of the course and excluded last week's lectures, as they had low reflection submissions.

Engagement measures

In this experiment, reflection specificity score and reflection text length were used as an engagement measures. The description of them are as follows:

Reflection specificity score

The application asks the students to reflect on the confusing or interesting aspects after each lecture. Students must answer four reflection questions after one another in a single reflection submission, as shown in table 12.

Table 12: Reflection questions details

Question order	**Question prompt**	**Question type**
1	Describe what was confusing or needed more details in today's class?	Open-ended question
2	How confusing? [Please pick one] 1.Slightly 2. Somewhat 3. Moderately 4. Mostly 5. Completely	Survey question
3	Describe what you found most interesting in today's class?	Open-ended question
4	How interesting? [Please pick one] 1. Slightly 2. Somewhat 3. Moderately 4. Mostly 5. Completely	Survey question

To inform the study, we utilized a natural language processing (NLP) algorithm (see detailed description in appendix A) to calculate the reflection specificity score of the students' open-ended reflections submission. This reflection specificity score provides the extent to which reflection is relevant to prompt and lecture (Butt et al., 2022). This score provides valuable insight into the student's cognitive engagement with the app, as it reflects how well they can synthesize their learning experience in the lecture and reflect on it. For instance, when students receive a high reflection specificity score, it indicates that their reflection is directly relevant to the lecture content and reflection prompts. This, in turn, suggests that the student is highly engaged with the application. Therefore, the reflection specificity score can be used as a cognitive app measure for the analysis.

My colleagues and I evaluated the agreement of the score evaluated by the NLP algorithm and human coders when evaluating the specificity of the reflection data. The result revealed a high level of agreement using Cohen's kappa as indicated by κ (NLP) = 0.775 and κ (NLP) = 0.773 with human coding for each set of reflection data discussing the interesting and confusing lecture's aspect (Butt et al., 2022), respectively. This high agreement indicates that the reflection specificity algorithm can be used to measure specificity score of the students' reflection. Hence, for this experiment, I used average reflection specificity scores of students' reflections evaluated by the NLP algorithm as a cognitive app engagement measure.

Reflection text length

Throughout the semester, after each lecture, students are prompted to provide textual responses to two open-ended questions, as shown in table 12. The study used the length (i.e., number of words) of the students' written reflections to measure app engagement. The reflection text length provides valuable insights into how frequently students are interacting with the app (in our case, reflection writing activity) and their willingness to take the time to reflect on what they have learned.

Lengthier students' reflections suggest that they are engaged with the app and are taking the time to reflect on their learning. The effectiveness of CourseMIRROR apps depends on students' readiness to put in the time and effort necessary to reflect in-depth on their learning experience. On the other hand, students writing shorter or shallow detailed reflections may show that the app is not holding the student's focus or that they are not completely engaged. Therefore, reflection text length can be used as an app engagement measure to effectively understand the students' meaningful interaction in the CourseMIRROR app. By analyzing the length of students' written reflections, the study can better assess the impact of the designed nudges on students' cognitive and behavioral app engagement.

Nudge interventions

This study explored the impact of the following two nudge interventions to improve the reflection specificity and text length of the students.

Throttling mindless nudge

In this nudge intervention, the users are given a pause to re-evaluate their decision (Wang et al. 2014). This study used this nudge intervention by providing students with a pause using an alert dialogue box, containing a scaffolding message in the application. Also, there are two options in the alert dialogue box to choose from: "Proceed", and "Ok! I will revise it" (see example in the figure 7). The students can select one option to move forward. If they choose "proceed" option, they will move to the next question. On the other hand, if they choose the other option ("Ok! I will revise it"), the students will stay on the same question interface and revise the reflection before moving on to the next question. This pause and the scaffolding message are intended to help students re-evaluate their response to the open-ended reflection question before proceeding to the next question. This can help students to think more deeply and critically about the question and the lecture content, which can improve the quality and length of their responses.

In addition, the scaffolding messages used in the alert dialogue box are designed to provide students with specific guidance to improve their reflection specificity score (see an example of nudge in figure 7). By providing tailored guidance, the scaffolding message can help students to write thoughtful, and detailed reflections, which can be seen in the text length and specificity of their reflection. Therefore, when a student responds to reflection question and proceeds to the next question in the reflection activity, the application evaluates the reflection specificity score using an NLP algorithm through an API. The API evaluates the students' textual reflections and scores them with a value ranging from 1 to 4 (4 being the high relevance and 1 being the shallow relevance to the lecture and reflection prompt). If the reflection quality score is less than or equal to 2, then the application gives an alert with a scaffolding message. The scaffolding message is decided based on the reflection specificity and question being asked, as shown in table 13.

Table 13: The messages used for the throttling mindless nudge

Reflection questions	**Reflection quality**	
	1	2
Describe what was confusing or needed more details in today's class?	Please tell us what you found confusing or unclear in today's class.	Your feedback is valuable, please provide additional details.
Describe what you found most interesting in today's class?	Please tell us what you found interesting or important in today's class"	Your feedback is valuable, please provide additional details.

Figure 7 shows the throttling mindless nudge in the CourseMIRROR application.

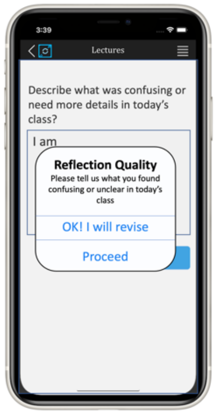

Figure 7: Alert for throttling mindless nudge

Scaffolding nudge

In this nudge intervention, users are provided with constant feedback to help them make informed decisions about their behavior (Fancsali et al., 2021). In this experiment, students are nudged by providing real-time feedback using the scaffolding messages, and color bar during their reflection writing process in the app, as shown in figure 7. By providing feedback during the reflection writing process, students can receive immediate guidance on improving their responses, reducing the cognitive load of having to self-assess their writing. Thus, guiding them to improve the reflection specificity and provide an in-depth reflection.

In addition, the constant feedback is based on the reflection specificity score by an NLP algorithm through API. This real-time scaffolding messages intends to help students write relevant, more thoughtful, and detailed reflection, which can be reflected in the text length and reflection specificity score. The API using NLP algorithm provides a specificity score ranging from 4 (highly relevant to the lecture and reflection prompt) and 1 (shallow relevance to the lecture and reflection prompt). The scaffolding messages and color bar changes in a real time changes based on these reflection questions and specificity score, as shown in table 14.

Table 14: Rubric to select color code and scaffolding message

Reflection specificity	Color code	Reflection question: Describe what was confusing or needed more details in today's class?	Reflection question: Describe what you found most interesting in today's class?
1	Red	Please tell us what you found confusing or unclear in today's class.	Please tell us what you found interesting or important in today's class
2	Yellow	Your feedback is valuable, please provide additional details.	
3	Blue	Sounds good, can you please tell us why it is confusing?	Sounds good, can you please tell us why it is interesting?
4	Green	Great, thanks!	

Figure 8 (a-d) shows the possible combination of scaffolding message and color coding for the confusing question, asking students to reflect on the confusing aspects of the lecture:

a) b) c) d)

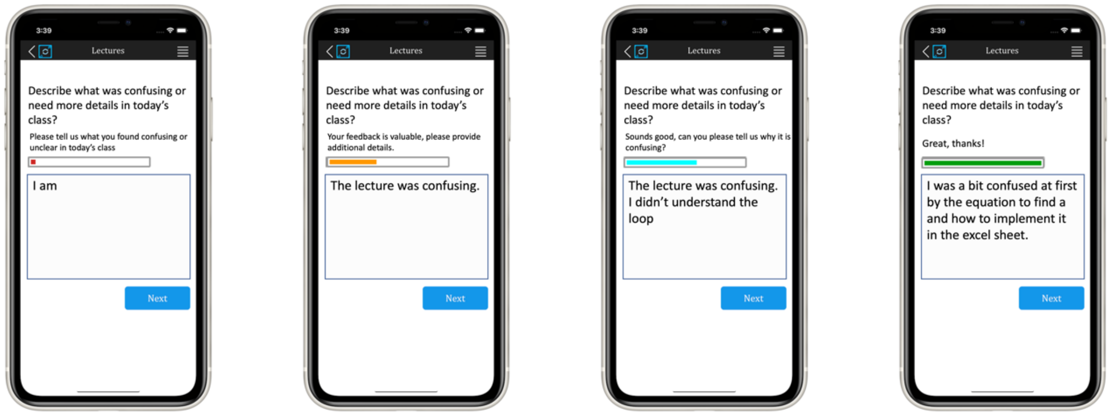

Figure 8: Example of scaffolding nudge while students reflect on the confusing aspect of the lecture

Procedure

In this experiment, six sections of the FYE class were divided into three conditions. The first condition, referred to as the baseline, received no intervention. The second condition, known as scaffolding nudge condition, received the scaffolding nudge. The third condition, throttling mindless nudge, received the throttling mindless nudge. I will refer the condition by their intervention name. Students were randomly assigned to one of the conditions as they enrolled in any of the course sections in the application. To inform our analysis, we selected 24 lectures in a semester for each course section. Table 15 shows the distribution of students in each experimental condition.

Table 15: Distribution of students in each condition

Condition	Sample size
Baseline	105
Scaffolding nudge	106
Throttling mindless nudge	106

Analysis and results for experiment 3

In the analysis, the effectiveness of the nudges was separately evaluated using both engagement measures: reflection specificity score and text length. For each engagement measure, the impact of the nudges was separately evaluated for students' reflections on the confusing and interesting aspects of the lectures. For convenience, I will refer to the students' reflections describing the interesting and confusing aspects of the lecture as interesting and confusing questions, respectively. The means and standard deviation for students' reflection specificity and text length (i.e., number of words) for both question types are presented in table 16.

Table 16: Means and standard deviation of engagement measures for each question type

	Interesting question		**Confusing question**	
Engagement measures	Mean	Standard deviation	Mean	Standard deviation
Reflection specificity score	3.04	0.61	2.76	0.71
Reflection text length	15.43	9.73	16.04	10.74

Reflection specificity score as a dependent variable

To inform the research questions, different analyses were conducted with the dependent variable being the reflection specificity score for the student's reflection to the interesting and confusing questions in the application.

For research questions 1-3, I conducted a one-way ANOVA to assess the impact of the nudge interventions on students' specificity scores for their reflections on both interesting and confusing questions in each condition. The within-subject factor for the analysis was the nudge conditions (as shown in table 15), and the dependent variable was the students' average reflection specificity score for the respective questions across 24 lectures. Since the data violated the homogeneity of variances assumption, as indicated by a significant Levene test ($p < 0.05$), I conducted a one-way Welch ANOVA to determine whether the student's reflection specificity score differed among the baseline, scaffolding nudge, and throttling mindless nudge conditions for both question types. The results revealed that students' reflection specificity scores for the interesting question were significantly different among conditions, with Welsh's $F (2, 207.245) = 9.235$, $p < 0.001$. However, for the confusing question, the students' reflection specificity scores were not significantly different among conditions, as indicated by Welsh's $F (2, 206.715) = 2.094$, $p = 0.126$.

The mean of students' reflections specificity score for the interesting question in both the scaffolding and throttling mindless nudge conditions was more than the baseline condition, as shown in figure 9.

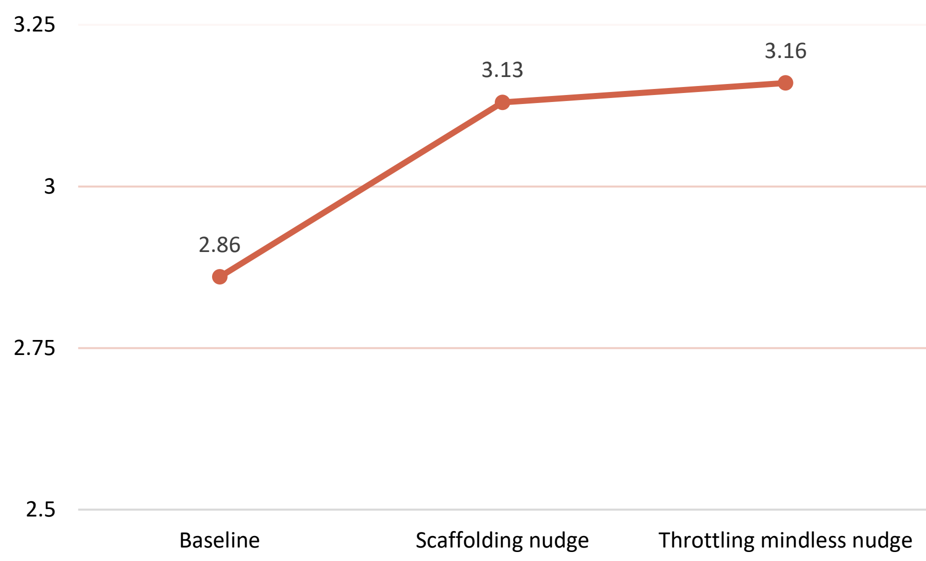

Figure 9: Mean of reflection specificity score for each condition

Additionally, I conducted pairwise comparisons using the Games-Howell test for the student's reflection specificity score of the interesting question. The post hoc analysis revealed that students' reflection specificity scores increased significantly from the baseline to the scaffolding nudge condition and from the baseline to the throttling mindless nudge condition. However, all other comparison combinations were not significant. Table 17 shows the mean differences and *p* values of all comparisons using the reflection specificity score of the interesting question.

Table 17: Interesting question - Result for pairwise comparison using Games Howell method

Comparisons	**Mean difference**	***p* value**
Scaffolding nudge vs Baseline	0.27	0.004*
Throttling mindless nudge vs Baseline	0.29	< 0.001**
Throttling mindless nudge vs Scaffolding nudge	0.02	0.952

*, ** indicate significant at $p < 0.05$ and $p < 0.001$, respectively.

Furthermore, I am interested to see the effectiveness of the nudging intervention on specificity scores of students' reflections on the interesting and confusing questions, both within (for research questions 1-3) and across each time point (research question 4). To inform this analysis, I divided the lecture data into three equal time points, namely, lectures 1-8 (time point 1), 9-16 (time point 2), and 17-24 (time point 3), to assess the effectiveness of the nudging interventions on students' reflection specificity scores. At each time point, I took the average specificity score of students' reflection submissions as the dependent variable. Students who did not submit reflections during two-time points were excluded from this analysis. However, multiple imputations were used to estimate the reflection specificity scores of students who did not submit any reflections at any one time point. This was done by taking the average of their average reflection specificity scores for the other two time points. Table 18 shows the descriptive statistics of each condition after the imputations.

Table 18: Distribution of students in each condition

Condition	**Sample size**
Baseline	97
Scaffolding nudge	103
Throttling mindless nudge	104

Interventions impact within each time point

For students' reflections on the interesting question, I conducted three separate one-way ANOVA to evaluate the impact of nudge interventions on the students' reflection specificity score in each condition within each time point. For ***time point 1*** and ***3***, I used Welsh one-way ANOVA as the data violated the homogeneity of variance assumption, as indicated by a significant Levene's test ($p < 0.05$). The results of ANOVAs revealed that students' reflection specificity score was statistically significant for each condition as indicated by Welch's $F(2,200.012) = 5.427$, $p = .005$, and Welch's $F(2, 199.613) = 6.191$, $p = .002$, for ***time point 1*** and ***3*** respectively. For ***time point 2***, the result of ANOVAs revealed that students' reflection specificity score was statistically significant for each condition as indicated by $F(2,301) = 7.884$, $p < 0.001$.

Furthermore, pairwise comparisons were conducted to see the differences in students' reflection specificity scores among conditions at each time point. For ***time point 1***, I conducted pairwise comparisons using Games-Howell. The analysis result revealed that the increase in the students' reflection specificity score was significantly higher from baseline to the scaffolding nudge condition and from baseline to the throttling mindless nudge condition. Table 19 shows the result of the Games-Howell post hoc analysis.

Table 19: Interesting question - Result for pairwise comparisons using Games Howell method for time point 1

Comparisons	**Mean difference**	***p* value**
Scaffolding nudge vs Baseline	0.25	0.03*
Throttling mindless nudge vs Baseline	0.28	0.009*
Throttling mindless nudge vs Scaffolding nudge	0.02	0.972

* indicate significant at p < 0.05.

For ***time point 2***, I conducted pairwise comparison using the Bonferroni post hoc method which revealed that the increase in the mean of students' reflection specificity score from baseline to the scaffolding nudge condition was statistically significant as well as from baseline to the throttling mindless nudge condition. Table 20 shows the result of the pairwise comparison using the Bonferroni method.

Table 20: Interesting question - Result for pairwise comparison using Bonferroni method for time point 2

Comparisons	**Mean difference**	***p* value**
Scaffolding nudge vs Baseline	0.34	0.002**
Throttling mindless nudge vs Baseline	0.33	0.002**
Throttling mindless nudge vs Scaffolding nudge	-0.01	1.00

** indicate significant at p < 0.001.

For ***time point 3***, I conducted Games-Howell post hoc analysis for the pairwise comparison, which revealed that the increase in the students' reflection specificity score from the baseline to the scaffolding nudge condition was statistically significant and from baseline to the throttling mindless nudge condition. Table 21 shows the result of the pairwise comparison using Games-Howell for ***time point 3***.

Table 21: Interesting question - Result for pairwise comparison using Games Howell method for time point 3

Comparisons	**Mean difference**	***p* value**
Scaffolding nudge vs Baseline	0.26	0.035*
Throttling mindless nudge vs Baseline	0.32	0.003**
Throttling mindless nudge vs Scaffolding nudge	0.05	0.867

*, ** indicate significant at p < 0.05 and p < 0.001, respectively.

Overall, figure 10 shows the decreasing trend of specificity score in the students' reflection to the interesting question within each time point for each condition.

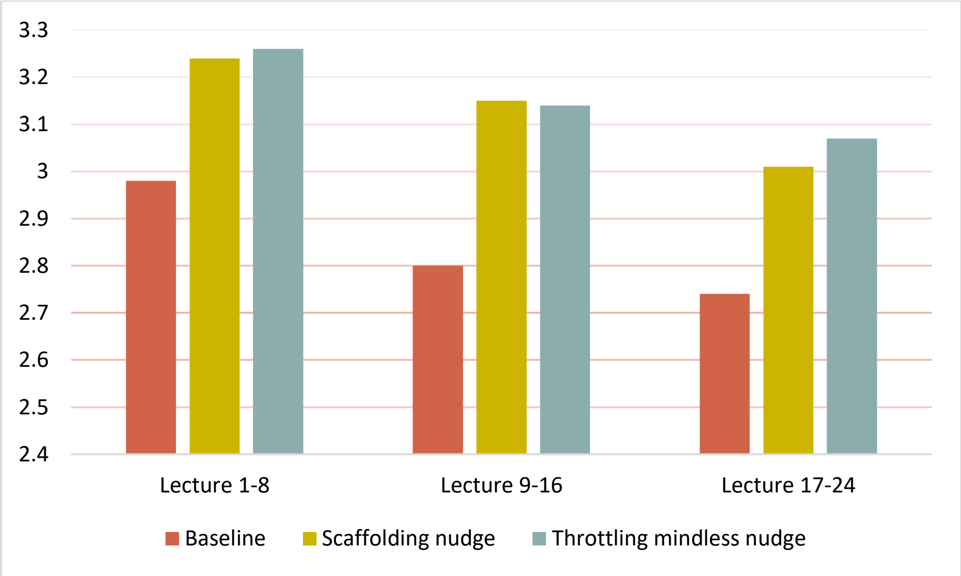

Figure 10: Interesting question - Mean of students' reflection specificity score within each time point for each condition

For reflections on the confusing question, I conducted three separate one-way ANOVA to evaluate the impact of nudge interventions on the students' reflection specificity score in each condition within each time point. For all time points, I used the Welsh one-way ANOVA as the data violated the homogeneity of variance assumption, indicated by a significant Levene's test (p < 0.05). The analysis revealed that the students' reflection specificity was not significantly different among conditions at ***time point 1*** for Welsh's F (2, 200.18) = 0.35, p = 0.705 while it was significantly different among conditions for ***time point 2*** and ***3*** as indicated by Welsh's F (2,199.710) = 3.88, p = 0.02, and Welch's F (2, 199.366) = 3.277, p = .040 respectively.

Pairwise comparisons were conducted for ***time points 2*** and ***3*** using students' reflection specificity as it showed the variation in students' reflection specificity scores between conditions. For ***time point 2***, I conducted Games-Howell post hoc analysis for the pairwise comparison, which revealed that the increase in the students' reflection specificity score from the baseline to the scaffolding nudge condition was statistically significant. Table 22 shows the result of the pairwise comparison using Games-Howell for ***time point 2***.

Table 22: Confusing question - Result for pairwise comparison using Games Howell method for time point 2

Comparisons	**Mean difference**	***p* value**
Scaffolding nudge vs Baseline	0.31	0.021*
Throttling mindless nudge vs Baseline	0.19	0.183
Throttling mindless nudge vs Scaffolding nudge	0.12	0.59

* indicates significant at p < 0.05.

For ***time point 3***, I conducted Games-Howell post hoc analysis for the pairwise comparisons which revealed that the increase in the students' reflection specificity score from the baseline to the throttling mindless nudge condition was statistically significant. Table 23 shows the result of the pairwise comparison using Games-Howell for ***time point 3***:

Table 23: Confusing question - Result for pairwise comparison using Games Howell method for time point 3

Comparisons	**Mean difference**	***p* value**
Scaffolding nudge vs Baseline	0.21	0.18
Throttling mindless nudge vs Baseline	0.27	0.04*
Throttling mindless nudge vs Scaffolding nudge	0.05	1.0

* indicates significant at $p < 0.05$.

Overall, figure 11 shows the decreasing trend of reflection specificity score within each time point for each condition.

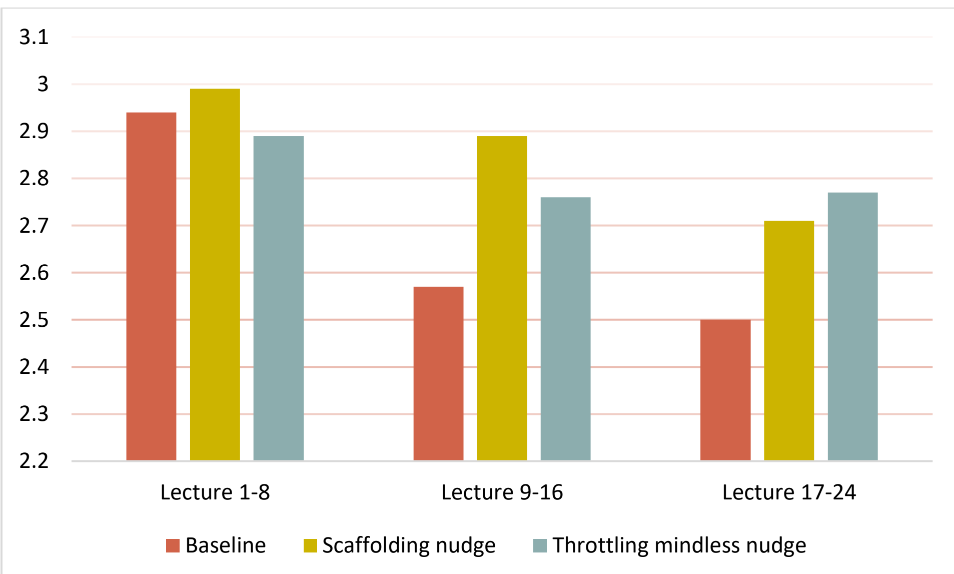

Figure 11: Confusing questions - Mean of students' reflection specificity score within each time point for each condition

Intervention impact over time Points

For reflection on the interesting question, I conducted three separate one-way repeated measures ANOVA tests for each condition to see if there was any change in reflection specificity score over three time points. All three ANOVAs revealed that there were significant differences in the reflection specificity score over time in each condition, indicated by the analysis results $F(2,192) = 7.74$, $p < 0.001$(baseline), $F(2,204) = 5.6$ $p = 0.004$ (scaffolding nudge), and $F(2,206) = 6.09$ $p = 0.003$ (throttling mindless nudge). Follow-up tests were conducted for each condition to evaluate the difference among the means between each time point. For the ***baseline*** condition, the reflection specificity scores significantly decreased from time point 1 to 2 and from time point 1 to 3, but not from time point 2 to 3. For the ***scaffolding nudge*** condition, the reflection specificity scores significantly decreased from time point 1 to 3, but not from time point 1 to 2, or from time point 2 to 3. For the ***throttling mindless nudge*** condition, the reflection specificity scores significantly decreased from time point 1 to 3 and not from time point 1 to 2, or not from time point 2 to 3. Table 24 shows the pairwise comparison of time points for each condition with mean differences, and *p* values:

Table 24: Interesting question - Pairwise comparison of time point using Bonferroni method for each condition.

Time point	**Baseline**		**Scaffolding nudge**		**Throttling mindless nudge**	
	Mean difference	*p* value	Mean difference	*p* value	Mean difference	*p* value
2 vs 1	0.177	0.013*	0.09	0.447	0.127	0.05
3 vs 1	0.235	0.002*	0.22	0.006*	0.194	0.005*
2 vs 3	0.058	0.972	0.134	0.16	0.067	0.73

* indicate significant at p < 0.05.

Overall, figure 12 shows the change in students' reflection specificity scores over time for each condition.

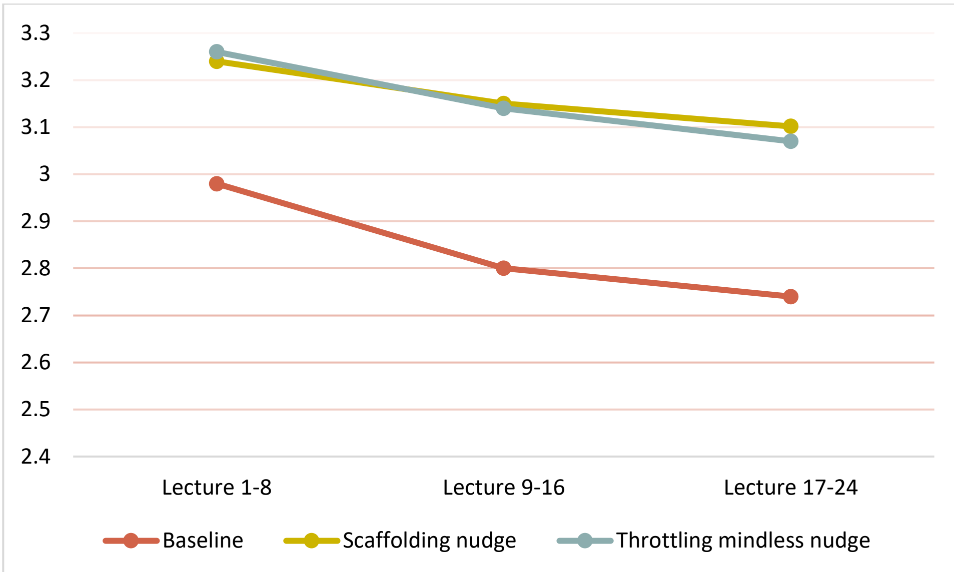

Figure 12: Interesting question - Mean of students' reflection specificity score over time for each condition

For reflection on the confusing question, I conducted three separate one-way repeated measures ANOVA tests for each condition to see if there were differences in students' reflection specificity score over three timepoints. For both ***baseline*** and ***scaffolding nudge*** condition, the ANOVAs result showed that there were significant differences in the reflection specificity score over time, indicated by the analysis results, F (2,192) = 5.479, p < .001, and F (1.87,190.986) = 4.980, p= 0.009 (with Greenhouse-Geiser correction). For the students in the ***throttling mindless nudge*** condition, the analysis revealed no significant differences in the reflection specificity score over time, F (2,206) = 5.44 p= 0.162. Additionally, I conducted a pairwise comparison with Bonferroni adjustment among time points to see the change in the reflection specificity score for the baseline and scaffolding nudge condition. For the ***baseline*** condition, the reflection specificity score significantly decreased from time point 1 to 2, and from time point 1 to 3, but not from time point 2 to 3. For the ***scaffolding nudge*** condition, the reflection specificity score was statistically significantly decreased from time point 1 to 3, but not from time point 1 to 2, or from time point 2 to 3. Table 25 shows the pairwise comparison of time points for ***baseline*** and ***scaffolding nudge condition*** with mean differences, and *p* values:

Table 25: Confusing question - Pairwise comparison of time point using Bonferroni method for each condition

Time points	**Baseline**		**Scaffolding nudge**	
	Mean difference	*p* value	Mean difference	*p* value
2 vs 1	0.37	0.013*	0.01	0.31
3 vs 1	0.44	0.002*	0.28	0.019*
2 vs 3	0.072	0.895	0.17	0.11

* indicate significant at $p < 0.05$.

Overall, Figure 13 shows the mean reflection specificity score over time for each condition.

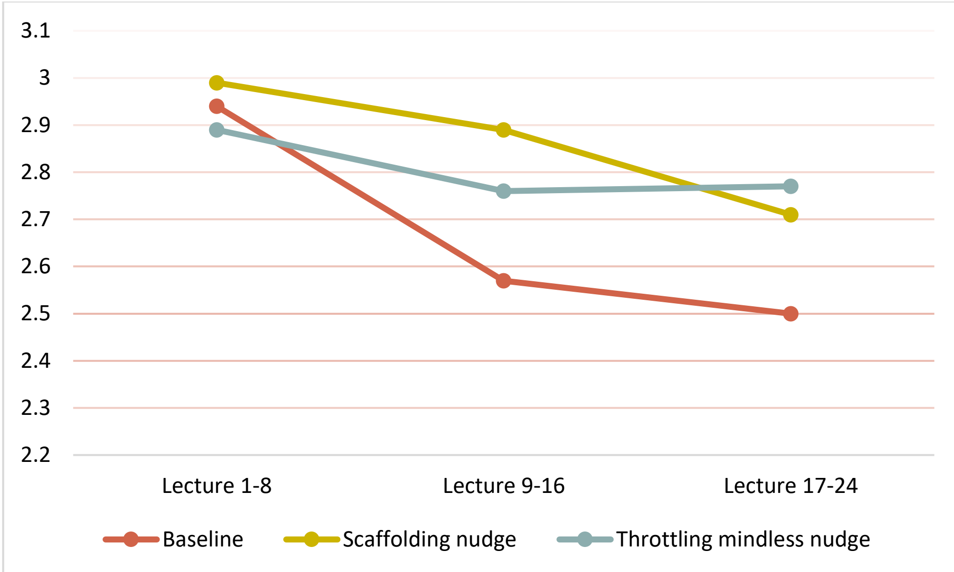

Figure 13: Confusing question - Mean of students' reflection specificity score over time for each condition

Reflection text length as a dependent variable

To inform research questions (1-3), different analyses were conducted to evaluate the impact of scaffolding and throttling mindless nudges on the students' reflection text length for the interesting and confusing aspect of the lecture for each condition. As the data violated the normality assumption, I conducted a non-parametric Kruskal-Wallis H test to inform the study. In the analysis, the within-subject factor was the nudge conditions (as shown in table 18), and the dependent variable was text length of students' reflections on both interesting and confusing reflection questions. The distribution of the reflection text length was not similar for all conditions, as assessed by visual inspection of a boxplot for both questions. The result revealed that the text length of reflections to interesting and confusing questions were significantly different among conditions, H (2) = 35.115, $p < 0.001$, and H (2) = 49.85, $p < 0.001$, respectively.

Furthermore, I conducted the pairwise comparison using Dunn's method for text length of students' reflections on both questions. *For the reflections on the interesting question*, the post hoc analysis revealed significant differences ($p < 0.001$) in students' reflection text length in the ***baseline*** condition (mean rank = 116.09) as compared to the ***scaffolding nudge*** condition (mean rank = 175.01) and ***throttling mindless nudge*** condition (mean rank = 185.50). Table 26 shows the result of the pairwise comparison using Dunn's method.

Table 26: Interesting question - Result for pairwise comparison of conditions using Dunn's method

Comparisons	**Mean rank difference**	***p* value**
Scaffolding nudge vs Baseline	58.94	< 0.001**
Throttling mindless nudge vs Baseline	69.41	< 0.001**
Throttling mindless nudge vs Scaffolding nudge	10.49	1.00

** indicate significant p < 0.001.

For reflection on the confusing question, post hoc analysis revealed statistically significant differences ($p < .001$) in students' reflection text length in the ***baseline*** condition (mean rank = 107.82) as compared to the ***scaffolding nudge*** (mean rank = 178.37) and ***throttling mindless nudge*** (mean rank = 190.32) conditions. Table 27 shows the result of the pairwise comparison using Dunn's method.

Table 27: Confusing question - Result for pairwise comparison using Dunn's method

Comparisons	**Mean rank difference**	***p* value**
Scaffolding nudge vs Baseline	70.54	< 0.001**
Throttling mindless nudge vs Baseline	82.49	< 0.001**
Throttling mindless nudge vs Scaffolding nudge	11.942	1.00

** indicate significant at p < 0.001.

To further understand the impact of nudging interventions in scaffolding students to write a lengthy reflection, I examined the differences in text length of the student's reflections on interesting and confusing questions, both within (for research questions 1-3) and across each time points (for research question 4). For the time points, I divided the lecture data into three equal time points, namely, lectures 1-8 (time point 1), 9-16 (time point 2), and 17-24 (time point 3). I excluded students who did not submit any reflection within any two time points. However, I included students whose average reflection text length for only one time point was missing. I used multiple imputations to estimate the average reflection text length of students who did not submit any reflections during any one time point by taking the average of their reflection text length in the other two time points. Table 28 shows the participants in each condition after multiple imputations.

Table 28: Distribution of participants in each condition

Condition	Sample size
Baseline	97
Scaffolding nudge	103
Throttling mindless nudge	104

Interventions impact within each time point

For reflection on the interesting question, I conducted three separate one-way ANOVAs to evaluate the impact of nudge interventions on the students' reflection text length for each condition within three time points. In this regard, I used Kruskal-Wallis H test for each time point separately, as the data did not meet the normality assumptions for one-way ANOVA. The analysis revealed that the students' reflection text length varied significantly among conditions at each time point, as indicated by $H(2) = 28.18$ ($p < 0.001$), $H(2) = 29.40$ ($p < 0.001$), and $H(2) = 31.34$ ($p < 0.001$) for ***time point 1, 2*** and ***3*** respectively.

Additionally, I performed a pairwise comparisons using Dunn's method for each time point. For ***time point 1***, there were significant differences in students' reflection text length between the scaffolding nudge (mean rank = 167.62) and baseline (mean rank = 113.58) conditions, and baseline and throttling mindless nudge (mean rank = 172.83) conditions, but not between the throttling mindless nudge and scaffolding nudge conditions. Table 29 shows the mean rank differences and *p*-value of all comparisons:

Table 29: Interesting question - Result for pairwise comparison using Dunn's method for time point 1

Comparisons	Mean rank difference	*p* value
Scaffolding nudge vs Baseline	54.03	0.00**
Throttling mindless nudge vs Baseline	60.2	0.00**
Throttling mindless nudge vs Scaffolding nudge	6.21	1.0

** indicate significant at p < 0.001.

For ***time point 2***, students' reflection text length is significantly different between the scaffolding nudge (mean rank = 164.43) and baseline (mean rank = 113.32) conditions, and baseline and throttling mindless nudge (mean rank = 177.23) conditions, but not between the throttling mindless nudge and scaffolding nudge conditions. Table 30 showed the mean rank difference and *p* value of all comparisons

Table 30: Interesting question - Result for pairwise comparison using Bonferroni method for time point 2

Comparisons	Mean rank difference	*p* value
Scaffolding nudge vs Baseline	51.10	0.00**
Throttling mindless nudge vs Baseline	63.91	0.00**
Throttling mindless nudge vs Scaffolding nudge	12.80	0.884

** indicate significant p < 0.001.

For ***time point 3***, students' reflection text length was significantly different between baseline (mean rank = 112.05) and both scaffolding nudge (mean rank = 164.82) and throttling mindless nudge (mean rank = 178.03) conditions. All the other comparisons were non-significant. Table 31 showed the mean rank difference and *p* value of all comparisons:

Table 31: Interesting question - Result for pairwise comparison using Bonferroni method for time point 3

Comparisons	**Mean rank difference**	***p* value**
Scaffolding nudge vs Baseline	52.77	0.00**
Throttling mindless nudge vs Baseline	65.98	0.00**
Throttling mindless nudge vs Scaffolding nudge	13.20	0.839

** indicate significant p < 0.001.

Overall, the figure 14 showed the students in the throttling mindless nudge, and scaffolding conditions wrote lengthier reflection as compared to the baseline condition, as evident by a mean rank score of students' reflection text length with-in each time point for conditions.

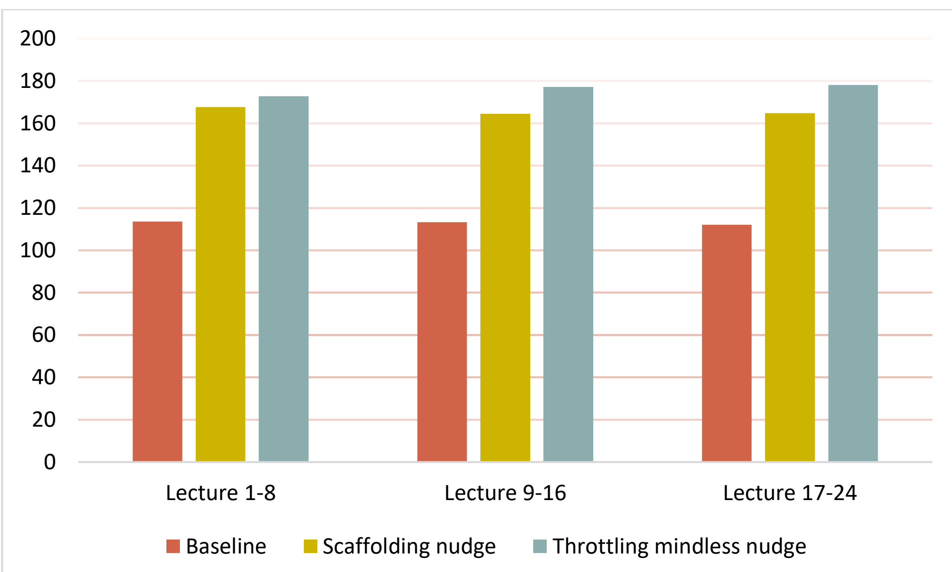

Figure 14: Interesting question - Mean rank score of students' reflection text length within each time points for each condition

For students' reflection on the confusing question, I conducted three separate one-way ANOVAs to evaluate the impact of nudge interventions on the students' reflection text length in each condition within each time point. In this regard, I used the Kruskal-Wallis H test for each time point separately as the data did not meet the assumptions for one-way ANOVA. The analysis revealed that the length of students' reflection text varied significantly among conditions, as indicated by H(2) = 29.69 ($p < 0.001$), H(2) = 39.01 ($p < 0.001$), and H(2) = 50.967 ($p < 0.001$) for ***time point 1, 2*** and ***3***, respectively.

Additionally, I performed a pairwise comparison using Dunn's method for each time point. For ***time point 1***, there are statistically significant differences in students' reflection text length between the scaffolding nudge (mean rank = 166.64) and baseline (mean rank = 107.28) conditions, and baseline and throttling mindless nudge (mean rank = 180.67) conditions, but not between the throttling mindless nudge and scaffolding nudge conditions. Table 32 showed the mean rank difference and *p* value of all comparisons.

Table 32: Confusing question - Result for pairwise comparison using Dunn's method for time point 1

Comparisons	**Mean rank difference**	***p* value**
Scaffolding nudge vs Baseline	59.35	0.00**
Throttling mindless nudge vs Baseline	73.38	0.00**
throttling mindless nudge vs Scaffolding nudge	14.02	1.0

** indicate significant at p < 0.001.

For ***time point 2***, students' reflection text length was found to be significantly different between the scaffolding nudge (mean rank = 167.13) and baseline (mean rank = 101.15) conditions, and baseline and throttling mindless nudge (mean rank = 185.90) conditions, but not between the throttling mindless nudge and scaffolding nudge conditions. Table 33 showed the mean rank difference and *p*-value of all comparisons.

Table 33: Confusing question - Result for pairwise comparison using Bonferroni method for time point 2

Comparisons	**Mean rank difference**	***p* value**
Scaffolding nudge vs Baseline	65.97	0.00**
Throttling mindless nudge vs Baseline	84.74	0.00**
Throttling mindless nudge vs Scaffolding nudge	18.76	0.753

** indicate significant at p < 0.001.

For ***time point 3***, students' reflection text was significantly different between baseline condition (mean rank = 112.05) and both scaffolding nudge (mean rank = 164.82) and throttling mindless nudge (mean rank = 178.03) conditions. All the other comparisons were non-significant. Table 34 showed the mean rank difference and *p* value of all comparison.

Table 34: Confusing question - Result for pairwise comparison using Bonferroni method for time point 3

Comparisons	**Mean rank difference**	***p* value**
Scaffolding nudge vs Baseline	52.77	0.00**
Throttling mindless nudge vs Baseline	65.98	0.00**
Throttling mindless nudge vs Scaffolding nudge	13.20	0.374

** indicate significant at p < 0.001.

Overall, figure 15 shows that the students in the throttling mindless nudge and scaffolding nudge conditions wrote lengthier reflection compared to the baseline condition, as indicated by the mean rank score of students' reflection text length with-in each time point for conditions.

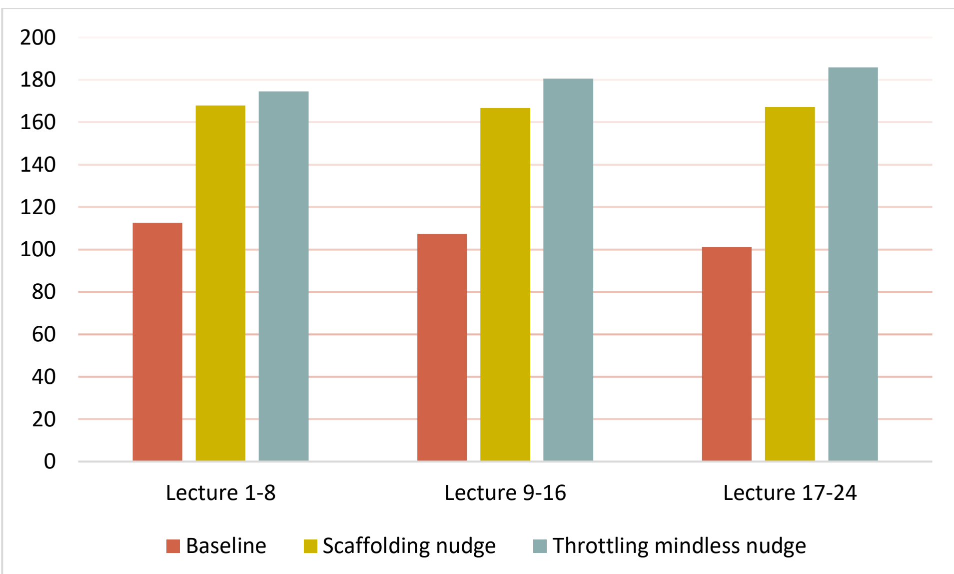

Figure 15: Confusing question - Mean rank score of students' reflection text length within each timepoints for each condition

Interventions impact over time points

For students' reflections on the interesting question, I explored if there were any variations in the students' reflection text length over three time points for each condition. For the ***baseline*** condition, I used Friedman test as the data violated the normality assumption of the One-way Repeated Measure ANOVA. The analysis result showed that there were significant differences in the reflection text length score over time, as indicated by $\chi2(2) = 10.90$, $p < 0.001$. Pairwise comparisons with a Bonferroni correction were also carried out for the ***baseline*** condition, which revealed that the reflection text length was significantly decreased ($p = 0.004$) from the time point 1 (median = 2.20) to 3 (median = 1.74), but no other comparison was significant.

For the ***scaffolding nudge*** condition, I used the one-way repeated measure ANOVA with Greenhouse-Geisser correction as the data didn't meet the sphericity assumption. The result revealed the significant differences in the reflection specificity score over time: $F(1.87,190.837) = 3.67$, $p = 0.030$. Post hoc analysis with a Bonferroni adjustment revealed that the reflection text length significantly decreased from time point 1 to 2 but not between time point 1 and 3, or between time point 2 and 3.

For the ***throttling mindless nudge*** condition, I used Friedman test as the data violated the normality assumption of the One-way Repeated Measure ANOVA. The analysis results showed that there were no significant differences in the reflection text length score, as indicated by $\chi2(2) = 1.65$, $p = 0.438$. Overall, figure 16 shows the mean of reflection text length score over time for each condition.

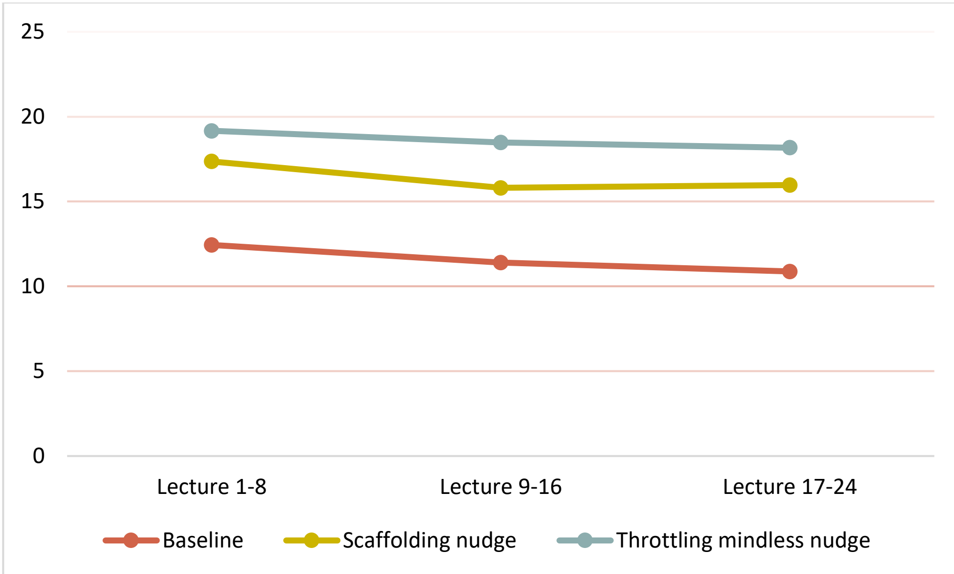

Figure 16: Interesting question - Mean of students' reflection text length over time for each condition

For students' reflections on the confusing question, I explored if there were any variations in the reflection text length across three different time points for each condition. Three Friedman tests for each condition were conducted as the data violated the normality assumption of the One-way Repeated Measure ANOVA. The analysis showed that there were significant differences in the reflection text length, as indicated by the results $\chi 2(2) = 37.97$, $p < 0.001$, $\chi 2(2) = 12.72$, $p < 0.001$, and $\chi 2(2) = 9.05$, $p < 0.001$ for ***baseline, scaffolding nudge***, and ***throttling mindless nudge*** conditions. Follow-up tests were conducted for all conditions in the experiment. For the ***baseline*** condition, pairwise comparisons with a Bonferroni correction were also carried out, which revealed that the reflection specificity score was significantly different from the time point 1 (median = 10.50) to the 2 (median = 7.25) ($p < 0.001$), and 3 (median = 6.0) ($p < 0.001$) but no other comparison was significant.

For the ***scaffolding nudge*** condition, pairwise comparisons with a Bonferroni correction were also carried out, which revealed that the reflection text length was significantly decrease from the time point 1 (median = 18.66) to the time point 2 (median = 15.00) (p = .02), and 3 (mean rank = 16.00) (p = .002) but no other comparison was significant. For the ***throttling mindless nudge*** condition, pairwise comparisons with a Bonferroni correction were also carried out, which revealed that the reflection text length was significantly decrease from the time point 1 (median = 18.75) to the time point 2 (mean rank = 17.91) (p = .023), and the time point 3 (median = 16.00) (p = .034) but no other comparison was significant. Figure 17 shows the mean reflection text length score over time for each condition.

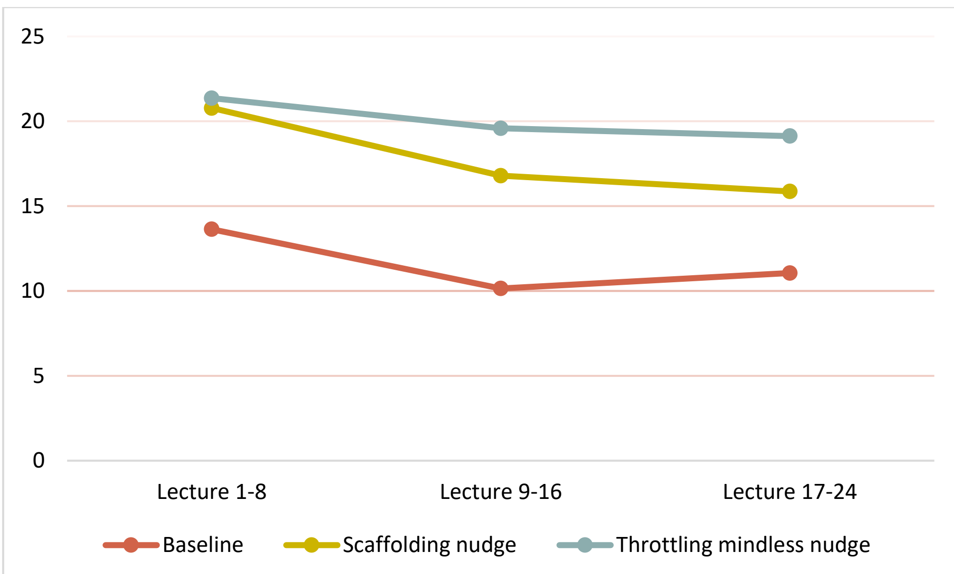

Figure 17: Confusing question - Mean of students' reflection text length over time for each condition

Discussion for experiment 3

The results of the experiment were mixed regarding the impact of nudging interventions on students' reflection specificity scores and text length in response to questions, asking about the interesting and confusing aspect of the lectures in the CourseMIRROR application.

The first analysis examined the impact of scaffolding and throttling mindless nudges on students' reflection specificity scores and text length of their response to interesting and confusing questions over semester. The results showed that students in the scaffolding and throttling mindless nudge conditions wrote more relevant reflections for the interesting question compared to the baseline condition. The students in both nudging intervention conditions wrote lengthy reflections for both the confusing and interesting questions compared to the baseline condition. On the whole, these nudges were successful in enhancing students' reflection specificity scores and text length which is consistent with the previous studies on the effect of nudges on the students' engagement in the reflection writing (Mohammadhassan et al., 2020; Nelson et al., 2012).

Furthermore, the result showed that the impact of nudging interventions could vary among conditions within and across three time points in the semester. ***Within time points***, the results indicated that, at time point 1, students in the scaffolding and throttling mindless nudge conditions had significantly higher reflection specificity scores compared to the baseline condition for the interesting question, but no significant differences were found for the confusing question. In case of reflection text length, students in the scaffolding nudge condition wrote significantly more for both questions at the time point 1. At time point 2, students in both nudging intervention conditions had significantly higher reflection specificity scores for the interesting question compared to the baseline condition. Moreover, the students in the scaffolding nudge condition had significantly higher scores for the confusing question compared to the baseline condition. Once again, the scaffolding nudge condition wrote significantly more reflection text for both question types at time point 2. Finally, at time point 3, students in both intervention conditions had significantly higher reflection specificity scores compared to the baseline condition for the interesting question. In addition, students in the throttling mindless nudge condition had significantly higher reflection specificity scores as compared to the baseline condition for the confusing question. Students in the scaffolding and throttling mindless nudge conditions wrote significantly lengthier reflections for both questions at time point 3. Overall, both interventions were effective at different time points in promoting students' app engagement by improving their reflection text length and the specificity of reflection questions.

The effectiveness of scaffolding and throttling mindless nudges on app engagement at different time points may be due to several factors. A possible explanation is that both nudge interventions were providing some form of support and guidance to the students during the reflection writing process either through real-time feedback or through a throttling mindless activity by enabling a pause through an alert dialogue box. This finding is consistent with the existing literature on the impact of nudging (Castleman & Page, 2015) and the first experiment's findings. Another possible explanation is that the constant feedback through nudges may have acted as a motivator, as they could see that they were improving by looking at the color bar or seeing no alert dialogue over time. This confidence in achievement might have reinforced their self-efficacy and encouraged them to remain engaged with the application, aligned with previous literature. Previous studies have also discussed the relationship of constant feedback with students' motivation in the classroom activity (Burgers et al., 2015), which in our case is reflection activity using the CourseMIRROR app.

Across time points, there was a significant decrease in students' reflection specificity scores of the ***baseline*** condition from time point 1 to 2 and 3 for both reflection questions. However, students in the ***baseline*** condition wrote significantly less reflection text from time points 1 to 3 for the interesting question and from time points 1 to 2 and then 3 for the confusing question. In the case of the ***scaffolding nudge*** condition, the students had a significant decrease in reflection specificity scores from time point 1 to 3 for both interesting and confusing questions and they wrote significantly less reflection text from time point 1 to 2 for the interesting question and from time point 1 to 2 to 3 for the confusing question. In the case of the ***throttling mindless nudge*** condition, there was a significant increase in reflection specificity scores from time point 1 to 2 and 3 for the interesting question. At the same time, there was no significant difference over time points for the confusing question. However, the ***throttling mindless nudge*** condition showed no significant difference in the reflection text length for the interesting question. However, there was a significant decrease from time point 1 to 2 to 3 for the confusing question. Overall, the study suggests that the students in both nudge intervention either decrease or stay consistent over time points.

The varied effects of the scaffolding and throttling mindless nudge interventions over time points are intriguing. In all conditions, students wrote highly relevant and lengthier reflections in the first time point. However, all students reflection specificity and text length reduced significantly as time progressed in the semester. In other words, nudging interventions were not able to constantly influence the students' behavior over the semester. One possible explanation could be the setting of the class where students start working in teams towards the end of the semester. Therefore, they might not feel the need for reflection activity and thus reduce the engagement with the application. Another explanation could be that the difficulty level of the class increases over time and they start spending more time on other graded classroom activities. Therefore, it is important to consider these factors when exploring the effectiveness of nudging interventions over time. Additionally, longitudinal studies can help identify the specific factors that influence the effectiveness of nudging interventions and help to develop more effective nudges that can sustain their impact over time.

The experiment's findings indicate that scaffolding and throttling mindless nudges can effectively promote reflection and improve the quality of student responses to reflection questions, particularly in enhancing reflection text length. These findings are consistent with the literature on nudging interventions, which suggests that small prompts or cues can have a significant impact on behavior (Thaler and Sunstein, 2008) and have a positive impact on students' (or general user) app engagement (e.g., Basu et al., 2015; Fahid et al., 2021; Fancsali et al., 2018; Zydney & Warner, 2016).

Prior studies have explored the scaffolding nudge's impact on the students' app engagement by improving their reflective or general writing. For instance, Looi et al. (2011) conducted a study using the MyDesk app, which includes scaffolding students with KWL tool while they reflection on their learning progress. The study revealed that students were able to reflect deeply, indicating a high level of cognitive app engagement. Another study by Hwang et al. (2012) investigated the efficacy of the Ubiquitous Scientific Device Trainer (USDT) app, which guided students on operating scientific devices within a museum. Students who used the app outperformed those who received instructor demonstrations, indicating their cognitive app engagement and better application of their knowledge. The authors suggest that providing users with guidance and resources can help to reduce the cognitive load associated with using a new app, making it more likely that users will continue to engage with it and have better learning experience.

While studies have found the scaffolding nudge effective in improving cognitive app engagement, some studies have shown either no effect or mixed results. For instance, Wingate (2010) used the scaffolding nudge by presenting the content feedback to improve the student's writing. While the nudge improved some students writing, there was no impact on many other students’ writings. They further found that the scaffolding nudge lessened their motivation, resulting in disengagement with the feedback while writing. Similarly, another study by Mitchell et al. (2019) showed that the scaffolding nudge didn’t improve the students’ writing quality, specifically students with low self-efficacy and high anxiety levels. A similar phenomenon is seen in the study (Wambsganss et al., 2022) where the students’ self-efficacy was the same, but the scaffolding nudge didn’t improve students’ reflection writing. Therefore, it would be interesting to see the impact of the scaffolding nudge by moderating the students’ motivational constructs (e.g., self-efficacy, or self-regulation).

Throttling mindless nudges has not yet been widely used in the context of educational apps, as indicated by the literature survey. However, the broader technology literature has frequently applied this technique to improve user engagement (i.e., make informed decision) with various types of apps (e.g., Bergram et al., 2020; Caraban et al., 2020; Grüning et al., 2023). In a study by Caraban et al. (2019), different effective nudge interventions were explored in the context of human-computer interaction (HCI). The study found that throttling mindless nudges can be effective in changing user behavior, as it provides a time buffer to allow users to reverse their uninformed actions. Similarly, Wang et al. (2014) investigated the effectiveness of throttling mindless nudges to improve the quality of Facebook posts by the user. They introduced mindless nudge by introducing a 10-second delay using a Chrome plugin before the publication of Facebook posts. The findings revealed that participants were more likely to revise or abandon their posts during the delay because of the nudge, resulting in more thoughtful and deliberate behavior.

Overall, the current experiment showed that both nudge interventions were effective to scaffold students during their reflection writing and improved their cognitive and behavioral app engagement. However, more research is needed to explore the long-term effectiveness of these interventions over time. Specifically, it would be interesting to see whether the positive effects of scaffolding and throttling mindless nudge diminish once students have received enough training. Furthermore, it is important to consider the duration and frequency of the interventions, as well as the specific scaffolding prompts used, to optimize the effectiveness of both nudging interventions. Further research could also examine the implementation of nudging interventions in diverse educational contexts and investigate potential factors that moderate their effectiveness. By addressing these gaps in the literature, we can better understand how to optimize the use of nudging techniques to promote learning and engagement in educational applications.

SUMMARY AND DISCUSSION FOR ALL THREE EXPERIMENTS

The findings of these experiments provide insights into the effectiveness of nudging interventions for improving the students' engagement with the primary features (i.e., purposes) of the CourseMIRROR app. Specifically, this study has found that the social comparison nudge and neutral reminder nudge are effective for improving the students' reflection submissions. Additionally, students receiving a social comparison nudge were more likely to submit their reflections in comparison to those who either just received a neutral reminder nudge or no nudge. However, the summary reminder and interface nudges were found to be ineffective in improving students' visits to the reflection summary interface, suggesting that these nudges may not be the best approach for improving students' engagement with the application.

Finally, the scaffolding and throttling mindless nudges were found to be effective in improving the specificity and text length of students' reflections, indicating that these nudges were successful in promoting students' deeper engagement during the reflection writing process in the app. However, the effectiveness of these nudges varied within and across different time points during a semester, suggesting that a nuanced approach is needed to optimize their effectiveness.

These findings are consistent with previous research on digital nudges in education. For example, a study by Franklin Jr et al. (2022) found that social comparison nudge was effective in improve the students engagement with an assessment tool suite (DAACS; Diagnostic Assessment and Achievement of College Skills). Similarly, another study (Sherr et al., 2019) found that reminder nudges are an effective tool to bring about the desired behavioral changes and improve the students' learning.

The findings regarding the summary reminder and interface nudges are surprising, as previous research suggests that these types of nudges can be effective in promoting behavioral change (Bowen et al., 2018; Dobronyi et al., 2019; Simmons et al., 2018). For example, a study by Gorissen et al. (2015) found interface nudge to be effective for encouraging students to engage more deeply with online course content. It is possible that the lack of effectiveness observed in this study is due to the specific design of the nudges or the context in which they were used.

The findings regarding the scaffolding and throttling mindless nudges are consistent with previous research on their use to promote students' (or user) cognitive and behavioral engagement in the digital environment. For example, a study by Wambsganss et al. (2022) found that scaffolding can be effective in improving the students' argumentative writing and thus improving their cognitive engagement. Similarly, a study by van Oldenbeek et al., (2019) found that the email-based feedback nudge increased the students' views and duration of new learning videos. Similarly, the effectiveness of throttling mindless nudges for influencing users' behavior in the digital environment has been discussed in the broader technology literature but scarcely in the educational setting (Damgaard & Nielsen, 2018). For instance, a study by Grüning et al. (2023) found throttling mindless nudges to be effective in reducing the usage of certain apps by the users. Similarly, Caraban et al. (2019) identified this nudge as one of the 23 effective nudges used in HCI literature to influence the user's decision making.

The findings of this study also highlight the importance of considering the specific context in which digital nudges are implemented. While I found reminder nudges to be effective in improving students' reflection submissions, they were not effective in refocusing the students' attention to visit the reflection summary interface. This suggests that different nudges may not be effective to bring about behavioral change in all contexts. Therefore, there is a need for studies to explore the type of nudges that can be effective in improving certain behaviors among students.

Furthermore, the varying results of the employed nudges within and across different timepoints during a semester highlight the need for further research to determine the most effective implementation strategies for digital nudges. This has been emphasized in the previous studies as well. For instance, Kizilcec et al. (2014) explored the impact of reminder nudges in online STEM courses and found varied impacts of nudging on short or long term student participation in a course. Similarly, a review by Damgaard & Nielsen (2018) on nudging in education also emphasized the need for studies to understand the sustainable impact of nudging on students' behavior.

Overall, the findings of this study showed that nudging interventions were effective in improving the students' app engagement in the CourseMIRROR application. This study also contributes to the growing body of research investigating the effectiveness of digital nudges to improve the students' app engagement. These findings will support the future research in designing engaging educational apps that could provide a better learning experience to students. Furthermore, the understanding of digital nudging can help to design cost-effective interventions to facilitate students' ability to regulate their study habits, enhancing their motivation, improving their focus, and providing personalized learning experience (Damgaard & Nielsen, 2018; Edwards & Li, 2020).

CONCLUSION & IMPLICATIONS

This study aimed to investigate the impact of nudging interventions on app engagement in a first-year engineering class using the CourseMIRROR application. Specifically, the study focused on designing and testing nudging interventions to facilitate the students' reflection submissions, supporting their visit to reflections summary interface, and scaffolding the students to write in-depth and comprehensive reflections. Three experiments were conducted to achieve these objectives. In the first experiment, I explored whether a social comparison nudge and a neutral reminder could improve students' reflection submissions. The results of the study indicate that both the neutral reminder and social comparison nudges had a significant impact on increasing students' reflection submissions compared to the baseline condition over the semester. However, a closer examination of the results within the three equally divided time points revealed that initially, there was no significant difference in reflection submission among the conditions. As time progressed, the students in the social comparison nudge condition submitted more reflection, and in the final time point, the reflection submission of students in both nudge intervention conditions was significantly higher than the baseline condition. Interestingly, students in the neutral reminder condition submitted reflection equally over time points, indicating that they may have relied on the reminder to submit their reflection rather than their intrinsic motivation. Overall, these findings highlight the effectiveness of nudges in increasing student engagement in CourseMIRROR application and suggest that social comparison nudges may be particularly effective in promoting sustained behavior change over time to submit reflections.

The second experiment investigated the impact of a summary reminder nudge and an interface nudge to encourage students to view reflection summaries interface in the app. The results of the study indicated that both nudge interventions were unable to increase the students' reflection summary views compared to the baseline condition over the semester. Also, I explored the impact of nudging interventions within three equally divided time points over the semester. The result revealed that there was no significant difference in students' reflection summary views among the conditions. Interestingly, when compared over there times, students in all conditions showed a decline in the reflection summary views, indicating that the employed nudge intervention could not refocus the students' attention toward reflection summary interface. Therefore, further research is needed to identify factors that can effectively increase students' engagement with reflection summary interface.

Finally, the third experiment examined the impact of scaffolding and throttling mindless nudges to scaffold students in writing specific reflections and encourage longer reflection texts. The results revealed that both interventions significantly impacted students' reflection specificity scores for the interesting question, leading to longer reflection text lengths than the baseline condition. However, the impact of nudging interventions varied across three time points, indicating that their effectiveness may depend on the type of reflection questions, or the scaffolding clues used to scaffold the students. Both scaffolding and throttling mindless nudges can effectively promote reflection and improve the quality of student responses to reflection questions.

Furthermore, the study contributed to the educational technology literature in several ways. First, this study provided empirical evidence on the impact of nudge interventions on students' app engagement. Second, this study is one of the few studies to the best of my knowledge in the context of students' app engagement that conceptualized the digital interventions as a nudge and provided a theoretical framework to guide the research process of the experiments. Previously, studies used digital nudge to enhance the students' app engagement without even realizing it (e.g., Pham et al., 2016) and even those which conceptualized them as nudge didn't ground their work in the literature (e.g., Wambsganss et al., 2022). Therefore, this study can serve as foundation for the future researchers working with nudge interventions to ensure that their study is well grounded in the established principles and concepts. Moreover, they can use this study to build their understanding in identifying digital nudges and students' cognitive biases to influence the students' behavior in the similar context. Third, previous studies have only investigated the impact of one or two nudge interventions on the students' behavior (e.g., Castleman et al., 2014; Mitrovic et al. 2019). However, this study employed three difference experiments, each introduced two commonly used nudge interventions aligned with the study's goal and explored their impact on the students' app engagement. Therefore, this study can provide a good repertoire of nudges that can be effective in improve students' engagement with the app. For instance, our finding showed that reminder with social comparison information can be effective to remind the students about a particular task because of their tendency to adhere to the social norm. Hence, this study provides valuable insights to future researchers and app designers, exploring different cost-effective digital nudges to influence the student's behavior.

Finally, this study contributed to the design-based research on educational app where I implemented a digital nudge to improve the students writing (in our case reflection wiring) using the technology. In this regard, I discussed the integration of a digital nudge (based on the NLP algorithms) in the CourseMIRROR app, and its' implication to improve the students' metacognitive skills (i.e., reflection writing). Hence this study provided a contribution to the ongoing effort that focuses on the integration of similar advance technology enhancement in the educational apps for improving other metacognitive skills such as comment writing (Mohammadhassan et al., 2022), problem solving (Winkler, et al., 2021), and empathy (Wambsganss et al., 2022). Hence, the research can further explore NLP techniques (e.g., sentiment analysis, data mining) and use them as a nudge to students and keep them engaged with the application.

In conclusion, the study provided valuable insights into the impact of nudging interventions on student engagement with the CourseMIRROR application. The findings indicate that nudges can effectively increase students' app engagement, especially used for involving students in the reflection activity. Specifically, social comparison nudges were found to be effective in promoting sustained behavior change over time, while scaffolding and throttling mindless nudges were effective in improving reflection specificity scores and encouraging longer reflection texts. The study also emphasizes the careful design and testing of nudges intervention to ensure their effectiveness and the significance of ongoing evaluation and refinement of interventions to maintain their efficacy over time.

LIMITATIONS AND FUTURE DIRECTIONS

The current study has several limitations that should be considered while understanding the impact of the result and their interpretation. Also, these limitations allow future research to address these gaps and comprehensively understand the impact of nudge interventions on students' app engagement. One limitation is that the study was designed for a first-year engineering course in a single institute. This limits the generalizability of the findings to other courses, institutions, or even educational apps. Future research could address this limitation by replicating the study in multiple courses or institutions and using various educational apps. Also, this study relied on quantitative approaches to inform the study, limiting the findings' depth and richness. Therefore, it would be beneficial if future studies could be designed using the mixed method approach that triangulates quantitative data (using data analytics) with qualitative data (using student interviews). This approach would provide a comprehensive understanding of how well nudge interventions alter student behavior while shedding light on why some interventions might be less successful than others. Also, this method can provide a thorough understanding of the underlying mechanism influencing the nudge intervention effectiveness. For example, a follow-up study could be conducted after experiment 2 by interviewing the students and trying to understand their perception of the employed nudge and why the nudge intervention did not result in improving students' visits to reflection summary interface. This study would provide a more thorough understanding of the factors affecting student app engagement and how interventions could be tailored to improve reflection summary views in the CourseMIRROR app.

Additionally, this study investigated the impact of nudge interventions at three different time points. However, it would be beneficial to investigate the nudge intervention in parallel with the activities taking place in the class. For example, classroom observations could be conducted to determine if the teacher emphasized using the application by consistently discussing their reflection summaries and answering students' questions about their reflections. These observations can help us identify if the instructor's frequent use of the app's findings in their pedagogy or emphasizing the students to use the application moderate nudging interventions' impact on the students' app engagement, as students might see the relevance and importance of the application to their learning experience. Moreover, the study only examined nudging interventions in short-term effects over a semester. It is unclear if these effects will be sustained over a longer period or may fade away after some time. Future research could address this limitation by conducting a longitudinal study to examine nudging interventions' long-term impact on students' behavior. Finally, the study did not account for potential student behavior variation due to their demographic profile (e.g., gender, socioeconomic status), motivation, or self-efficacy levels. Future research could address this limitation by examining the impact of nudging interventions controlling for these variables.

In conclusion, the study provides valuable insights into the effects of nudging interventions on students' engagement with the CourseMIRROR app. Still, several limitations need to be addressed in future research to enhance the validity and generalizability of the findings. In conclusion, the study provides valuable insights into the effects of nudging interventions on students' engagement with the CourseMIRROR app. Still, several limitations need to be addressed in future research to enhance the validity and generalizability of the findings.

REFERENCES

30 Best Examples of Nudge Marketing in eCommerce (2022, December 06). Convertcart. Retrieved March 12, 2023, from https://www.convertcart.com/blog/nudge-marketing-examples

Agapie, E., Golovchinsky, G., & Qvarfordt, P. (2013). Leading people to longer queries. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3019–3022. https://doi.org/10.1145/2470654.2481418

Al-Ameen, M. N., Wright, M., & Scielzo, S. (2015). Towards making random passwords memorable: Leveraging users' cognitive ability through multiple cues. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2315–2324. https://doi.org/10.1145/2702123.2702241

Alameri, J., Masadeh, R., Hamadallah, E., Bani, H., & Fakhouri, H. (2020). *Students' perceptions of e-learning platforms (Moodle, Microsoft Teams and Zoom platforms) in the University of Jordan education and its relation to self-study and academic achievement during COVID-19 pandemic*. *11*(5), 21–33.

Aljohani, N. R., & Davis, H. C. (2013). Learning analytics and formative assessment to provide immediate detailed feedback using a student centered mobile dashboard. *Proceedings of the 2013 Seventh International Conference on Next Generation Mobile Apps, Services and Technologies*, 262–267. https://doi.org/10.1109/NGMAST.2013.54

Anwar, S., Butt, A. A., & Menekse, M. (2022). Relationship between course engagement and educational application engagement in the context of first-year engineering students. *129th Annual Conference & Exposition, American Society of Engineering Education.* Minneapolis, Minnesota.

Applying behavioural insights to reduce fraud, error and debt. (2012). Cabinet Office Behavioural Insights Team. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/60539/BIT_FraudErrorDebt_accessible.pdf

Ariely, D. (2009). *Predictably irrational: The hidden forces that shape our decisions*. Harper Perennial.

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 267–270. https://doi.org/10.1145/2330601.2330666

Ashok, S. M. (2011, February 22). *Abu Dhabi University implements mobile education solution*. Tahawultech. https://www.tahawultech.com/features/abu-dhabi-university-implements-mobile-education-solution/

Barev, T. J. (2020). Theory-motivated design for developing privacy nudges. *15th International Conference on Wirtschaftsinformatik, Association for Information System.* Potsdam, Germany. https://www.alexandria.unisg.ch/259875/1/JML_768.pdf

Bartel, A., & Hagel, G. (2014). Engaging students with a mobile game-based learning system in university education. *2014 IEEE Global Engineering Education Conference (EDUCON)*, 957–960. https://doi.org/10.1109/EDUCON.2014.6826215

Barton, A., & Grüne-Yanoff, T. (2015). From libertarian paternalism to nudging–and beyond. *Review of Philosophy and Psychology*, *6*(3), 341–359. https://doi.org/10.1007/s13164-015-0268-x

Basu, S., Sengupta, P., & Biswas, G. (2015). A Scaffolding framework to support learning of emergent phenomena using multi-agent-based simulation environments. *Research in Science Education*, *45*(2), 293–324.

Ben-Joseph, K. (2021, April 21). *2021 Engagement Benchmarks for Education Apps*. Alchemer. https://www.alchemer.com/resources/blog/2021-engagement-benchmarks-for-education-apps/

Bergram, K., Bezençon, V., Maingot, P., Gjerlufsen, T., & Holzer, A. (2020). Digital Nudges for Privacy Awareness: From consent to informed consent? *Proceedings of the 28th European Conference on Information Systems (ECIS), Association for Information System*. Marrakesh, Morocco.

Bergram, K., Djokovic, M., Bezençon, V., & Holzer, A. (2022). The digital landscape of nudging: A systematic literature review of empirical research on digital nudges. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3517638

Bettinger, E. P., & Baker, R. B. (2014). The Effects of student coaching: An Evaluation of a randomized experiment in student advising. *Educational Evaluation and Policy Analysis*, *36*(1), 3–19. https://doi.org/10.3102/0162373713500523

Bilyalova, A. A., Salimova, D. A., & Zelenina, T. I. (2020). Digital transformation in education. In T. Antipova (Ed.), *Integrated Science in Digital Age* (pp. 265–276). Springer International Publishing.

Bouta, H., & Retalis, S. (2013). Enhancing primary school children collaborative learning experiences in maths via a 3D virtual environment. *Education and Information Technologies*, *18*(4), 571–596. https://doi.org/10.1007/s10639-012-9198-8

Bowen, A., Ellis, J., & Chaparro, B. (2018). Long nav or short nav?: Student responses to two different navigational interface designs in libguides version 2. *The Journal of Academic Librarianship*, *44*(3), 391–403. https://doi.org/10.1016/j.acalib.2018.03.002

Brewer, R. N., & Jones, J. (2015). Pinteresce: Exploring reminiscence as an incentive to digital reciprocity for older adults. *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work and Social Computing*, 243–246. https://doi.org/10.1145/2685553.2699017

Brown, M. G., Schiltz, J., Derry, H., & Holman, C. (2019). Implementing online personalized social comparison nudges in a web-enabled coaching system. *The Internet and Higher Education, 43,* 100691. https://doi.org/10.1016/j.iheduc.2019.100691

Brown, S., White, S., & Power, N. (2016). Cluster analysis of assessment in anatomy and physiology for health science undergraduates. *International Journal of Teaching and Learning in Higher Education*, *28*, 102–109.

Bruehlman-Senecal, E., Hook, C. J., Pfeifer, J. H., FitzGerald, C., Davis, B., Delucchi, K. L., Haritatos, J., & Ramo, D. E. (2020). Smartphone app to address loneliness among college students: Pilot randomized controlled trial. *JMIR Mental Health, 7*(10), e21496. https://doi.org/10.2196/21496

Bustillo, J., Rivera, C., Guzmán, J. G., & Acosta, L. R. (2017). Benefits of using a mobile application in learning a foreign language. *Sistemas & Telemática*, *15*(40), 55–68. https://doi.org/10.18046/syt.v15i40.2391

Butt, A. A., Anwar, S., & Menekse, M. (2022). Work in progress: Investigating the relationship

between FYE students' reflections and academic performance across gender.*13th Annual First-Year Engineering Experience Conference, American Society of Engineering Education.* East Lansing, Michigan.

Cai, S., & Xu, Y. (2008). Designing product lists for e-commerce: The effects of sorting on consumer decision making. *International Journal of Human–Computer Interaction, 24*(7), 700–721. https://doi.org/10.1080/10447310802335730

Camilleri, M. A., & Camilleri, A. C. (2017). Digital Learning resources and ubiquitous technologies in education. *Technology, Knowledge and Learning, 22*(1), 65–82. https://doi.org/10.1007/s10758-016-9287-7

Caraban, A., Ferreira, M. J., Gouveia, R., & Karapanos, E. (2015). Social toothbrush: Fostering family nudging around tooth brushing habits. *UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing,* 649–653. https://doi.org/10.1145/2800835.2809438

Caraban, A., Karapanos, E., Gonçalves, D., & Campos, P. (2019). 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3290605.3300733

Caraban, A., Konstantinou, L., & Karapanos, E. (2020). The nudge deck: A design support tool for technology-mediated nudging. *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 395–406. https://doi.org/10.1145/3357236.3395485

Carr, A. (2013, December 12). *How Square register's UI guilts you into leaving tips*. Fast Company. https://www.fastcompany.com/3022182/how-square-registers-ui-guilts-you-into-leaving-tips

Castleman, B. L., & Page, L. C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Behavioral Economics of Education*, *115*, 144–160. https://doi.org/10.1016/j.jebo.2014.12.008

Castleman, B. L., & Page, L. C. (2016). Freshman year financial aid nudges: An experiment to increase fafsa renewal and college persistence. *The Journal of Human Resources*, *51*(2), 389–415.

Castleman, B. L., & Page, L. C. (2017). Parental influences on postsecondary decision making: Evidence from a text messaging experiment. *Educational Evaluation and Policy Analysis*, *39*(2), 361–377. JSTOR.

Castleman, B. L., Page, L. C., & Schooley, K. (2014). The forgotten summer: Does the offer of college counseling after high school mitigate summer melt among college-intending, low-income high school graduates? *Journal of Policy Analysis and Management*, *33*(2), 320–344. https://doi.org/10.1002/pam.21743

Cavus, N. (2011). Investigating mobile devices and LMS integration in higher education: Student perspectives. *World Conference on Information Technology*, *3*, 1469–1474. https://doi.org/10.1016/j.procs.2011.01.033

Cavus, N., & Sekyere-Asiedu, D. (2021). A comparison of online video conference platforms: Their contributions to education during COVID-19 pandemic. *World Journal on Educational Technology: Current Issues*, *13*, 1180–1191. https://doi.org/10.18844/wjet.v13i4.6329

Chapman, P. M. (1997). *Models of engagement: Intrinsically motivated interaction with multimedia learning software* [Doctoral dissertation, University of Waterloo].

Cheng, J., Kulkarni, C., & Klemmer, S. (2013). Tools for predicting drop-off in large online classes. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion*, 121–124. https://doi.org/10.1145/2441955.2441987

Chittaro, L. (2016). Tailoring web pages for persuasion on prevention topics: Message framing, color priming, and gender. In A. Meschtscherjakov, B. De Ruyter, V. Fuchsberger, M. Murer, & M. Tscheligi (Eds.), *Persuasive Technology* (pp. 3–14). Springer International Publishing.

Cockburn, A., Quinn, P., & Gutwin, C. (2015). Examining the peak-end effects of subjective experience. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 357–366. https://doi.org/10.1145/2702123.2702139

Cutrell, E., Czerwinski, M., & Horvitz, E. (2001). Notification, disruption, and memory: Effects of messaging interruptions on memory and performance. *Proceedings of IFIP INTERACT01: Human-Computer Interaction, International Conference on Human-Computer Interaction*. Tokyo, Japan.

Damgaard, M. T., & Nielsen, H. S. (2017). *The use of nudges and other behavioural approaches in education*. European Expert Network on Economics of Education (EENEE). https://eenee.eu/wp-content/uploads/2021/05/EENEE_AR29.pdf

Damgaard, M. T., & Nielsen, H. S. (2018). Nudging in education. *Economics of education review*, *64*, 313–342. https://doi.org/10.1016/j.econedurev.2018.03.008

D'Angelo, C., Rutstein, D., Harris, C., Bernard, R., Borokhovski, E., & Haertel, G. (2014). Simulations for STEM learning: Systematic review and meta-analysis. *Menlo Park: SRI International*, *5*(23), 1–5.

Davis, D., Jivet, I., Kizilcec, R. F., Chen, G., Hauff, C., & Houben, G.-J. (2017). Follow the successful crowd: Raising MOOC completion rates through social comparison at scale. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 454–463. https://doi.org/10.1145/3027385.3027411

Dawood, R., Muchallil, S., & Munadi, K. (2013). Meuruno: An SMS-based learning system that nudge students to review previously taught materials. *Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*, 321–324. https://doi.org/10.1109/TALE.2013.6654453

Disappointing donor week: Majority of Dutch say "no" to being an organ donor (2016, October 19). Nltimes. Retrieved March 11, 2023, from https://nltimes.nl/2016/10/19/disappointing-donor-week-majority-dutch-say-organ-donor

Dobronyi, C. R., Oreopoulos, P., & Petronijevic, U. (2019). Goal setting, academic reminders, and college success: A Large-Scale field experiment. *Journal of Research on Educational Effectiveness*, *12*(1), 38–66. https://doi.org/10.1080/19345747.2018.1517849

Doherty, K., & Doherty, G. (2018). Engagement in HCI: Conception, Theory and Measurement. *ACM Computing Survey*, *51*(5). https://doi.org/10.1145/3234149

Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing behaviour: The mindspace way. *Journal of Economic Psychology*, *33*(1), 264–277. https://doi.org/10.1016/j.joep.2011.10.009

Dweck, C. (2016). What having a "growth mindset" actually means. Harvard Business Review. https://hbr.org/2016/01/what-having-a-growth-mindset-actually-means

Eckles, D., Wightman, D., Carlson, C., Thamrongrattanarit, A., Bastea-Forte, M., & Fogg, B. J. (2009). Social Responses in mobile messaging: Influence strategies, self-disclosure, and source orientation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1651–1654. https://doi.org/10.1145/1518701.1518954

Edwards, S. H., & Li, Z. (2020). A Proposal to use gamification systematically to nudge students toward productive behaviors. *Proceedings of the 20th Koli Calling International Conference on Computing Education Research,* 1-8. https://doi.org/10.1145/3428029.3428057

Egebark, J., & Ekström, M. (2016). Can indifference make the world greener? *Journal of Environmental Economics and Management*, *76*, 1–13. https://doi.org/10.1016/j.jeem.2015.11.004

Einfeld, C. (2019). Nudge and evidence based policy: Fertile ground. *Evidence & Policy: A Journal of Research*, *Debate and Practice*, *15*(4), 509–524. https://doi.org/10.1332/174426418X15314036559759

Esposito, G., Hernández, P., van Bavel, R., & Vila, J. (2017). Nudging to prevent the purchase of incompatible digital products online: An experimental study. *PloS One*, *12*(3), e0173333.

Fahid, F. M., Rowe, J. P., Spain, R. D., Goldberg, B. S., Pokorny, R., & Lester, J. (2021). Adaptively scaffolding cognitive engagement with batch constrained deep Q-networks. *Artificial Intelligence in Education: 22nd International Conference,* 112-124. https://doi.org/10.1007/978-3-030-78292-4_10.

Falloon, G. (2013). Young students using iPads: App design and content influences on their learning pathways. *Computers & Education*, *68*, 505–521. https://doi.org/10.1016/j.compedu.2013.06.006

Fan, X., Luo, W., Menekse, M., Litman, D., & Wang, J. (2015). CourseMIRROR: Enhancing Large Classroom Instructor-Student Interactions via Mobile Interfaces and Natural Language Processing. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 1473–1478. https://doi.org/10.1145/2702613.2732853

Fan, X., Luo, W., Menekse, M., Litman, D., & Wang, J. (2017). Scaling reflection prompts in large classrooms via mobile interfaces and natural language processing. *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, 363–374.

Fancsali, C., Tigani, L., Toro Isaza, P., & Cole, R. (2018). A Landscape study of computer science education in NYC: Early findings and implications for policy and practice. *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, 44–49. https://doi.org/10.1145/3159450.3159467

Fancsali, S. E., Pavelko, M., Fisher, J., Wheeler, L., & Ritter, S. (2021). Scaffolds and Nudges: A Case Study in Learning Engineering Design Improvements. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (pp. 441–445). Springer International Publishing.

Ferreira, M. J., Caraban, A. K., & Karapanos, E. (2014). Breakout: Predicting and breaking sedentary behaviour at work. *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, 2407–2412. https://doi.org/10.1145/2559206.2581330

Fitz-Walter, Z., Tjondronegoro, D., & Wyeth, P. (2012). A gamified mobile application for engaging new students at university orientation. *Proceedings of the 24th Australian Computer-Human Interaction Conference,* 138–141.

Francom, G. M., Schwan, A., & Nuatomue, J. N. (2021). Comparing Google classroom and d2l brightspace using the technology acceptance model. *TechTrends*, *65*(1), 111–119. https://doi.org/10.1007/s11528-020-00533-0

Franklin Jr, D. W., Bryer, J., Lui, A. M., Andrade, H. L., & Akhmedjanova, D. (2022). The effects of nudges on students' use of the diagnostic assessment and achievement of college skills. *Online Learning*, *26*(2), 218–240.

Fritz, J. (2010). Video demo of UMBC's" Check My Activity" tool for students. *Educause Quarterly, 33*(4). https://www.learntechlib.org/p/109175/.

Fritz, J. (2017). Using analytics to nudge student responsibility for learning. *New Directions for Higher Education*, *2017*(179), 65–75. https://doi.org/10.1002/he.20244

Geissinger, H. (1997). Educational software: Criteria for evaluation. *Proceedings of ASCILITE'97 on What Works and Why?*, 219–225. https://www.ascilite.org/conferences/perth97/papers/Geissinger/Geissinger.html

Gielen, A. C., Bishai, D. M., Omaki, E., Shields, W. C., McDonald, E. M., Rizzutti, N. C., Case, J., Stevens, M. W., & Aitken, M. E. (2018). Results of an RCT in two pediatric emergency departments to evaluate the efficacy of an m-health educational app on car seat use. *American Journal of Preventive Medicine*, *54*(6), 746–755. https://doi.org/10.1016/j.amepre.2018.01.042

Godwin-Jones, R. (2011). Mobile apps for language learning. *Language Learning & Technology*, *15*(2), 2–11.

Gorissen, P., van Bruggen, J., & Jochems, W. (2015). Does tagging improve the navigation of online recorded lectures by students? *British Journal of Educational Technology*, *46*(1), 45–57. https://doi.org/10.1111/bjet.12121

Gouveia, R., Pereira, F., Karapanos, E., Munson, S. A., & Hassenzahl, M. (2016). Exploring the design space of glanceable feedback for physical activity trackers. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 144–155. https://doi.org/10.1145/2971648.2971754

Greene, B. A. (2015). Measuring cognitive engagement with self-report scales: Reflections from over 20 years of research. *Educational Psychologist*, *50*(1), 14–30. https://doi.org/10.1080/00461520.2014.989230

Grguric, M. (2023, February 16). Mobile App Retention: The Cold Truth Behind Customer Loyalty. Udonis. https://www.blog.udonis.co/mobile-marketing/mobile-app-retention

Grudin, J. (1992). Utility and usability: Research issues and development contexts. *Interacting with Computers*, *4*(2), 209–217. https://doi.org/10.1016/0953-5438(92)90005-Z

Grüning, D. J., Riedel, F., & Lorenz-Spreen, P. (2023). Directing smartphone use through the self-nudge app one sec. *Proceedings of the National Academy of Sciences*, *120*(8), e2213114120.

Haggerty, D. (2019). *App Retention: 25% of Apps Are Only Used Once*. Applause. https://www.applause.com/blog/app-retention-rates-localytics

Halderman, L. K., Finn, B., Lockwood, J. R., Long, N. M., & Kahana, M. J. (2021). EEG correlates of engagement during assessment. *ETS Research Report Series, 2021*(1), 1–17. https://doi.org/10.1002/ets2.12312

Hansen, P. G., & Jespersen, A. M. (2013). Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation*, *4*(1), 3–28.

Harbach, M., Hettig, M., Weber, S., & Smith, M. (2014). Using personal examples to improve risk communication for security and privacy decisions. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2647–2656. https://doi.org/10.1145/2556288.2556978

Harley, D., Winn, S., Pemberton, S., & Wilcox, P. (2007). Using texting to support students' transition to university. *Innovations in Education and Teaching International*, *44*(3), 229–241. https://doi.org/10.1080/14703290701486506

Haslam, J. (2019). *What makes a good retention rate?*. Adjust. Retrieved March 12, 2022, https://www.adjust.com/blog/what-makes-a-good-retention-rate/

He, M. (2018). *Designing an educational app to teach Chinese characters for heritage language education* [Doctoral dissertation, University of Delaware]. https://udspace.udel.edu/handle/19716/23813

Hewett, T. T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G., & Verplank, W. (1992). *ACM SIGCHI curricula for Human-Computer Interaction*. Association for Computing Machinery.

Hirano, S. H., Farrell, R. G., Danis, C. M., & Kellogg, W. A. (2013). WalkMinder: Encouraging an active lifestyle using mobile phone interruptions. *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, 1431–1436. https://doi.org/10.1145/2468356.2468611

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "Educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest*, *16*(1), 3–34. https://doi.org/10.1177/1529100615569721

Hoxby, C., Turner, S., & others. (2013). Expanding college opportunities for high-achieving, low income students. *Stanford Institute for Economic Policy Research Discussion Paper*, *12*, 014.

Hummel, D., Schacht, S., & Maedche, A. (2017). Designing adaptive nudges for multi-channel choices of digital services: A laboratory experiment design. *Proceedings of the 25th European Conference on Information Systems*, 2677–2688. https://aisel.aisnet.org/ecis2017_rip/20

Hwang, G.-J., Tsai, C.-C., Chu, H.-C., Kinshuk, K., & Chen, C.-Y. (2012). A context-aware ubiquitous learning approach to conducting scientific inquiry activities in a science park. *Australasian Journal of Educational Technology*, *28*(5). https://doi.org/10.14742/ajet.825

Hwang, Y., & Vrongistinos, K. (2012). Using blackboard and skype for mentoring beginning teachers. *American Journal of Distance Education*, *26*, 172–179. https://doi.org/10.1080/08923647.2012.697019

Islas Sedano, C., Leendertz, V., Vinni, M., Sutinen, E., & Ellis, S. (2013). Hypercontextualized learning games: fantasy, motivation, and engagement in reality. *Simulation & Gaming*, *44*(6), 821–845. https://doi.org/10.1177/1046878113514807

Ivan Stojšić, Anđelija Ivkov Džigurski, Olja Maričić, Ljubica Ivanović Bibić, & Smiljana Đukičin Vučković. (2017). Possible application of virtual reality in geography teaching. *journal of subject didactics*, *1*(2). https://doi.org/10.5281/zenodo.438169

Leonard, Jay. (2022, December 19). *8 Tips for boosting app retention: Create an app for the long run*. Business 2 community. https://www.business2community.com/mobile-apps/8-tips-boosting-app-retention-create-app-long-run-01172687

Jesse, M., & Jannach, D. (2021). Digital nudging with recommender systems: Survey and future directions. *Computers in Human Behavior Reports*, *3*, 100052. https://doi.org/10.1016/j.chbr.2020.100052

Jiang, R., Qie, X., Li, Z., Zhang, H., Li, X., Yuan, S., Liu, M., Sun, Z., Srivastava, A., Liu, M., Ma, Z., & Lv, G. (2020). Luminous crown residual vs. Bright space segment: Characteristical structures for the intermittent positive and negative leaders of triggered lightning. *Geophysical Research Letters*, *47*(21), e2020GL088107. https://doi.org/10.1029/2020GL088107

John, peter, smith, graham, & stroker, gerry. (2009). Nudge nudge, think think: Two strategies for changing civic behaviour. *The Political Quarterly*, *80*(3), 361–370. https://doi.org/10.1111/j.1467-923X.2009.02001.x

Jung, D., Erdfelder, E., & Glaser, F. (2018). *Nudged to win: Designing robo-advisory to overcome decision inertia*. https://aisel.aisnet.org/ecis2018_rip/19

Kammerer, Y., & Gerjets, P. (2014). The role of search result position and source trustworthiness in the selection of web search results when using a list or a grid interface. *International Journal of Human–Computer Interaction*, *30*(3), 177–191. https://doi.org/10.1080/10447318.2013.846790

Kankane, S., DiRusso, C., & Buckley, C. (2018). Can we nudge users toward better password management? An initial study. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3170427.3188689

Kaptein, M., Markopoulos, P., de Ruyter, B., & Aarts, E. (2015). Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies*, *77*, 38–51. https://doi.org/10.1016/j.ijhcs.2015.01.004

Karlsen, R., & Andersen, A. (2019). Recommendations with a Nudge. *Technologies*, *7*(2), 45. https://doi.org/10.3390/technologies7020045

Kavanagh, S., Luxton-Reilly, A., Wuensche, B., & Plimmer, B. (2017). A systematic review of Virtual Reality in education. *Themes in Science and Technology Education*, *10*(2), 85–119.

Kayalar, F. (2016). Cross-cultural comparison of teachers' views upon integration and use of technology in classroom. *Turkish Online Journal of Educational Technology-TOJET*, *15*(2), 11–19.

Kelders, S. M., Van Zyl, L. E., & Ludden, G. D. (2020). The concept and components of engagement in different domains applied to ehealth: A systematic scoping review. *Frontiers in Psychology*, *11*, 926. https://doi.org/10.3389/fpsyg.2020.00926

Kim, S., & Baek, T. H. (2018). Examining the antecedents and consequences of mobile app engagement. *Telematics and Informatics*, *35*(1), 148–158. https://doi.org/10.1016/j.tele.2017.10.008

Kim, S., Chang, M., Deater-Deckard, K., Evans, M. A., Norton, A., & Samur, Y. (2017). Educational games and students' game engagement in elementary school classrooms. *Journal of Computers in Education*, *4*(4), 395–418. https://doi.org/10.1007/s40692-017-0095-4

Kim, S. W. (2016). Design and implementation of socket-based multi-player game service for elementary mathematics learning. *Journal of Service Research and Studies*, *6*(2), 175–184.

Kincaid, J. P., Hamilton, R., Tarr, R. W., & Sangani, H. (2003). Simulation in education and training. In M. S. Obaidat & G. I. Papadimitriou (Eds.), *Applied system simulation: methodologies and applications* (pp. 437–456). Springer US. https://doi.org/10.1007/978-1-4419-9218-5_19

Kizilcec, R., Schneider, E., Cohen, G., & McFarland, D. (2014). Encouraging forum participation in online courses with collectivist, individualist and neutral motivational framings. *ELearning Papers*, *37*, 13–22.

Konstantinou, L., Caraban, A., & Karapanos, E. (2019). Combating misinformation through nudging. In D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, & P. Zaphiris (Eds.), *Human-Computer interaction – interact 2019* (pp. 630–634). Springer International Publishing.

Korableva, O., Durand, T., Kalimullina, O., & Stepanova, I. (2019). Studying user satisfaction with the MOOC platform interfaces using the example of coursera and open education platforms. *Proceedings of the 2019 International Conference on Big Data and Education*, 26–30. https://doi.org/10.1145/3322134.3322139

Larkin, K. (2015). "An App! An App! My Kingdom for An App": An 18-Month quest to determine whether apps support mathematical knowledge building. In T. Lowrie & R. Jorgensen (Zevenbergen) (Eds.), *Digital games and mathematics learning: Potential, promises and pitfalls* (pp. 251–276). Springer. https://doi.org/10.1007/978-94-017-9517-3_13

Lee, M. K., Kiesler, S., & Forlizzi, J. (2011). Mining behavioral economics to design persuasive technology for healthy choices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 325–334. https://doi.org/10.1145/1978942.1978989

Leonard, Jay. (2022, December 19). *8 Tips for boosting app retention: Create an app for the long run*. Business 2 community. https://www.business2community.com/mobile-apps/8-tips-boosting-app-retention-create-app-long-run-01172687

Lembcke, T.-B., Engelbrecht, N., Brendel, A., & Kolbe, L. (2019). To nudge or not to nudge: Ethical considerations of digital nudging based on its behavioral economics roots. Proceedings of the 25th European Conference on Information Systems. https://aisel.aisnet.org/ecis2019_rp/95

Lemken, D. (2021). Options to design more ethical and still successful default nudges: A review and recommendations. *Behavioural Public Policy*, 1–33. Cambridge Core. https://doi.org/10.1017/bpp.2021.33

Lim, T. M., & Yunus, M. M. (2021). Teachers' perception towards the use of quizizz in the teaching and learning of english: A Systematic review. *Sustainability*, *13*(11). https://doi.org/10.3390/su13116436

Lin, C. Y., Wu, T. F., & Hou, W. H. (2018). The effect of health educational app on improving stroke related knowledge and quality of life in patients with stroke: A randomized controlled trial. *12th World Congress of the International Society of Physical and Rehabilitation Medicine,* e521–e522. https://doi.org/10.1016/j.rehab.2018.05.1214

Littlemore, K., & Farmer, R. (2014). Using Blackboard Learn to create an open online course. *Blackboard Teaching and Learning Conference, University College Dublin*. Dublin, Ireland.

Looi, C.-K., Zhang, B., Chen, W., Seow, P., Chia, G., Norris, C., & Soloway, E. (2011). 1:1 mobile inquiry learning experience for primary science students: A study of learning effectiveness. *Journal of Computer Assisted Learning*, *27*(3), 269–287. https://doi.org/10.1111/j.1365-2729.2010.00390.x

Loughry, M. L., Ohland, M. W., & Woehr, D. J. (2013). Assessing teamwork skills for assurance of learning using CATME team tools. *Journal of Marketing Education*, *36*(1), 5–19. https://doi.org/10.1177/0273475313499023

Luo, W., Fan, X., Menekse, M., Wang, J., & Litman, D. (2015). Enhancing instructor-student and student-student interactions with mobile interfaces and summarization. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, 16–20. https://doi.org/10.3115/v1/N15-3004

Magooda, A., Elaraby, M., & Litman, D. (2021). Exploring Multitask Learning for Low-Resource AbstractiveSummarization. *ArXiv Preprint ArXiv:2109.08565*.

Magooda, A., Litman, D., Ashraf, A., Menekse, M. (2022). Improving the quality of students' written reflections using natural language processing: Model design and classroom evaluation. In: Rodrigo, M.M., Matsuda, N., Cristea, A.I., Dimitrova, V. (Eds.), Artificial intelligence in education: Lecture notes in computer science (pp. 519-525). Springer. https://doi.org/10.1007/978-3-031-11644-5_43

Manser, M. (2016, August 24). *3 Mobile engagement tactics to better connect with students*. Airship. https://www.airship.com/blog/3-mobile-engagement-tactics-to-better-connect-with-students/

Matz, R., Schulz, K., Hanley, E., Derry, H., Hayward, B., Koester, B., Hayward, C., & McKay, T. (2021). Analyzing the efficacy of ECoach in supporting gateway course success through tailored support. *LAK21: 11th International Learning Analytics and Knowledge Conference*, 216–225. https://doi.org/10.1145/3448139.3448160

Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology*, *70*(1), 531–549. https://doi.org/10.1146/annurev-psych-010418-102744

McEvoy, D. (2016). Loss aversion and student achievement. *Economics Bulletin*, *36*(3), 1762–1770.

McKay, T., Miller, K., & Tritz, J. (2012). What to do with actionable intelligence: E^2coach as an intervention engine. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 88–91. https://doi.org/10.1145/2330601.2330627

Melcher, J., Camacho, E., Lagan, S., & Torous, J. (2022). College student engagement with mental health apps: Analysis of barriers to sustained use. *Journal of American College Health*, *70*(6), 1819–1825. https://doi.org/10.1080/07448481.2020.1825225

Mele, C., Russo Spena, T., Kaartemo, V., & Marzullo, M. L. (2021). Smart nudging: How cognitive technologies enable choice architectures for value co-creation. *Journal of Business Research*, *129*, 949–960. https://doi.org/10.1016/j.jbusres.2020.09.004

Menekse, M. (2020). The Reflection-informed learning and instruction to improve students' academic success in undergraduate classrooms. *The Journal of Experimental Education*, *88*(2), 183–199. https://doi.org/10.1080/00220973.2019.1620159

Menekse, M., Anwar, S., & Purzer, S. (2018). Self-Efficacy and mobile learning technologies: a case study of CourseMIRROR. In C. B. Hodges (Ed.), *Self-Efficacy in instructional technology contexts* (pp. 57–74). Springer International Publishing. https://doi.org/10.1007/978-3-319-99858-9_4

Menon, D. (2022). Uses and gratifications of educational apps: A study during COVID-19 pandemic. *Computers and Education Open*, *3*, 100076. https://doi.org/10.1016/j.caeo.2022.100076

Meske, C., & Potthoff, T. (2017). The DINU-model–a process model for the design of nudges. *25th European Conference on Information Systems (ECIS)*, 2587–2597. https://aisel.aisnet.org/ecis2017_rip/11

Mirsch, T., Lehrer, C., & Jung, R. (2017). Digital nudging: Altering user behavior in digital environments. *Proceedings of the 13th International Conference on Business Information Systems (WI 2017)*, 634–648.

Mitchell, K. M., McMillan, D. E., & Rabbani, R. (2019). An exploration of writing self-efficacy and writing self-regulatory behaviours in undergraduate writing. *The Canadian Journal for the Scholarship of Teaching and Learning*, *10*(2).

Mkpojiogu, E. O., Hussain, A., & Hassan, F. (2018). A systematic review of usability quality attributes for the evaluation of mobile learning applications for children. *AIP Conference Proceedings*, *2016*, 020092. https://doi.org/10.1063/1.5055494

Mohammadhassan, N., Mitrovic, A., Neshatian, K., & Dunn, J. (2020). Automatic assessment of comment quality in active video watching. In H.J SO, J.M, M.M Rodrigo, J. Mason, A. Mitrovic (Eds.), *Proceedings of the 28th international conference on computers in education* (pp. 1-10). Asia-Pacific Society for Computers in Education (APSCE). https://apsce.net/upfile/icce2020/ICCE%202020%20Proceedings%20-%20Volume%20I%20v4.pdf

Motz, B. A., Mallon, M. G., & Quick, J. D. (2021). Automated educative nudges to reduce missed assignments in college. *IEEE Transactions on Learning Technologies*, *14*(2), 189–200. https://doi.org/10.1109/TLT.2021.3064613

Munson, S. A., Krupka, E., Richardson, C., & Resnick, P. (2015). Effects of public commitments and accountability in a technology-supported physical activity intervention. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1135–1144. https://doi.org/10.1145/2702123.2702524

Muslimin, M. S., Nordin, N. M., Mansor, A. Z., & Yunus, M. M. (2017). The design and development of MobiEko: A mobile educational app for microeconomics module. *Malaysian Journal of Learning and Instruction*, 221–255. https://doi.org/10.32890/mjli.2017.7804

Nekmat, E. (2020). Nudge Effect of Fact-Check Alerts: Source influence and media skepticism on sharing of news misinformation in social media. *Social Media + Society*, *6*(1). https://doi.org/10.1177/2056305119897322

Nelson, K. J., Quinn, C., Marrington, A., & Clarke, J. A. (2012). Good practice for enhancing the engagement and success of commencing students. *Higher Education*, *63*(1), 83–96. https://doi.org/10.1007/s10734-011-9426-y

Neuhaus, T. (2020). Nudging education – The (potential) role of decision architectures in improving educational settings. *International Journal of Education and Research*, *8*(9). http://www.ijern.com/journal/2020/September-2020/07.pdf

Nguyen, Q. T., & Vo, N. H. K. (2021). Educational nudge: Field interventions to enhance students' engagement in the english classrooms during the Sars-Cov-2 pandemic. *Proceeding of the 9th Open TESOL International Conference*, 61-82.

O'Brien, H. (2016). Theoretical perspectives on user engagement. In H. O'Brien & P. Cairns (Eds.), *Why engagement matters: Cross-disciplinary perspectives of user engagement in digital media* (pp. 1–26). Springer International Publishing. https://doi.org/10.1007/978-3-319-27446-1_1

O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, *59*(6), 938–955. https://doi.org/10.1002/asi.20801

O'Hara, R. E., & Sparrow, B. (2019). A summer nudge campaign to motivate community college STEM students to reenroll. *AERA Open*, *5*(3), 2332858419875715. https://doi.org/10.1177/2332858419875715

Oinas-Kukkonen, H. (2010). Behavior change support systems: A research model and agenda. *International Conference on Persuasive Technology*, 4–14.

O'Neill, V. E. M. (2019). Using Push Technology to Improve Retention in Undergraduate STEM Students (Publication No. 2454687023) [Doctoral Dissertation, New Jersey City University]. ProQuest Dissertations Publishing

Pachler, N., Bachmair, B., & Cook, J. (2009). *Mobile learning: Structures, agency, practices*. Springer.

Papadakis, S., & Kalogiannakis, M. (2017). Mobile educational applications for children: What educators and parents need to know. *International Journal of Mobile Learning and Organisation*, *11*(3), 256–277.

Papadakis, S., Kalogiannakis, M., & Zaranis, N. (2018). Educational apps from the Android Google Play for Greek preschoolers: A systematic review. *Computers & Education*, *116*, 139–160. https://doi.org/10.1016/j.compedu.2017.09.007

Patterson, R. W. (2018). Can behavioral tools improve online student outcomes? Experimental evidence from a massive open online course. *Journal of Economic Behavior & Organization, 153*, 293–321. https://doi.org/10.1016/j.jebo.2018.06.017

Paunov, Y., Wänke, M., & Vogel, T. (2019). Ethical defaults: Which transparency components can increase the effectiveness of default nudges? *Social Influence*, *14*(3–4), 104–116. https://doi.org/10.1080/15534510.2019.1675755

Pechenkina, E., Laurence, D., Oates, G., Eldridge, D., & Hunter, D. (2017). Using a gamified mobile app to increase student engagement, retention and academic achievement. *International Journal of Educational Technology in Higher Education*, *14*, 1–12.

Pham, X. L., & Chen, G. D. (2018). PACARD: A New interface to increase mobile learning app engagement, distributed through app stores. *Journal of Educational Computing Research*, *57*(3), 618–645. https://doi.org/10.1177/0735633118756298

Pham, X., Nguyen, T., Hwang, W., & Chen, G. (2016). Effects of push notifications on learner engagement in a mobile learning app. *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, 90–94. https://doi.org/10.1109/ICALT.2016.50

Pichert, D., & Katsikopoulos, K. V. (2008). Green defaults: Information presentation and pro-environmental behaviour. *Journal of Environmental Psychology*, *28*(1), 63–73. https://doi.org/10.1016/j.jenvp.2007.09.004

Pielot, M., Church, K., & de Oliveira, R. (2014). An in-situ study of mobile phone notifications. *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices Services*, 233–242. https://doi.org/10.1145/2628363.2628364

Purdue Engineering Degree Programs & Enrollment 2021. (n.d.). College of Engineering - Purdue University. Retrieved February 26, 2023, from https://engineering.purdue.edu/Engr/AboutUs/FactsFigures/EnrollmentAndDegrees/2021-enrollment

Quinn, A. (2018). Learning to program with SoloLearn. *The Mathematics Teacher*, *112*(3), 226–230.

Rocha, C. V., & Coutinho, C. P. (2015). Assessing pedagogical dimensions of an educational app: A tool for teachers and educators. In D. Garbett & A. ovens (Eds.), *Teaching for Tomorrow Today* (pp. 123-132). Edify Ltd.

Rogers, Y., Hazlewood, W. R., Marshall, P., Dalton, N., & Hertrich, S. (2010). Ambient influence: Can twinkly lights lure and abstract representations trigger behavioral change? *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, 261–270. https://doi.org/10.1145/1864349.1864372

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Schmidt, A. T., & Engelen, B. (2020). The ethics of nudging: An overview. *Philosophy Compass*, *15*(4), e12658.

Schneider, A. L. J., & Graham, T. C. N. (2017). Nudging and shoving: Using in-game cues to guide player exertion in exergames. *Entertainment Computing*, *19*, 83–100. https://doi.org/10.1016/j.entcom.2017.01.002

Schneider, C., Weinmann, M., & Brocke, J. vom. (2018). Digital nudging: Guiding online user choices through interface design. *Communications of the ACM*, *61*, 67–73. https://doi.org/10.1145/3213765

Sherr, G., Akkaraju, S., & Atamturktur, S. (2019). Nudging students to succeed in a flipped format gateway biology course. *Journal of Effective Teaching in Higher Education*, *2*(2), 57–69.

Shuler, C. (2012, January 12). *ILearn II: An Analysis of the education category on apple's app store*. The Joan Ganz Cooney Center.

Simmons, L., Crook, A., Cannonier, C., & Simmons, C. (2018). There's an app for that: The impact of reminder apps on student learning and anxiety. *Journal of Education for Business*, *93*(5), 185–195. https://doi.org/10.1080/08832323.2018.1441120

Singler, K., Roth, T., Beck, S., Cunningham, M., & Gosch, M. (2016). Development and initial evaluation of a point-of-care educational app on medical topics in orthogeriatrics. *Archives of Orthopaedic and Trauma Surgery*, *136*(1), 65–73.

Soe, T. H., Nordberg, O. E., Guribye, F., & Slavkovik, M. (2020). Circumvention by design—Dark patterns in cookie Consent for online news outlets. *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society,* 1-12. https://doi.org/10.1145/3419249.3420132

Song, Y., Wen, Y., Yang, Y., & Cao, J. (2022). Developing a 'Virtual Go mode' on a mobile app to enhance primary students' vocabulary learning engagement: An exploratory study. *Innovation in Language Learning and Teaching*, 1–10. https://doi.org/10.1080/17501229.2022.2047693

Soricone, L., & Endel, B. (2019). *Nudging to STEM success*. Persistance Plus. https://jfforg-prod-prime.s3.amazonaws.com/media/documents/Nudging_to_STEM_Success.pdf

Sruthi, P., & Mukherjee, S. (2020). Byju's the learning app: An investigative study on the transformation from traditional learning to technology based personalized learning. *International Journal of Scientific and Technology Research*, *9*(3), 5054–5059.

Sticklen, J., Amey, M., Eskil, T., Hinds, T., & Urban-Lurain, M. (2004). Application of scaffolding to introductory matlab. *Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition, American Society of Engineering Education.* Salt Lake City, Utah.

Szaszi, B., Palinkas, A., Palfi, B., Szollosi, A., & Aczel, B. (2018). A systematic scoping review of the choice architecture movement: Toward understanding when and why nudges work. *Journal of Behavioral Decision Making*, *31*(3), 355–366. https://doi.org/10.1002/bdm.2035

Teuber, M., Leyhr, D., Moll, J., & Sudeck, G. (2022). Nudging digital physical activity breaks for home studying of university students—A randomized controlled trial during the COVID-19 pandemic with daily activity measures. *Frontiers in Sports and Active Living*, *4*.

Thaler, R., & Sunstein, C. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Tuckman, B. W. (1998). Using tests as an incentive to motivate procrastinators to study. *The Journal of Experimental Education*, *66*(2), 141–147. https://doi.org/10.1080/00220979809601400

Tularam, G. A. (2018). Traditional vs non-traditional teaching and learning strategies-the case of e-learning!. *International Journal for Mathematics Teaching and Learning*, *19*(1), 129–158.

Turland, J., Coventry, L., Jeske, D., Briggs, P., & van Moorsel, A. (2015). Nudging towards security: Developing an application for wireless network selection for android phones. *Proceedings of the 2015 British HCI Conference*, 193–201.

Unkovic, C., Sen, M., & Quinn, K. M. (2016). Does encouragement matter in improving gender imbalances in technical fields? Evidence from a randomized controlled trial. *PloS One*, *11*(4), e0151714.

User retention rate for mobile apps and websites (2017, December 4). Mixpanel. Retrieved March 12, 2023, from https://mixpanel.com/blog/whats-a-good-retention-rate/

van der Laan, L. N., & Orcholska, O. (2022). Effects of digital just-in-time nudges on healthy food choice – A field experiment. *Food Quality and Preference*, *98,* 104535. https://doi.org/10.1016/j.foodqual.2022.104535

Van Lange, P. A. M., Higgins, E. T., & Kruglanski, A. W. (2011). *Handbook of theories of social psychology*. SAGE Publications. http://digital.casalini.it/9781446250068

van Oldenbeek, M., Winkler, T. J., Buhl-Wiggers, J., & Hardt, D. (2019). Nudging in blended learning: Evaluation of email-based progress feedback in a flipped-classroom information systems course. *27th European Conference on Information Systems (ECIS)*. https://aisel.aisnet.org/ecis2019_rp/186

Wambsganss, T., Janson, A., & Leimeister, J. M. (2022). Enhancing argumentative writing with automated feedback and social comparison nudging. *Computers & Education*, *191*, 104644. https://doi.org/10.1016/j.compedu.2022.104644

Wanchid, R. (2013). The use of self-correction, paper-pencil peer feedback and electronic peer feedback in the EFL writing class: Opportunities and challenges. *Academic Journal of Interdisciplinary Studies*, *2*(3), 157.

Wang, Y., Leon, P. G., Acquisti, A., Cranor, L. F., Forget, A., & Sadeh, N. (2014). A field trial of privacy nudges for facebook. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2367–2376. https://doi.org/10.1145/2556288.2557413

Weijers, R. J., de Koning, B. B., & Paas, F. (2020). Nudging in education: From theory towards guidelines for successful implementation. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-020-00495-0

Weinmann, M., Schneider, C., & Brocke, J. vom. (2016). Digital nudging. *Business & Information Systems Engineering*, *58*(6), 433–436. https://doi.org/10.1007/s12599-016-0453-1

Weston, A., Morrison, L., Yardley, L., Kleek, M. V., & Weal, M. (2015, May). Measurements of engagement in mobile behavioural interventions?. *Digital Health*. https://eprints.soton.ac.uk/377613/

Wilkinson, D., Sivakumar, S., Cherry, D., Knijnenburg, B. P., Raybourn, E. M., Wiseniewski, P., & Sloan, H. (2017). *Work in progress: User-tailored privacy by design*. Sandia National Lab. https://www.ndss-symposium.org/ndss2017/usec-mini-conference-programme/work-progress-user-tailored-privacy-design/

Williams, R. (2021, February 23). *How a "Nudge" Can Change Behaviour*. Ray Williams. https://raywilliams.ca/nudge-can-change-behaviour/

Wingate, U. (2010). The impact of formative feedback on the development of academic writing. *Assessment & Evaluation in Higher Education*, *35*, 519–533. https://doi.org/10.1080/02602930903512909

Wu, H.-K., Lee, S. W.-Y., Chang, H.-Y., & Liang, J.-C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, *62*, 41–49. https://doi.org/10.1016/j.compedu.2012.10.024

Zamprogno, L., Holmes, R., & Baniassad, E. (2020). Nudging student learning strategies using formative feedback in automatically graded assessments. *Proceedings of the 2020 ACM SIGPLAN Symposium on SPLASH-E*, 1–11. https://doi.org/10.1145/3426431.3428654

Zhang, J., & Liao, B. (2015). Learning on the Fingertips: The opportunities and challenges of educational apps. *Journal of Education and Practice*, *6*(20), 62–67.

Zhao, F. (2019). Using Quizizz to integrate fun multiplayer activity in the accounting classroom. *International Journal of Higher Education*, *8*(1), 37–43.

Zydney, J. M., & Warner, Z. (2016). Mobile apps for science learning: Review of research. *Computers & Education*, *94*, 1–17. https://doi.org/10.1016/j.compedu.2015.11.001

APPENDIX A: REFLECTION SPECIFICITY PREDICTION ALGORITHM

In the CourseMIRROR mobile application, students were scaffolded through some prompts to improve the reflection based on the specificity score evaluated using a Natural language processing algorithm during the reflection writing process. In this study, I used the natural language processing (NLP) algorithm proposed by Magooda et al. (2022) to evaluate the specificity score and design my intervention based on the specificity score. The proposed NLP approach consists of a simple model where we generated feature based on the students' reflections and then use the classification module to produce the reflection specificity score. For feature generation, the transformer-based bidirectional deep contextual language model was used to automatically generate numerical features from raw students' reflection text using the DistilBERT model. The DistilBERT model used in our implementation is a distilled version of the original BERT transformer-based encoder. By reducing the number of parameters to approximately 60% of the original BERT, DistilBERT is faster and better suited for real-time quality prediction.

Furthermore, we used a logistic regression classifier that operated on the generated features in the previous step and produced a reflection specificity score of 1,2,3, or 4. We only trained the logistic regression classification module to minimize the training load and kept the DistilBERT parameters fixed. This approach also enables us to fine-tune the model with new samples over time easily.