DISSERTATION

TIMING IS EVERYTHING: INVESTIGATING USERS' FEEDBACK TIMING AND CONTENT PREFERENCES TO ENHANCE COMPUTER-BASED TRAINING

Submitted by

Rebecca L. Pharmer

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2026

Doctoral Committee:

Advisor: Matthew G. Rhodes

Christopher D. Wickens
Niccole A. Nelson
Franisco R. Ortega

ABSTRACT

TIMING IS EVERYTHING: INVESTIGATING USERS' FEEDBACK TIMING AND CONTENT PREFERENCES TO ENHANCE COMPUTER-BASED TRAINING

Feedback is a central instructional component of computerized training systems. Research has shown that the timing and content of feedback impacts learning but less is currently known about how user preferences for the way feedback is received during training impacts its overall effectiveness. Across three experiments, this work examined whether users demonstrate a preference for the timing of feedback they receive (Experiment 1), whether honoring that preference improves their learning outcomes (Experiment 2), and whether honoring learner preferences for the content within their feedback (i.e., whether it is outcome based or explanatory) likewise impacts learning outcomes (Experiment 3). Across experiments, learners consistently preferred immediate over delayed feedback and explanatory over outcome feedback. However, aligning feedback with these preferences did not reliably improve learning outcomes and, in some cases, non-preferred conditions produced equal or better learning performance. Preference alignment more consistently influenced subjective experience, with differences in perceived mental workload, frustration, and performance despite similar objective outcomes. Overall, the results suggest that stable learner preferences do not reliably predict which feedback conditions best support learning. Instead, learning depends more on feedback design than preference matching.

ACKNOWLEDGEMENTS

I first want to sincerely thank my advisor, Dr. Matt Rhodes, for his guidance, patience, and unwavering support throughout my doctoral training. Thank you for taking me on as a mentee and encouraging my growth as a researcher. Your example as both a researcher and mentor has profoundly shaped how I approach my work, and I strive to carry that same dedication and thoughtfulness forward in my own career. I am also incredibly grateful for my committee members' support and mentorship. Notably, Dr. Chris Wickens has pushed me to grow as an applied scientist throughout my time here and I am honored to have absorbed even a fraction of his expertise in the field of Human Factors. Dr. Francisco Ortega has encouraged me to be confident in my abilities and provided me with opportunities and support to pursue research in adaptive training. Dr. Niccole Nelson has been instrumental in my growth as an instructor, giving me the opportunity to mentor undergraduates through their first experiences with research methods and statistics – a role through which I discovered a passion for teaching I did not know I had. Additionally, I would like to thank my master's advisor, Dr. Ben Clegg, for believing in my potential to complete this program and for laying the foundation for my success from the very beginning.

I would also like to extend my gratitude to the individuals who supported me throughout my graduate education. While there are too many to name, I'll do my best to highlight a few. First, my peers in the Cognitive Psychology program, with whom I have been able to celebrate successes, navigate failures, and build a supportive, lifelong community. My dear friends, Brooke Carlaw, Haley McCoy, Dr. Colleen Patton, and Dr. Amelia Warden – I cannot see a world in which I completed this program without your friendship and support. Moreover, I would like to acknowledge Jesse Hazel, for being a constant source of support and a trusted confidant.

I am equally grateful to my family. My mom, Lorrie Pharmer, not only encouraged my pursuit of higher education but also demonstrated a strong work ethic and confidence in her expertise. My dad, Dr. Jim Pharmer, inspired me to pursue my PhD in the first place and has been a listening ear to empathize about the trials and tribulations of graduate school. Finally, I must acknowledge my husband, Braddock McDonald. It is impossible to put into words how much I deeply value you and your support, not only during this program but throughout the past decade. I am so proud of the life we have built together, and a consistent source of motivation in finishing this degree has been the future we work towards side-by-side. I love you with my whole heart, and I look forward to our next chapter.

TABLE OF CONTENTS

CHAPTER 1 – INTRODUCTION

Computerized training systems are now woven into nearly every corner of modern learning – from classrooms that rely on learning platforms, to corporate and military training programs, to consumer-facing tools like language learning mobile apps. The National Center for Education Statistics (NCES; 2023) reported that approximately 60% of undergraduate students in the United States were enrolled in at least one distance learning course in 2021, with 30% learning exclusively online. Although computer-based training offers scalable, flexible, and personalized instruction, their effectiveness depends on more than their flexibility. Successful learning relies on instructional design choices within these platforms that moderate the content presented to the learner and the support they receive. One of the most critical design choices concerns feedback or how the system responds when a learner makes an error (Kluger & DiNisi, 1996; Durlach & Ray, 2011; Vandewaetere, 2011).

Feedback during learning provides more than simply correcting mistakes. Feedback directs attention, reinforces knowledge, and influences learner motivation (Durlach & Spain, 2014; Durlach, 2019; Johnson, et al., 2017). Poorly designed feedback interventions can lead to frustration or overload for the learner, potentially discouraging them from further practice. Conversely, well-designed feedback can aid knowledge and skill acquisition and build confidence (Durlach & Spain, 2014; Durlach, 2019; Kluger & DiNisi, 1996; Lansberg et al., 2010). In traditional methods of instruction, either in the classroom or one-on-one tutoring, feedback can be provided in real-time. However, in computerized systems, feedback must be programmed in advance – requiring consideration of each possible error and the most optimal response. Moreover, not all learners will want or benefit from the same type of feedback. Some may prefer immediate correction of their mistakes as they happen while others may find it more beneficial to reflect on their performance after-the-fact. Although preferences may not always align with what is empirically demonstrated to lead to better learning (Bjork et al., 2013; Deslauriers et al., 2019; Soderstrom & Bjork, 2015), they influence enjoyment and willingness to persist in training (Dunlosky & Rawson, 2012; Durlach & Spain, 2014; Metcalfe & Finn, 2008).

As computerized training becomes more sophisticated and widely adopted, this intersection—between the science of effective instruction and the realities of user experience—offers both practical and theoretical insights. A closer look at feedback design can help ensure computerized training systems are not just efficient, but also engaging and responsive to the learners who use them.

Adaptive Training Systems

Adaptive training systems (ATS) are computerized training programs that provide personalized tailoring of instructional content to learners (Durlach & Ray, 2011; Park & Lee, 2004). These systems are capable of tracking and analyzing behavior patterns via response time or accuracy data gathered during training, assessing these patterns to identify learners' strengths and weaknesses in a given domain (Durlach, 2019). Further, the fundamental strength of ATS is their ability to adjust training based on the learner's behavior. Typical interventions within these systems include adjusting the difficulty of the task based on accuracy or performance, removal and replacement of scaffolding, spacing of instruction, and error-sensitive feedback (Durlach & Ray, 2011). Meta-analyses examining the effectiveness of ATS have shown that they can be as effective as one-on-one human tutoring, often considered the 'gold-standard' of instruction due to a tutor's ability to personalize training in real time (Ma et al., 2014; VanLehn, 2011). This personalized approach to instruction may increase engagement and effectively challenge the learner at their level of expertise, without overloading them with difficulty beyond their skill level (Durlach & Ray, 2011; Ma et al., 2014). Of particular relevance to applied domains, adaptive training systems have been shown to increases the efficiency of training, allowing for a quicker return on investment of training resources (Durlach & Ray, 2011). However, these systems often require a substantial upfront investment for development and testing to identify an effective adaptive schedule appropriate to the focal training task. There is presently no generalized adaptive framework that assures optimal learning outcomes (Durlach, 2019) and designers of ATS must weigh considerations such adaptation schedules, feedback, and system usability. Durlach (2019) argues that the effectiveness of ATS reflects how well they are designed and that research concerning usability or user preferences within the domain is sparse compared to research on the effectiveness of instruction. Given the many variables tied to both presentation and content, error sensitive feedback is among the most critical and complex interventions,.

Cognitive Demands During Learning

To guide decisions for developing effective adaptive trainers, developers must consider frameworks outlining how instructional interventions, like feedback, may impact cognitive resource demands on the learner to ensure optimal learning outcomes. Humans have limited working memory capacity (Baddeley, 2000; Cowan, 1999; Wickens, 2008; Wickens, 2021) and learning imposes a heavy demand on the learner's attentional and memory resources. Sweller's Cognitive Load Theory (CLT) outlines three types of cognitive load related to learning (Chandler & Sweller, 1991; Paas & Sweller, 2014; Sweller, 2011). The first, *intrinsic load*, refers to task complexity. Sweller (2011) notes that task complexity stems from 'element interactivity' or the number of elements learners must hold and integrate into working memory to learn a concept (e.g., remembering chemical symbols in chemistry). If there are several elements that must be learned/integrated simultaneously, then the load is high (e.g., balancing chemical equations). Such load is inherent to the task and may only change as the learner becomes more experienced and holds elements of the learned task as mental representations. Therefore, intrinsic load should be appropriate for the skill level of the learner. *Extraneous load* describes load imposed on the learner by suboptimal instructional design. This occurs when learners are forced to split attention through the same processing channels or when the coherence of instruction is poor or lacks contiguity (Mayer & Moreno, 1998a). Finally, *germane load* relates to the cognitive resources devoted to learning. Germane load is deemed "good" cognitive load and is akin to the concept of "desirable difficulties" or effortful learning strategies that ultimately enhance learning (Bjork & Bjork, 2020). Accordingly, this theory suggests a benefit to increasing difficulty in adaptive training systems to optimally increase workload instead of only aiming to reduce it (Bjork & Bjork, 2020; Metcalfe, 2011; Sweller, 2011; Vygotsky, 1987).

Mayer's (1997; 2024) Cognitive Theory of Multimedia Learning (CTML) assumes that learners engage in three core cognitive processes during learning: selecting, organizing, and integrating. Learners must *select* important or relevant information from either visual or auditory representations of content being taught, *organize* that information into a coherent mental representation of the concept, and then *integrate* this new information with prior knowledge to learn effectively. This process engenders active and intentional processing (Mayer, 1997), thus occupying a learner's mental resources, leading to a set of guidelines for instructional design. Mayer and Moreno (1998b) note that instructional system designers must manage attentional demands on the learner and reduce any unnecessary processing. CTML and CLT provide complementary perspectives on working memory and attentional constraints on learners and highlight system design as a mechanism for managing cognitive load. Adaptive training systems possess the unique capacity to manage mental demand imposed on learners due to their ability to assess learner states and adjust training in real-time (Durlach & Ray, 2011; Ma, et al., 2014; VanLehn, 2011). Demand can be assessed on several performance metrics such as response time, accuracy, and user ratings of difficulty or metacognitive predictions. Using such performance metrics, systems can create a representation of the learner's current level of competency in the domain or skill to be trained, informing the decision to adapt training to further challenge or remediate the learner.

When integrating feedback into the system, the learner's processing of feedback should be considered an additional cognitive event. Another fundamental assumption of CTML is that learners have separate processing channels for visual and auditory information and learning will suffer if either processing channel is overloaded (Baddeley, 2000; Mayer, 1997; 2024). This notion comports with other theories regarding dividing attention into multitasking scenarios. Wickens' (2008; 2021) Multiple Resource Theory (MRT) refers to *resource overlap* as the extent to which concurrent tasks require shared resources, or processing channels, with greater overlap leading to increased demand and potential for overload – ultimately resulting in reduced ability to effectively timeshare between tasks and hurting concurrent task performance. Mayer and Moreno's (1998a; 1998b) guidelines for multimedia instructional design suggest that instructional content must consider the modality of the task such that it does not require the same processing channel as the task itself. Fiorella et al. (2012) demonstrated this through a comparison of either auditory or text-based feedback presented concurrently during a primarily visual simulation-based training system. Findings from this study indicated that performance was superior when instructional content was presented auditorily during training. Importantly, this study showed that guidelines for the design of to-be-learned content outlined by Mayer and

Moreno (1998a; 1998b) extend to the feedback domain, where feedback requiring the same processing channel as the instructional content can diminish learning performance by overloading the learner (Johnson, et al. 2017).

In short, introducing feedback requires an inherent attention shift from the learning content to diagnosing one's own performance and integrating that information into their current mental model of the concept (Kluger & DeNisi, 1996; Wickens, 2008; 2021). Moreover, it should guide learners by informing how they interact with instructional content moving forward (Johnson et al., 2017; Johnson & Priest, 2014). This poses additional load on the learner, highlighting the value of considering tradeoffs between feedback interventions.

Feedback in Adaptive Training Systems

Shute (2008) defines feedback as information communicated to the learner with the intent to modify their thinking or behavior for the purpose of learning. That is, feedback can be a mechanism for supporting a learner's strategy when engaging with content. Following this broad definition, there are several considerations regarding the ways feedback can be communicated. The first is when to present feedback in learning or, more formally, the timing of feedback presentation. Second, the content of feedback represents the type and amount of information provided to a learner after their response (e.g., whether a simple score is provided or an explanation of a concept is included). A third important consideration is the modality through which feedback is delivered, such as visual, auditory, or multimodal formats. Although feedback can also vary in terms of the modality through which it is delivered (e.g., visual, auditory), the present work focuses primarily on timing and content, acknowledging that modality may interact with these factors. Although feedback is highly beneficial for learning, findings for the effectiveness of different feedback interventions are mixed (Johnson et al., 2017).

Timing of Feedback

In the context of adaptive training, feedback timing refers to when corrective feedback is presented to the learner and is typically referred to as either *immediate* (i.e., delivered when the learner makes an error) or *delayed* (i.e., delivered after a response, following an intervening period during which the learner completes additional test or study items; Johnson et al., 2017). However, immediate and delayed feedback are operationally defined inconsistently in the literature, particularly regarding the interval that distinguishes immediate from delayed presentation. In their review of research on the design of feedback for serious games, Johnson et al. (2017) noted how literature on best practices for the timing and presentation of feedback is sparse and findings are generally inconsistent, likely reflecting inconsistencies in operational definitions of immediate or delayed feedback (cf. Kulik & Kulik, 1988; Johnson et al., 2017; Metcalfe et al., 2009). For example, immediate feedback could refer to corrective information presented after a learner makes a selection or it could refer to timed intervals throughout training when feedback is presented. The same could be true of delayed feedback, as some studies have used it to refer to feedback presented after a training phase or test is completed (Guzman-Munoz & Johnson, 2007) while others have examined delayed feedback as being presented days later (Butler et al., 2007).

An early meta-analysis by Kulik and Kulik (1988) indicated that delayed feedback often produced better learning in laboratory studies, whereas immediate feedback resulted in better learning performance in classrooms. Evidence from a large-scale study by Fyfe et al. (2021) comparing immediate versus delayed feedback on Canvas courses across 38 classrooms found no effect of feedback timing on course assessment performance. However, the authors noted a non-significant, but underpowered, trend of delayed feedback outperforming immediate feedback when tests were retrieval based (e.g., fill-in-the-blank) and in courses where students had more opportunities for feedback (e.g., number of quizzes and/or quiz questions). Although these potential moderating effects of feedback timing did not reach statistical significance, their limited representation in the study, combined with inconsistencies in the broader literature, highlights the need for further investigation into the boundary conditions under which feedback timing is most effective.

Kluger and DeNisi (1996) discussed how presenting feedback immediately after an error occurs versus presenting it after each training block may affect the attentional resources of the learner. From a cognitive load perspective, immediate feedback may reduce the working memory demands of learning a new task, as learners do not have to maintain errors until the end of a task in order to integrate feedback (Mayer & Moreno, 1998; Bjork et al., 2013). Moreover, other findings suggest that immediate feedback is better for learning because it keeps mistakes from being incorrectly encoded into memory and ultimately facilitates the integration of concepts into a more accurate mental model (Bangert-Drowns et al., 1991; Mayer, 1997; Johnson & Priest; 2014). In contrast, conflicting theories suggest that immediate feedback can instead overload the learner by splitting attention between feedback and concurrent task performance, as outlined in the discussion of cognitive demands during learning (Fiorella et al., 2012; Johnson et al., 2017; Mayer, 1997). This instead indicates an advantage of delayed feedback, as it does not require time sharing between attentional resources (Wickens, 2008; 2021). Delayed feedback inherently integrates spaced presentation during learning, requiring learners to engage in retrieval processes to integrate and update information in memory (Johnson et al., 2017). This spacing effect introduces a desirable difficulty that can ultimately enhance learning outcomes by facilitating essential processing of to-be-learned information (Bjork, 2020; Mayer, 2024).

Support for delayed over immediate feedback was demonstrated by Guzman-Munoz and Johnson (2007), whereby learners were tasked with learning geographic locations of cities on a map and with feedback given immediately after making a selection of the city name or delayed until all requested cities had been selected. Learners given delayed feedback outperformed those who received immediate feedback on both immediate and delayed retention tests.

Metcalfe et al. (2009) argue that the benefits of delayed feedback may reflect a confound in many prior studies whereby the retention interval from the last presentation of feedback to test is shorter in delayed than immediate feedback conditions. In an initial study using grade school children learning vocabulary, Metcalfe et al. (2009) compared learning performance between immediate and delayed feedback when this confound was present relative to when they controlled for it. They found that delayed feedback produced better final test performance than immediate feedback, regardless of whether they controlled for the time from last feedback presentation to final test. However, in a subsequent study that asked college students to learn GRE-level vocabulary, Metcalfe et al. (2009) observed that immediate and delayed feedback conditions performed similarly at test when the retention interval was held constant. Despite the theoretical implications of delayed feedback being confounded with retention interval, the authors describe this as a practical advantage of delaying feedback, as it creates a shorter retention interval to the test. Accordingly, it may be beneficial to design systems in a way that delays feedback closer to testing.

Content of Feedback

In addition to the timing of feedback, additional consideration must be given to the content presented in feedback. Feedback can convey general to very detailed performance information. Johnson and Priest (2014) defined *outcome feedback* as informing learners of the accuracy of their response or performance, while *explanatory feedback,* also referred to as *process* feedback*,* is defined as providing the learner with an explanation as to why their response was accurate or inaccurate. That is, explanatory feedback aims to direct the learner why their response was correct or incorrect, and sometimes on ways to change their strategies to improve performance, while outcome feedback aims to make the learner aware of their performance. An example and comparison of each type of feedback is presented in Table 1. Outcome feedback is easy to interpret and requires little cognitive effort during training. However, explanatory feedback is often favored in the literature as it is more supportive of processes involved in selecting, organizing, and integrating information outlined in CTML (Mayer, 1997; Johnson & Priest, 2014; Johnson et al., 2017).

Table 1**.** Example Comparison of Outcome and Explanatory Feedback

Outcome Feedback	**Explanatory Feedback**
Problem: Solve for (x): (2x + 5 = 13) **Learner Response:** (x = 3) **Feedback:** *"Incorrect. The correct answer is (x = 4)."*	**Problem:** Solve for (x): (2x + 5 = 13) **Learner Response:** (x = 3) **Feedback:** *"Incorrect. To solve (2x + 5 = 13), subtract 5 from both sides to obtain (2x = 8), then divide by 2 to get (x = 4)."*

Note: Outcome feedback provides accuracy information only whereas explanatory feedback includes procedural information intended to support understanding.

Much prior work supports the use of explanatory feedback over outcome feedback (Bangert-Drowns et al., 1991; Schute, 2008). For example, Astwood et al. (2008) investigated how feedback type influenced decision making in a call-for-fire simulation game requiring participants to decide which targets to destroy first based on a set of prioritization criteria. Participants who were given explanatory feedback on how to improve their decisions outperformed participants who received outcome feedback (i.e., accuracy scores with no instruction on how to tailor their decision making or strategy). The researchers attributed this finding to explanatory feedback better informing learners about how they should be making decisions. Mayer and Johnson (2010) had participants learn about electric circuits in a game-based task. They received either auditory cues corresponding to a 'correct' or 'incorrect' response during training or received those auditory cues paired with an explanation as to why their response was correct or incorrect. Results indicated that those who received explanatory feedback performed better not only at test, but also during training and on a transfer task.

The effectiveness of explanatory feedback can sometimes be impacted by learner characteristics such as prior domain knowledge, spatial ability (Moreno & Mayer, 1999), gender (Landsberg et al., 2010), and prior domain knowledge (see Vandewaeter et al, 2011 for a review). Further, the feedback principle of multimedia learning, introduced by Johnson and Priest (2014), posits that novice learners benefit more from explanatory feedback than outcome-based feedback alone, whereas expert learners perform best when feedback reflects performance information. This principle aligns with the expertise reversal effect reported by Kalyuga and colleagues, which describes the finding that more instructional support helps novice learners but may impose undue extraneous load on expert learners (Kalyuga, 2009; Kalyuga, 2021; Lee et al., 2014). Despite evidence of individual differences in feedback effectiveness, there is limited evidence as to whether learners might be aware of how different forms of feedback differentially affect their learning.

Learner Preferences and Self-Regulated Learning

In cognitive psychology, the term *metacognition* describes people's ability to think about, monitor, and adjust their own thinking processes (Flavell, 1979; Nelson, 1996; Rhodes, 2019).

This ability allows individuals to reflect on their own learning to set goals and create strategies to further their learning. Metacognition is fundamental in self-regulated learning, as students' understanding of their own learning influences their learning behavior (cf. Rhodes & Castel, 2009). To adopt effective learning strategies, learners need an accurate representation of their own knowledge (Bjork et al., 2013; Dunlosky & Rawson, 2012; Nelson & Narens, 1990). However, people's judgements about their own learning are often flawed as they can be influenced by factors that are not representative of one's actual learning state (Clegg et al., 2024; Rhodes & Castel, 2008; see Rhodes, 2016, for a review). For example, learners generally prefer instructional techniques that feel easy, such as rehearsal or re-reading, rather than those that adequately challenge them and lead to better learning, such self-testing (Dunlosky & Rawson, 2012; Roediger & Karpicke, 2006).

Soderstrom and Bjork (2015) attribute this dissociation between the strategies learners prefer and those that will benefit retention to the metacognitive illusion that occurs when information feels accessible in the moment of studying. Due to this illusion, learners may overestimate their understanding of a newly learned concept. Bjork and Bjork's (1992) theory of disuse describes this disparity between learning and performance as reflecting retrieval strength rather than storage strength. Storage strength refers to the degree to which knowledge and procedures are integrated or entrenched with other memory representations and is typically increased through effortful practice from active learning techniques. Retrieval strength represents the current ease of access or activation of those memory representations given the learner's current mental or environmental cues. Passive learning may temporarily enhance retrieval strength, but such representations fade over time, diminishing the contribution to storage strength. This poses a problem for effective control over study as overconfidence in one's knowledge can lead to less study than needed (Dunlosky & Rawson, 2012). Conversely, a lack of confidence in one's own abilities has implications for motivation and disengagement from learning opportunities that would provide feedback and direction (Bjork et al., 2013; Metcalfe et al., 2025; Putnam, et al., 2016). To enhance the accuracy of metacognitive judgments, students need instances where they can gain insight into their own learning (Bjork et al., 2014). This notion has been well demonstrated in prior research, including in seminal work by Roediger and Karpicke (2010), where they found evidence that judgements of learning (i.e., predictions of the likelihood of recalling information on later test) are more accurate when learners employ better learning strategies (e.g., engaging in retrieval practice) where they can gain real insight into whether their knowledge is complete.

Taken together, this body of research suggests that learners' difficulties in regulating their own learning does not stem from lack of effort but reflects a lack of external information aiding them in assessing their own learning state. Feedback offers a mechanism for calibrating learning strategies by providing learners with task-relevant information that supports a more accurate assessment of their performance (Durlach & Spain, 2014). By highlighting discrepancies between perceived and actual performance, feedback can improve metacognitive accuracy and support more adaptive control of learning.

Preferences for Feedback

If feedback serves as a critical mechanism for improving metacognitive accuracy, an important question concerns how learners decide which forms of feedback they prefer and whether these preferences are guided by their metacognitive judgments. There is evidence that students and learners value feedback and view receiving feedback responses as beneficial to learning (Lefreve & Cox, 2016; 2017; Rowe & Wood, 2008). Most research on the intersection of metacognition and feedback examines the impact of feedback on learners' monitoring and control of their own learning (Bjork et al; 2014; Carolan et al., 2014; Kornell & Rhodes, 2013; Metcalfe et al., 2025; Metcalfe & Finn, 2008; Sitzman et al., 2016). However, the inverse is largely unaddressed: How do learners' perceptions of what they believe will benefit their learning influence their selection of instructional strategies, such as the choice of feedback?

Qualitative research has shown a strong preference for immediate feedback (Lefreve & Cox, 2016; 2017; Rowe & Wood, 2008). Immediate feedback is perceived as less effortful than delayed feedback, as delayed feedback may impose demands on working memory and require the learner to restudy or recall a prior answer. However, as previously noted, prior work has shown a benefit of delayed feedback for these very reasons, potentially reflecting a desirable difficulty for the learner (Butler et al., 2007; Guzman-Muñoz & Johnson, 2007; Kluger & DeNisi, 1996; Metcalfe et al., 2009), suggesting a disconnect in learner preferences versus performance (Soderstrom & Bjork, 2015). In regard to the content of feedback, learners often value explanatory feedback over simple outcome feedback (Lefreve and Cox, 2016; 2017; Rowe & Wood, 2008). However, learners also indicate that feedback should be succinct and to-the-point. For example, Lefevre and Cox (2017) reported that learners indicated a preference for explanatory feedback but, when given the choice of which type of feedback to view, they most frequently elected to view outcome feedback. While these findings provide some indication that learners' subjective preferences do not match their behavior nor their current learning state, they highlight a need for further investigation.

Pridemore and Klein (1991) directly examined learner control over feedback in a computerized training system by allowing learners to choose whether to view feedback related to their response accuracy during lessons on psychological measurement. In a between-subjects design, learners completed lessons on psychological measurement and received either outcome (correct-answer) feedback or explanatory feedback. This manipulation was crossed with learner control, with some participants assigned to a participant-controlled condition in which they decided whether feedback was displayed following an error and others to a system-controlled condition in which the system determined when feedback was presented. Although the authors found an overall benefit of explanatory feedback, they observed no differences in learning outcomes between learner-controlled and system-controlled feedback conditions. However, because feedback type was manipulated between subjects, this design does not support conclusions about whether learners' preferences for feedback content—or their metacognitive judgments about its utility—translate into improved learning outcomes when they are given agency to decide on the type of feedback they receive.

To test whether learning outcomes are superior when the student has input into their feedback presentation, we must examine whether differences in learning outcomes exist when their preferences are both honored and dishonored (Pashler et al, 2008). Prior research on learner preferences and metacognitive judgments in self-directed learning has shown that this honor–dishonor methodology provides insight into whether discrepancies exist between what learners believe will best support their learning and what actually produces optimal learning outcomes (Kornell & Metcalfe, 2006). Simply gauging learners' subjective preferences limits conclusions to what learners think will help them learn best rather than providing causal links to better learning outcomes. In contrast, the honor-dishonor methodology allows a clear comparison of how metacognitive control impacts learning. That is, if dishonoring a learner's choice leads to better retention, it suggests that learners' self-directed choices may not be optimal, revealing gaps in monitoring and control processes (Kornell & Metcalfe, 2006; Rhodes et al., 2013).

In a series of experiments Kornell and Metcalfe (2006) had participants choose whether to restudy semantic or paired-associate items before test and the researchers either honored, partially honored (granted restudy for half of the items selected), or dishonored those choices. Their first experiment showed that learning was best when choices to restudy items were honored. A second experiment showed that this effect was more pronounced when participants had additional exposure to the content being learned, showing an effect of expertise driving the accuracy of these metacognitive judgements. Finally, in their third experiment, participants were required to choose which items to restudy from a set of items they had not yet learned. Results showed that participants prioritized easier items over more difficult ones and performance was superior when those choices were honored. In contrast, when participants' choices to restudy the most difficult items were honored, test performance suffered more than when those choices were dishonored, providing direct evidence that learner preferences can be maladaptive and that honor–dishonor manipulations are necessary to distinguish beneficial from harmful metacognitive control strategies. As metacognitive judgements have been shown to have some influence over study control (Dunlosky et al., 2013; Nelson & Narens, 1990; Rhodes, 2016; Rhodes et al., 2013), it follows that learners' awareness of their own learning processes could extend to how they engage with feedback. That is, if learners have insight into the feedback presentation strategies that will work best for them, honoring their preferences should lead to better retention than dishonoring them (Kornell & Metcalfe, 2006).

Gaps in the Literature

The literature surrounding the utility of feedback for learning – whether in traditional or computerized learning environments – is abundant (Durlach, 2019; Johnson et al., 2017). However, as noted, there are mixed findings on how the design of feedback influences learning.

These mixed findings on feedback timing suggest other moderating factors may operate. An that underexplored factor is the role of learner preferences in feedback utility, with some prior work indicating that learners express a strong preference for immediate feedback even when delayed feedback has been shown to enhance retention (Guzman-Muñoz, 2007; Johnson et al., 2017; Lefevre & Cox, 2016; Rowe & Wood, 2008). Similar discrepancies are present in research regarding the content of feedback. Learners frequently report preferring explanatory feedback yet, when given a choice, they may select simpler outcome feedback (Lefevre & Cox, 2016). Taken together, such data reveal a clear discrepancy between learners' reported preferences, their actual decisions, and what promotes effective learning. This gap suggests that feedback interventions in adaptive training systems may benefit from incorporating partial learner regulation—allowing input into timing and type—while balancing empirical guidance on effective learning strategies.

Although adaptive training systems increasingly personalize instruction, they often rely on system-driven predictions rather than learner choice. Feedback, as a central instructional component, may benefit from preference-based tailoring. However, empirical evidence on how learner preferences interact with feedback timing and content remains sparse – particularly regarding the performance differences that occur when preferences are honored or dishonored. Understanding how learner preferences interact with metacognitive judgments and feedback design is therefore critical for creating systems that maximize both learning outcomes and user satisfaction.

Present Experiments

Across three experiments, I addressed these gaps by first investigating whether learners show a preference for immediate or delayed feedback (Experiment 1) and whether the preference for immediate feedback reported by Lefevre and Cox (2016) studies holds true. Next, I extended their findings by investigating whether accommodating that preference leads to better learning outcomes and more favorable perceptions of training (Experiment 2), and lastly whether the same findings extend to accommodating user preferences for the content of feedback (Experiment 3). Findings from these experiments could address recommendations by Durlach (2019) that incorporating user-preferred design features into instructional systems could benefit performance outcomes.

To empirically test these research questions, participants learned concepts from an introductory psychology textbook (Myers & DeWall, 2018) and were trained on the concepts covered in each paragraph through multiple-choice questions adapted from a previous study (Bates et al., 2025). Participants were first introduced to the two feedback conditions they would be experiencing, using standardized instructions and visual depictions, after which they reported their initial feedback preference. Participants then completed two learning blocks in which they studied short psychology-related passages, answered comprehension questions, and received either immediate feedback (following each response) or delayed feedback (presented after completing all questions in a block). Finally, they were tested on their retention after a short delay, and a 2-day delay in Experiments 2 and 3.

Primary Research Questions and Hypotheses

Experiment 1: Do participants show a preference for feedback timing?

- H1: Participants will show a preference for immediate feedback over delayed feedback.

Experiment 2: Does providing learners with a choice for immediate feedback or delayed feedback impact their learning outcomes?

- H2: Learners will show an overall preference for immediate feedback.

- H3: Delayed feedback will yield stronger learning outcomes than immediate feedback, particularly at delayed tests.
- H4: Matching learner preference (immediate vs. delayed) will enhance learning, reduce workload, and improve satisfaction compared to mismatching.

Experiment 3: Does providing learners with a choice for the context of their feedback, either explanatory or outcome feedback, show similar effects to matching their preference for timing of feedback?

- H5: Learners will show an overall preference for explanatory feedback over outcome feedback.
- H6: The effectiveness of feedback in learning will depend on feedback content, such that explanatory feedback will lead to stronger learning gains than outcome feedback, particularly at delayed tests.
- H7: Matching learner preference (explanatory vs. outcome) will enhance learning, reduce workload, and improve satisfaction compared to mismatching.

CHAPTER 2 – EXPERIMENT 1

Previous research has shown a strong preference for immediate over delayed feedback (Lefreve & Cox, 2016; 2017; Rowe & Wood, 2008). Despite findings from qualitative studies regarding learner preferences, empirical work typically shows enhanced benefits for retention and learning from delaying feedback presentation (Guzman-Muñoz, 2007; Johnson et al., 2017; Lefevre & Cox, 2016; Rowe & Wood, 2008). The source of the apparent benefit of delayed feedback is up for debate. Some research indicates that it reflects a shorter retention interval between the final presentation of feedback and the test relative to immediate feedback conditions whereas alternative accounts suggest that delayed feedback enhances learning by introducing desirable difficulties or increasing germane cognitive load (Butler et al., 2007; Guzman-Muñoz & Johnson, 2007; Kluger & DeNisi, 1996; Metcalfe et al., 2009; Sweller, 2011).

Immediate feedback is often experienced as less cognitively demanding, as delayed feedback requires the learner to engage in effortful retrieval to integrate feedback into previously held information (Mayer & Moreno, 1998; Bjork et al., 2013). Research on metacognitive monitoring and control indicates that learners tend to favor strategies that feel easier over those that require greater effort (Dunlosky & Rawson, 2012; Roediger & Karpicke, 2006; Soderstrom & Bjork, 2015). Therefore, learners may prefer the less cognitively demanding feedback, as the discrepancy between learners' preference for immediate feedback and the observed benefits of delayed feedback may reflect differences in perceived cognitive effort or subjective perceptions of one's own learning performance.

The goal of Experiment 1 was twofold. The first goal was to determine whether participants showed a preference for immediate or delayed feedback within the context of the present task. Second, Experiment 1 investigated the subjective perceptions underlying these preferences, focusing on whether perceived workload and judgments of learning during training were sensitive to feedback timing. Accordingly, the current study solicited preferences before the experiment begins to gauge pre-existing subjective preferences for feedback timing as well as a post-experiment preference indication of whether preference changed after experiencing the task. Thus, preference was assessed after describing feedback timing and following exposure to the learning task.

Method

Participants

Eighty-one participants (Cis-men = 47, Cis-women = 32, Nonbinary = 1, Prefer not to say = 1) were recruited online via Prolific and compensated $7.50 for their participation in Experiment 1. The sample ranged between 20 to 61 years old (M = 32.46, SD = 9.47). The experiment took an average of 33.01 (SD = 5.13) minutes to complete. All participants were located in the United States, all were fluent English speakers, and all reported normal or corrected to normal vision. The sample size was determined using an a priori power analysis for a paired-samples t-test examining differences in learning outcomes between immediate and delayed feedback conditions. The analysis indicated that 80 participants would be required to detect an effect size of d = .40 with 80% power, assuming an alpha level of .05 and a two-tailed test.

Study Design

This study manipulated feedback timing (Immediate vs Delayed) within-subjects.

Materials

Instructional Content. Participants studied paragraphs relating to different concepts from a college level Introduction to Psychology textbook (Myers & DeWall, 2018) and were trained on the concepts covered in each paragraph via multiple-choice questions adapted from Bates et al. (2025). Instructional content consisted of 6 instructional passages on psychology-related concepts (see Appendix A; Experiment 1 used passages 1-6). Passages were designed to be comparable in length and vocabulary. Each paragraph was paired with two corresponding study questions assessing comprehension of the material; these study questions were also used as posttest items to assess retention of concepts (see Appendix B). In addition, two transfer test items per paragraph were developed to assess participants' ability to apply the learned concepts to novel but related contexts (see Appendix C).

Feedback Instructions and Examples. Participants were instructed on the difference between immediate and delayed feedback and given a visual depiction representing these differences, before being asked to indicate their initial preference for feedback timing (see Figure 1). They were instructed that they would experience both types of feedback during the study and that their selection would not impact the way feedback was presented to them.

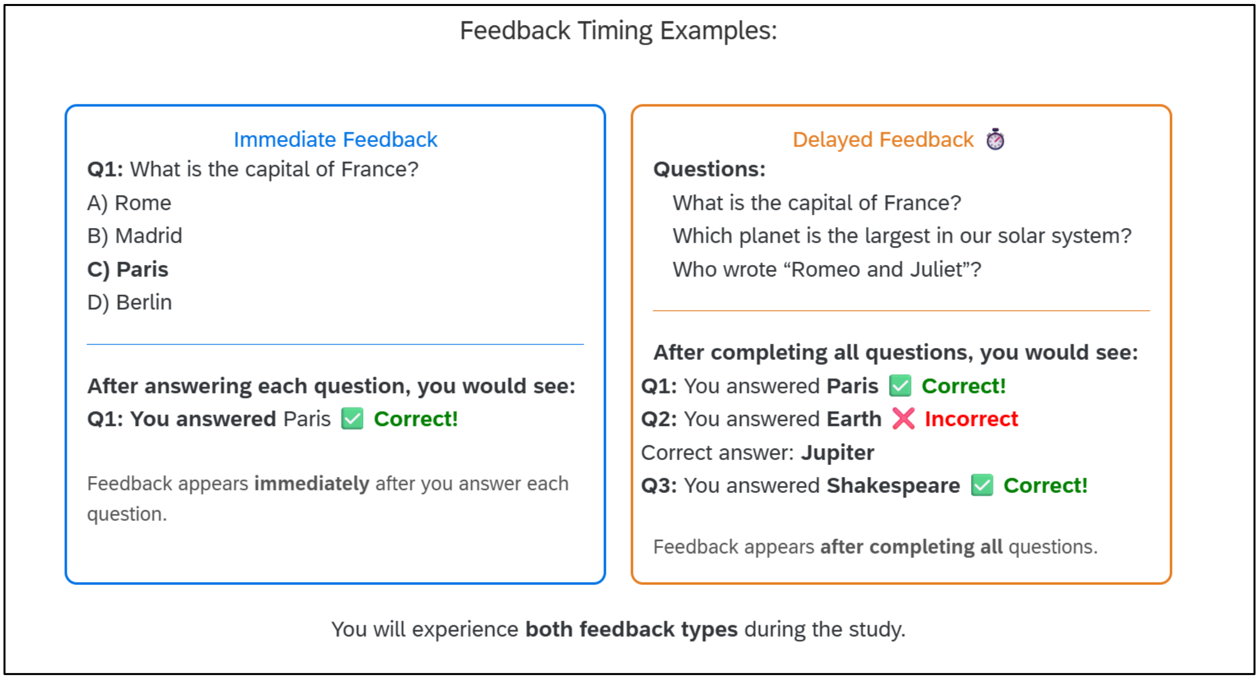

Figure 1. Visual Depiction representing immediate versus delayed feedback shown to participants. This content is unrelated to the content participants are instructed on.

Feedback messages. Participants were given outcome-based feedback regarding the accuracy of their response to each question. In the immediate feedback condition, participants were presented with feedback after each response to the study questions. In the delayed feedback condition, participants were presented with feedback to all items, one at a time, after they completed responses to all study questions. Feedback messages were formatted consistently across items and feedback timing was held constant. Feedback messages restated the question, indicated whether their input was correct or incorrect, and provided the correct answer if incorrect. Visual examples of feedback messages are provided in Figure 2.

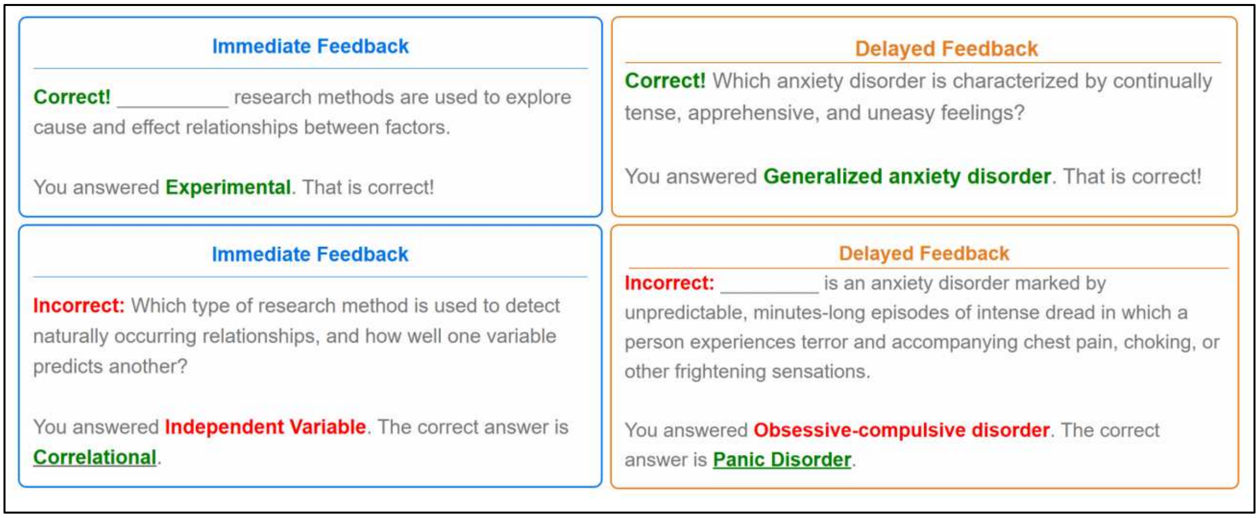

Figure 2. Examples of feedback messages for both immediate and delayed conditions.

Judgements of Learning (JOLs). Participants made Judgments of learning (JOLs) to indicate perceptions of their own learning. After each study phase, participants estimated how many questions they would answer correctly (out of 6 possible) on the upcoming posttest. JOL prompts are provided in Appendix D.

NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). Subjective workload was assessed using the NASA Task Load Index, which is a widely used multidimensional measure of perceived workload. The NASA-TLX evaluated workload across five subscales in the present study: mental demand, temporal demand, effort, frustration, and perceived performance (the physical demand subscale was not included, as it was not relevant to the computerized task used in the present experiments). Participants rated each dimension using a 20-point scale, with higher values indicating greater perceived demand (or, for the performance subscale, greater perceived success). The NASA-TLX was administered following each set of study blocks to assess participants' subjective experience of workload under different feedback conditions. Appendix E includes all subscales used.

Distractor Task. Participants completed a brief distractor task consisting of arithmetic problems. Each problem required a combination of addition, subtraction, and multiplication operations (e.g., $(2 + 7) \times 9 - 4$). Participants were given 20 seconds to solve each problem before automatically advancing to the next. The complete set of distractor task items is presented in Appendix F.

Perceptions of Training Questionnaire. To assess participants' general satisfaction with the learning content provided in the experiment, participants rated their overall perceptions of the training on a 7-point Likert scale, ranging from 1-"Strongly Disagree" to 7-"Strongly Agree". Items on the scale are included in Appendix G.

Procedure

Figure 3 presents a flowchart of the experimental procedures for Experiment 1. After completing the consent process, participants were introduced to the distinction between immediate and delayed feedback through four practice trivia items presented in the same format as the experimental stimuli but using unrelated content. Participants received immediate feedback for two items and delayed feedback for the remaining two. Following this practice, they were shown a visual depiction illustrating the difference between feedback types, accompanied by definitions of each (see Figure 1). Participants then indicated their initial preference for feedback timing during learning. They were informed that both types of feedback would be administered during the study and that their stated preference would not influence how the task was presented.

Participants then entered a two-block experimental phase. Each block began with a three-minute reading period, during which participants read three of the learning passages. Once three minutes had elapsed, they progressed to the study phase where they responded, at their own pace, to 6 study questions (two related to each learning passage). Feedback timing condition was manipulated between blocks, such that participants received immediate or delayed feedback for all 6 study questions in a single block. In the immediate feedback condition, participants were shown a single feedback message following each response, which remained on screen for 15 seconds before automatically advancing to the next study question. In the delayed feedback condition, feedback was presented only after participants completed all six study questions within a block. Each feedback message was displayed individually for 15 seconds.

After completing the study questions in each block, participants were prompted to provide their JOL and then completed the NASA-TLX. Participants then proceeded to the three-minute distractor task consisting of math problems, with 20 seconds to respond to each question before the task auto-advanced. Next, they completed a posttest consisting of the same 6 questions they responded to during study. Finally, they ended each block with a transfer test consisting of 6 novel questions that were not identical to those in the study phase but addressed the same concepts covered to assess transfer of training.

Feedback timing was counterbalanced across blocks, as well as across concepts, to control order effects, such that half of the participants receive immediate feedback in the first block and delayed feedback in the second block, and the other half receive the reverse order. After completing both learning blocks, participants again viewed the visual depicting feedback timing and indicated their final feedback preference. Finally, participants completed the Training Perceptions Exit Survey and a short demographics form.

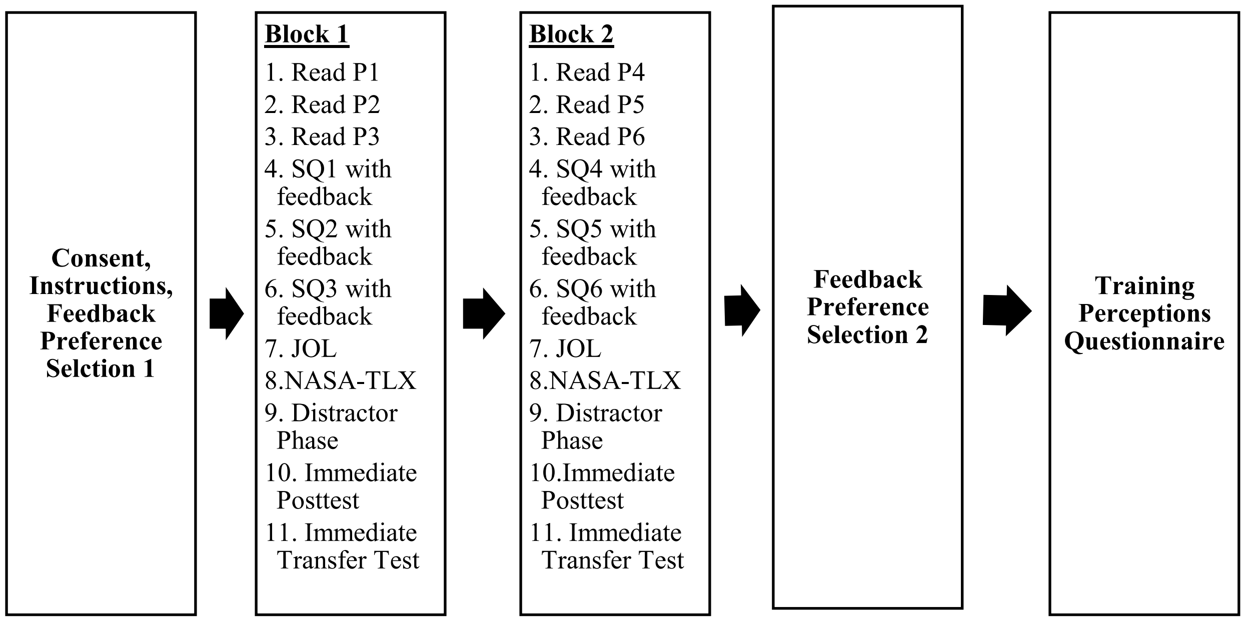

Figure 3. Flowchart of experimental procedures for Experiment 1.P = Passage (Appendix A); P SQ = Study Question (Appendix B); JOL = Judgements of Learning (Appendix D)

Results

Feedback Preference Selection

To assess whether the proportion of participants preferring immediate feedback differed significantly from chance (0.5), a binomial test was conducted for both their initial preference indication as well as their preference selection after completing all trials. For the initial selection, 93.83% (n = 76, N = 81) of participants indicated a preference for immediate feedback. In their final selection, 83.95% (n = 68, N = 81) of participants indicated a preference for immediate feedback. Significantly more participants indicated a preference for immediate over delayed feedback both at their initial judgement and final judgement (p < .001).

Figure 4 provides a visual representation of the percentage of participants who indicated a preference for either type of feedback across both timepoints. Eight participants adjusted their preference from immediate feedback at initial selection to delayed feedback upon completion of the experiment. Only 1 participant, who initially indicated a preference for delayed feedback, indicated a change in preference to immediate feedback at the final timepoint. To examine changes in feedback preference over time, an exact McNemar's test was conducted. This test is designed for paired categorical data and assesses whether the proportion of participants preferring immediate versus delayed feedback differs across timepoints, with the exact version used to account for small or uneven cell frequencies. Results of McNemar's indicated a significant change in preference (p = .021), such that participants were more likely to switch from immediate to delayed feedback than vice versa (OR = 9.09, 95% CI [1.25, 333.33]).

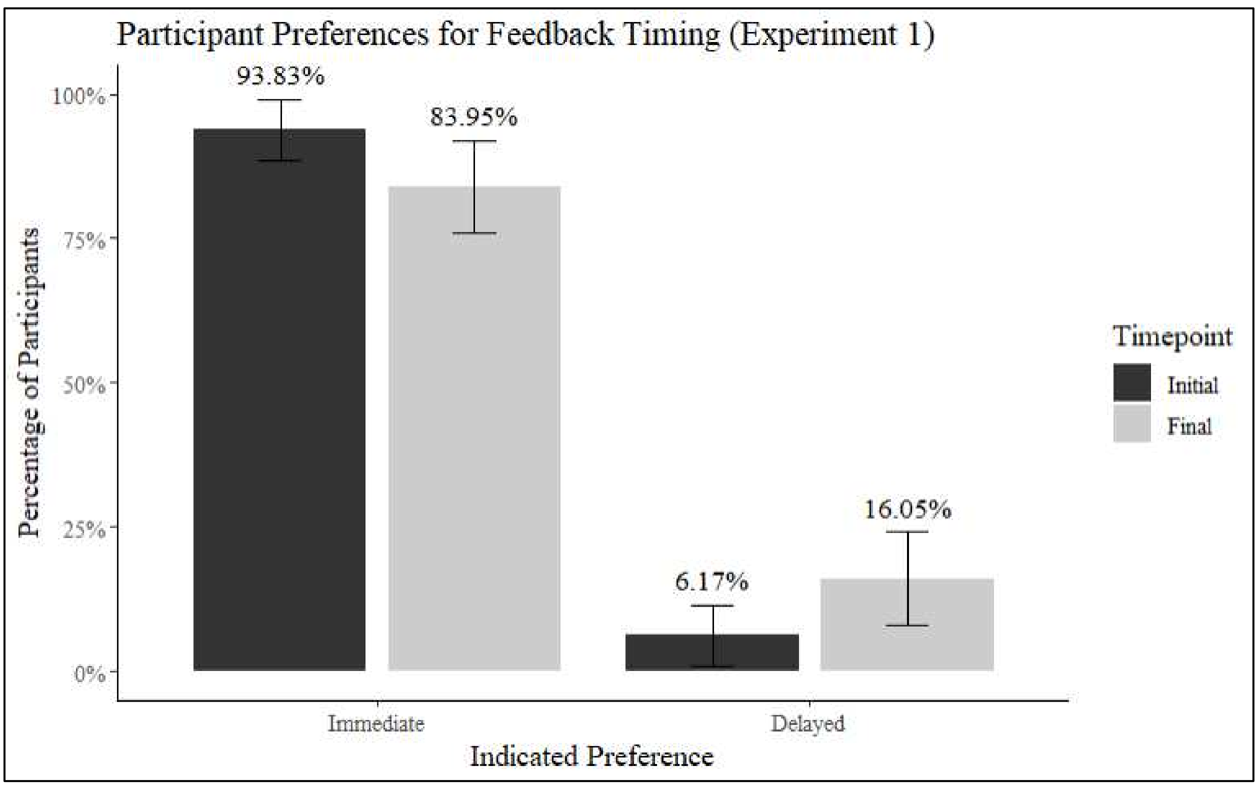

Figure 4. Percentages of participants who indicated a preference for immediate versus delayed feedback across time points for Experiment 1. Error bars represent 95% confidence intervals for associated percentages.

Given the small number of participants who indicated a preference for delayed feedback at either timepoint, inferential comparisons were not conducted for this subgroup. That is, inferential analyses were restricted to participants who indicated a preference for immediate feedback and paired-samples *t*-tests were used to compare accuracy across preference-honoring conditions within this group. Performance data for the delayed-preference group is reported descriptively. For all performance measures, accuracy is reported as proportion of items correct (i.e., number of correct responses divided by total trials).

Study Question Accuracy

Overall Accuracy. Study question accuracy was tallied for both immediate and delayed feedback blocks for each participant, regardless of indicated preference. Average accuracy during study was similar across immediate (M = 0.82, SD = 0.18) and delayed (M = 0.80, SD = 0.21) blocks. A paired samples t-test revealed no significant difference in average study accuracy by feedback timing condition, $t_{(80)} = 0.81$, $p = .417$, $d = 0.09$.

Accuracy as a Function of Feedback Timing Preference. Group means and standard deviations are reported in Table 2. As can be seen, participants who preferred immediate feedback scored similarly on study items when their preference was honored versus when it was dishonored, $t_{(75)} = 0.62$, $p = .540$, $d = 0.07$.

Table 2. Experiment 1: Study Question Accuracy as a Function of Feedback Preference

Feedback Condition	Preference Condition	*M* (*SD*)
Immediate	Honored	0.81 (0.18)
Immediate	Dishonored	0.80 (0.21)
Delayed	*Honored*	*0.83 (0.17)*
Delayed	*Dishonored*	*0.90 (0.09)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred immediate feedback due to the small number of participants indicating a delayed feedback preference.

Posttest Accuracy

Overall Accuracy. Regardless of indicated preference, participants responded accurately to slightly more posttest items, on average, for items subject to delayed feedback (*M* = 0.92, *SD* = 0.15) than immediate feedback (*M* = 0.89, *SD* = 0.14). A paired samples *t*-test indicated that this difference failed to meet the threshold for significance, $t_{(80)} = 1.75$, $p = .083$, $d = 0.19$.

Accuracy as a Function of Timing Preference. Group means and standard deviations are reported in Table 3. Participants who preferred immediate feedback scored descriptively higher at posttest when their preferences were dishonored rather than honored, although this difference did not reach statistical significance, $t_{(75)} = 1.92$, $p = .058$, $d = 0.22$.

Table 3. Experiment 1: Posttest Accuracy as a Function of Feedback Timing Preference

Feedback Condition	Preference Condition	*M* (*SD*)
Immediate	Honored	0.89 (0.14)
Immediate	Dishonored	0.93 (0.15)
Delayed	*Honored*	*0.87 (0.14)*
Delayed	*Dishonored*	*0.90 (0.15)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred immediate feedback due to the small number of participants indicating a delayed feedback preference.

Transfer Test Accuracy

Overall Accuracy. A paired samples *t*-test comparing transfer test accuracy by feedback timing condition found that accuracy was significantly higher following delayed feedback (*M* = 0.78, *SD* = 0.22) than immediate feedback (*M* = 0.70, *SD* = 0.26), $t_{(80)} = 2.15$, $p = .034$, $d = 0.24$.

Accuracy as a Function of Feedback Timing Preference. Group means and standard deviations are reported in Table 4. Participants who indicated a preference for immediate feedback scored significantly lower on the transfer test for items where their preference was honored than when it was dishonored, $t_{(75)} = 2.31$, $p = .024$, $d = .26$.

Table 4. Experiment 1: Transfer Test Accuracy as a Function of Feedback Preference Across Study Blocks

Feedback Condition	Preference Condition	*M* (*SD*)
Immediate	Honored	0.69 (0.25)
Immediate	Dishonored	0.78 (0.23)
Delayed	*Honored*	*0.80 (0.22)*
Delayed	*Dishonored*	*0.83 (0.29)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred immediate feedback due to the small number of participants indicating a delayed feedback preference.

Judgements of Learning (JOLs)

JOL. Participants' JOLs reflected their predicted proportion correct on the posttest (i.e., the proportion of the six posttest questions they expected to answer correctly) for each feedback condition. A paired samples *t*-test comparing JOL ratings by feedback condition indicated that, on average, JOLs were significantly higher during immediate feedback ($M = 0.96$, $SD = 0.13$) than delayed feedback ($M = 0.85$, $SD = 0.18$), $t_{(80)} = 5.88$, $p < .001$, $d = 0.65$.

An additional paired samples *t*-test was used to assess the absolute accuracy of participants' JOLs relative to their actual performance (see Figure 5). Participants predicted they would correctly answer a greater proportion of posttest questions when given immediate feedback ($M = 0.96$, $SD = 0.13$) than they answered correctly at posttest ($M = 0.89$, $SD = 0.15$), $t_{(80)} = 3.61$, $p < .001$, $d = 0.17$. The opposite pattern was observed in the delayed feedback condition, as participants predicted they would correctly answer a lower proportion of posttest questions ($M = 0.85$, $SD = 0.18$) than they ultimately answered correctly at posttest ($M = 0.92$,

$SD = 0.15$), $t_{(80)} = 3.20$, $p = .002$, $d = 0.36$.

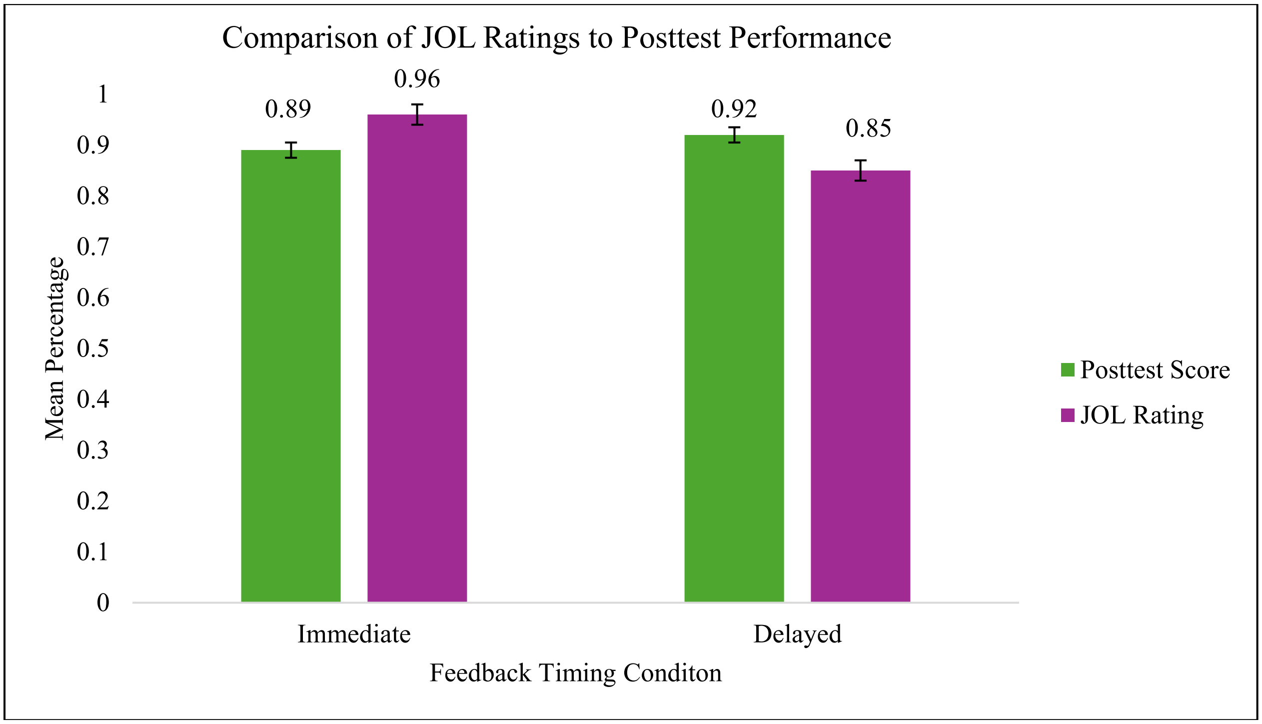

Figure 5. Visual comparison of participants' judgments of learning (JOLs) and posttest performance. JOLs reflect predicted proportion correct at posttest, whereas posttest scores reflect actual performance.

JOLs as a Function of Feedback Timing Preference. Group means are reported in Table 5. Participants who indicated a preference for immediate feedback predicted that they would correctly answer a significantly higher proportion of posttest items when their preference was honored than when it was dishonored, $t_{(75)} = 5.68$, $p < .001$, $d = 0.65$.

Table 5. Experiment 1: Participant JOL as a Function of Feedback Preference

Feedback Condition	Preference Condition	*M* (*SD*)
Immediate	Honored	0.96 (0.14)
Immediate	Dishonored	0.85 (0.18)
Delayed	*Honored*	*0.87 (0.22)*
Delayed	*Dishonored*	*1 (0)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred immediate feedback due to the small number of participants indicating a delayed feedback preference. The standard deviation for the delayed-dishonored condition is 0 because all participants in the delayed-preference group ($n = 5$) predicted perfect performance.

Perceived Workload (NASA-TLX)

Overall Workload. NASA-TLX ratings were averaged across subscales for both the immediate and delayed feedback blocks. Descriptive statistics and inferential comparisons between feedback timing conditions for NASA-TLX subscales for all experiments are presented in Appendix H (Table H1). To assess the impact of feedback timing on perceptions of workload during training, ratings for each subscale were compared by feedback condition via paired-samples *t*-tests. No significant effect of feedback timing was found on any of the five subscales.

Workload as a Function of Feedback Timing Preference. Average ratings on the NASA-TLX subscales were also examined as a function of participants' feedback timing preference to assess whether honoring learner preferences influenced subjective workload. Descriptive statistics for both the immediate and delayed preference groups, as well as inferential comparisons between preference-honored and preference-dishonored conditions for the immediate preference group, are reported in Appendix I (Tables I1–I2). No significant differences were observed across any of the five NASA-TLX subscales within the immediate-preference group as a function of whether participants' feedback timing preference was honored or dishonored.

Perceptions of Training

Descriptive statistics for all perceptions of training measures are provided in Appendix J (Table J1).

Discussion

The primary goal of Experiment 1 was to replicate prior findings demonstrating that learners generally prefer immediate feedback over delayed feedback (Lefevre & Cox, 2016, 2017; Rowe & Wood, 2008). An overwhelming majority of participants (93.83%) indicated a preference for receiving feedback immediately after responding to each item, consistent with hypothesis (H1). Interestingly, this preference decreased slightly from the initial preference assessment, which was based on descriptions and examples of the feedback conditions, to the final preference assessment following experience with the learning task itself (83.95%). Additionally, participants were significantly more likely to switch their preference from immediate feedback to delayed feedback after experiencing both feedback timing conditions. This pattern suggests that learners' initial feedback preferences may reflect assumptions about what will support learning, whereas direct experience with the task and feedback conditions may lead learners to reevaluate those preferences. Therefore, learners' feedback preferences may become more informed once they have firsthand experience with how different feedback timings affect learning. Indeed, learners need an accurate mental model of their own knowledge to correctly judge what strategies will be most effective for them (Bjork et al., 2013; Dunlosky & Rawson, 2012; Nelson & Narens, 1990). After being provided with feedback during the actual task, it is possible that they have a more informed representation of which timing condition best supports their learning (Durlach & Spain, 2014).

The second goal of Experiment 1 was to investigate subjective perceptions underlying feedback timing preferences. Participants' JOLs indicated greater confidence in future performance during immediate feedback conditions relative to delayed feedback conditions. However, these perceptions were not aligned with actual learning outcomes. Participants overestimated their performance following immediate feedback, predicting greater posttest accuracy than they ultimately achieved. In contrast, participants underestimated their performance following delayed feedback, predicting lower posttest accuracy than they ultimately demonstrated. These findings suggest that immediate feedback may create inflated perceptions of learning, whereas delayed feedback may feel less effective during learning despite producing stronger transfer performance. Moreover, participants who preferred immediate feedback predicted greater posttest accuracy when that preference was honored relative to when it was dishonored. Because no training accuracy differences emerged between timing conditions, these findings indicate that feedback timing influenced participants' perceptions of learning independent of actual performance during training.

One explanation frequently proposed for the benefits of delayed feedback is that delaying feedback imposes additional working memory demands that ultimately support stronger long-term learning (Bjork, 2020; Bjork et al., 2013; Mayer, 2024; Mayer & Moreno, 1998). However, analyses of subjective workload using the NASA-TLX did not indicate meaningful differences in perceived workload between feedback timing conditions. Thus, learners may not be consciously sensitive to the additional cognitive demands imposed by varying feedback timing and, consequently, may not base their feedback preferences on perceived workload differences.

Although the relatively small number of participants preferring delayed feedback limit inferential comparisons involving this group, the predominance of immediate feedback preferences allows comparisons within the immediate-preference group to be interpreted as broadly representative of the larger sample. Experiment 1 also provided preliminary insight into how honoring or dishonoring learners' feedback timing preferences influenced learning outcomes. No overall differences in posttest performance emerged between immediate and delayed feedback conditions, and this pattern remained when examining whether participants' preferences for immediate feedback were honored or dishonored. However, delayed feedback produced significantly higher transfer test performance than immediate feedback. Notably, participants who preferred immediate feedback demonstrated greater transfer performance when that preference was dishonored. These findings are consistent with prior work suggesting that learner preferences do not necessarily align with the instructional conditions that best support learning (Clegg et al., 2024; Rhodes & Castel, 2008; see Rhodes, 2016, for a review). Taken together, these results suggest that learners may not be fully metacognitively aware of how feedback timing influences their learning outcomes.

Although Experiment 1 provides important insight into how learners form and evaluate feedback timing preferences, several limitations should be acknowledged. Learner performance outcomes were measured, but a relatively short retention interval was used, limiting the conclusions that can be drawn regarding long term effects on retention (Schmidt & Bjork, 1992). Additionally, learners were informed that their preference would not have an impact on how the task was presented to them and that they would be experiencing both feedback timing conditions. Therefore, preferences were not necessarily honored or dishonored. Because of this, analyses comparing preference honoring versus dishonoring should be interpreted as preliminary.

To address these limitations, Experiment 2 was designed to more directly examine how honoring or dishonoring learner feedback preferences influences both subjective perceptions of learning and objective learning outcomes after learners had gained experience with the task itself. In addition, Experiment 2 incorporated a longer retention interval to better assess the extent to which feedback timing influences longer-term retention and transfer performance. By manipulating whether learners' experienced preferences were honored or dishonored, Experiment 2 allows for a clearer examination of when learner preferences support learning and when they may instead reflect inaccurate metacognitive beliefs about effective instruction.

CHAPTER 3 – EXPERIMENT 2

Experiment 1 replicated prior findings from Laverne and Cox (2016) demonstrating that learners exhibit a clear preference for the timing of feedback they receive, with most indicating a preference for immediate over delayed feedback. These preferences were also predictive of learners' subjective preferences of their own learning, providing preliminary evidence that learners may be demonstrating a mismatch between what feels easy and what is best for their learning. Building on these findings, the primary goal of Experiment 2 was to examine whether learning outcomes improve when learners are given input into how their feedback is presented.

Although studying learners' preferences and subjective judgements is informative, these measures only capture what learners believe to be effective rather than what is actually beneficial, which may diverge from what is most beneficial for learning (Bjork & Bjork, 1992; Rhodes & Castel, 2008; see Rhodes, 2016; Soderstrom & Bjork, 2015). To determine whether these preferences correspond to optimal learning outcomes, it is necessary to compare performance under conditions in which preferences are either honored or dishonored (Kornell & Metcalfe, 2006; Pashler et al, 2008). With respect to feedback timing, if learners believe that immediate feedback best supports their learning, then honoring this preference should result in superior performance under immediate, relative to delayed, feedback conditions.

Empirical work directly examining the effects of honoring versus dishonoring feedback timing preferences is limited. More broadly, research on learner control over feedback presentation is sparse and indicates that stated preferences may not always align with behavioral choices during learning (Lefevre & Cox, 2016). However, preliminary analyses from Experiment 1 suggested that performance during learning did not consistently differ as a function of stated preference. Drawing on the broader honor–dishonor literature, Kornell and Metcalfe (2006) demonstrated that honoring learners' decisions about which items to restudy enhanced subsequent memory and test performance, particularly when learners had sufficient exposure to the material. This finding suggests that accurate metacognitive control depends, in part, on learners' ability to assess their own mastery of the content (Bjork et al., 2013; Dunlosky & Rawson, 2012; Durlach & Spain, 2014; Nelson & Narens, 1990). Consistent with this assertion, Experiment 1 revealed that a significant number of participants changed their initial feedback timing preference from pre-to-post experimental session.

Accordingly, Experiment 2 gave participants exposure to the learning materials prior to making preference-based decisions that influence feedback presentation during training. Participants' decisions to continue using immediate or delayed feedback after exposure to task stimuli were either honored or dishonored, allowing a clear comparison of how metacognitive control impacts learning. That is, if dishonoring a learner's choice leads to better retention, it suggests that learners' self-directed choices may not be optimal, revealing gaps in monitoring and control processes (Kornell & Metcalfe, 2006; Rhodes et al., 2013). Moreover, the goal of refining feedback design should be to promote long term retention (Kluger & DeNisi, 1996; Shute, 2008). Evidence suggests that the effects of instructional inventions can be inflated when assessed in the same session as acquisition or training of the material, and longer retention intervals are more indicative of the actual impact of instructional interventions on learning (Schmidt & Bjork, 1992). Therefore, participants in Experiment 2 also completed a second assessment of their learning two days after the first.

Method

Participants

The final sample included in the analysis was approximately 114 participants (Cis-men = 53, Cis-women = 53, Trans-women = 1, Nonbinary = 3, Agender = 1, Prefer not to say = 3). All participants were located in the United States, all were fluent English speakers, and all reported normal or corrected to normal vision. The sample ranged between 18 to 79 years old (M = 41.45, SD = 13.55). Participants were recruited through Prolific and compensated $15 for their participation in session 1 and an additional $7.50 for their participation in session 2. The first session of the experiment took an average of 58.66 (SD = 17.72) minutes to complete, and the second session of the experiment took an average of 20.31 (SD = 10.26) minutes to complete. On average, participants completed the second session 55.37 (SD = 8.56) hours after completing the first session.

An a priori power analysis for a repeated-measures ANOVA indicated that a total sample of 100 participants would be required to detect an effect size of f = .15 with 95% power, assuming an alpha level of .05 and a two-tailed test. To ensure sufficient power across conditions and account for possible attrition, 121 participants were recruited. To ensure data quality, participants whose total completion time fell 3 standard deviations above or below the sample mean were removed prior to analysis (n = 3). Additionally, participants who failed to return for the second session were also removed from the analysis (n = 4).

Task

Experiment 2 used the same task as Experiment 1 with a key methodological modification. The two-block design was extended to a four-block design. The first two blocks followed the format of Experiment 1 and were followed by a feedback preference selection. The third and fourth blocks either honored or dishonored that selection, such that one block was consistent with the participant's choice (honored) and one was inconsistent (dishonored). These blocks also included an additional study paragraph and corresponding question set.

Study Design

The study employed a 2 (Feedback Timing: immediate vs. delayed) × 2 (Preference Honoring: honored vs. dishonored) within-subjects design. In addition, performance was assessed across multiple study blocks (Blocks 1-2 vs. Blocks 3-4) and test timepoints (e.g., initial and final assessments). This design ensured that participants made an informed feedback preference based on prior exposure to both feedback types in the first two blocks and allowed for a direct test of whether learning performance was influenced by alignment between feedback and stated preference. Learning performance was assessed using posttest and transfer test measures administered after each block, consistent with Experiment 1. To assess longer-term retention, Experiment 2 also included a delayed posttest and transfer test administered in a second session following a two-day delay.

Materials

Judgments of learning (JOLs). JOLs were again used to assess participants' metacognitive predictions about their future performance. After each study phase, participants were asked to estimate how many questions they would answer correctly on both the upcoming posttest and the transfer test. Specifically, participants reported their predicted number of correct responses out 4 out of 4 possible questions for each test for Blocks 1 and 2, and out of 6 for Blocks 3 and 4. These ratings were converted to proportion correct for analysis to maintain consistency with performance measures. JOL prompts are included in Appendix D.

Procedure

Figure 6 provides a flowchart of the experimental procedure for Experiment 2. The first session followed the same general structure as Experiment 1, with the addition of Blocks 3 and 4 to examine performance under preference-honoring and preference-dishonoring conditions. Participants began by completing four practice trivia items unrelated to the experimental materials to familiarize them with both immediate and delayed feedback. They were then shown a visual representation of the two feedback timing conditions (see Figure 1) and asked to indicate their preferred feedback type (immediate vs. delayed). After this instructional phase, participants completed the experimental training phase.

In Blocks 1 and 2, the procedure was identical to Experiment 1. Training consisted of two learning passages, four study questions per passage, four posttest questions identical to the study questions, and four transfer test questions that assessed the same concepts using novel wording. Following completion of these blocks, participants again viewed the feedback timing visualization and were asked to confirm their preference.

Participants then completed Blocks 3 and 4, which were structurally identical but included an additional learning passage and corresponding question set per block (resulting in three learning passages per block). These final blocks were used to implement preference-honoring and preference-dishonoring conditions, with feedback timing counterbalanced across participants to ensure that each participant experienced both matching and mismatching of their stated preference. The first session concluded with the Perceptions of Training Exit Survey. The second experimental session occurred two days later. Participants were notified via Prolific and given 24 hours to complete the session. This session consisted of the posttests and transfer tests for all blocks (see Appendix B).

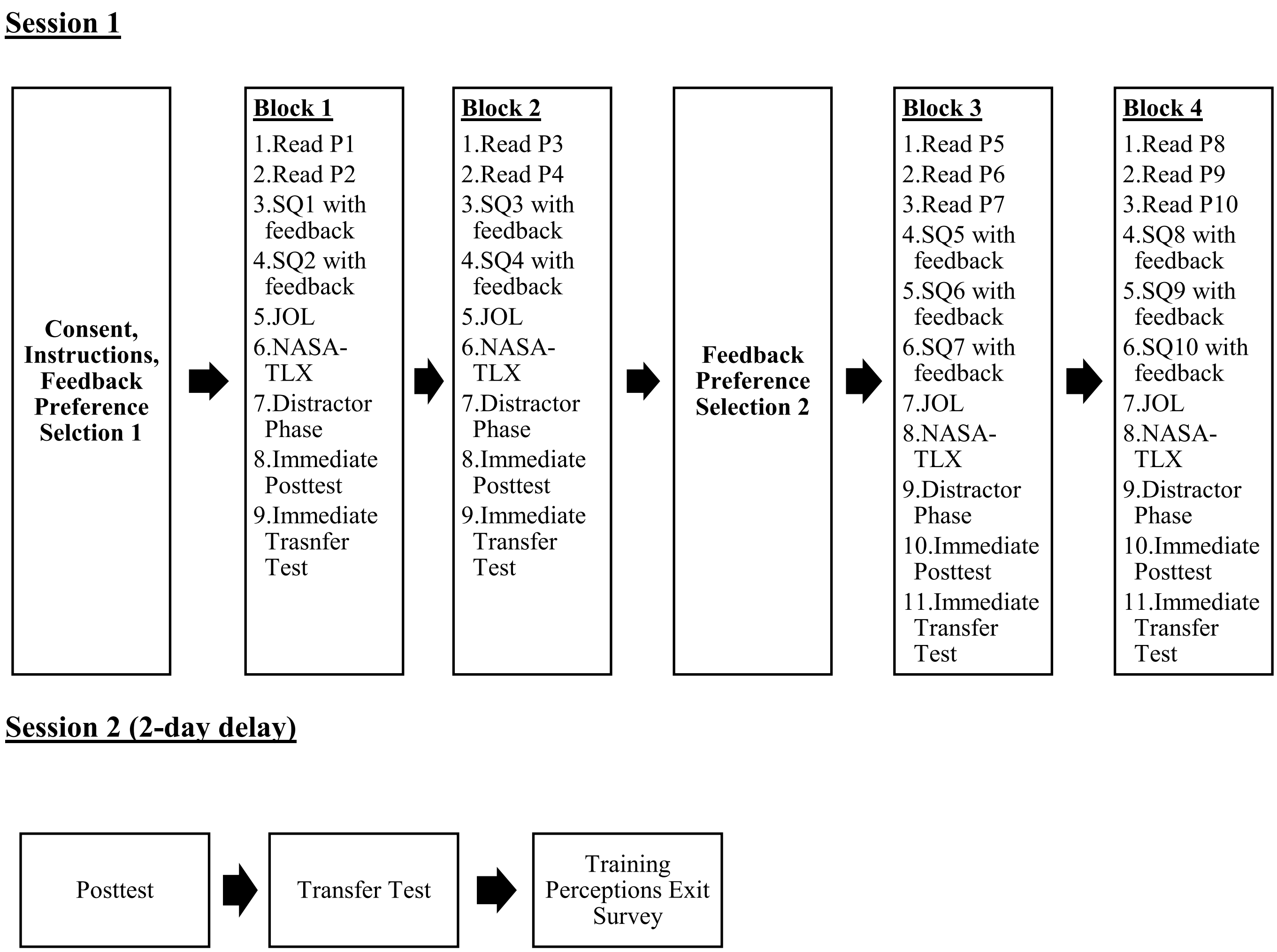

Figure 6. Flowchart of the experimental procedure for Experiments 2 and 3.

Results

Feedback Preference Selection

As in Experiment 1, a binomial test was conducted to examine whether the proportion of participants preferring immediate feedback differed significantly from chance (0.5). For the initial selection, 90.35% (n = 103) of participants indicated a preference for immediate feedback. In their final selection, 89.47% (n = 102) of participants indicated that they still preferred immediate feedback. For both timepoints, significantly more participants indicated a preference for immediate over delayed feedback ($p < .001$)

Figure 7 provides a visual representation of the percentage of participants who indicated a preference for either type of feedback across both timepoints. Seven participants adjusted their preference from immediate feedback at initial selection to delayed feedback upon completion of the experiment. Only 6 participants who initially indicated a preference for delayed feedback, indicated a change in preference to immediate feedback at the final timepoint. To examine changes in feedback preference over time, an exact McNemar's test was conducted. In contrast to Experiment 1, results of McNemar's indicated no significant change in preference ($p > .05$), such that there were not significantly more people who changed their preference from immediate to delayed feedback or vice versa (OR = 0.86, 95% CI [0.24, 2.98]).

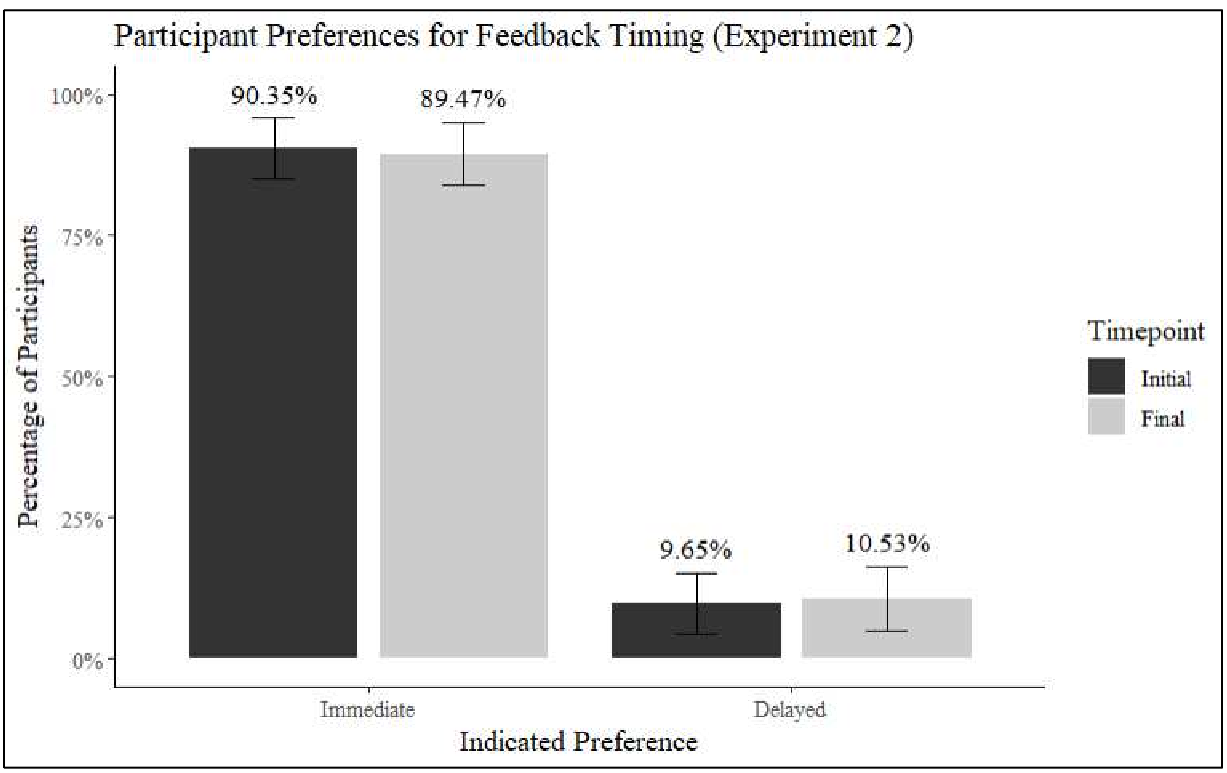

Figure 7. Percentages of participants who indicated a preference for immediate versus delayed feedback across time points for Experiment 2. Error bars represent 95% confidence intervals for

associated percentages.

As in Experiment 1, analyses were restricted to participants who indicated a preference for immediate feedback and paired-samples *t*-tests were used to compare accuracy across preference-honoring conditions within this group. Performance data for the delayed-preference group is reported descriptively due to the small sample size in this subgroup. For all performance measures, accuracy is reported as proportion of items correct (i.e., number of correct responses divided by total trials) to account for differences in the number of observations across study phases.

Study Question Accuracy

Overall Accuracy by Feedback Timing. In Blocks 1 and 2, study question accuracy was significantly higher when participants received delayed feedback ($M = 0.88$, $SD = 0.18$) than immediate feedback ($M = 0.77$, $SD = 0.23$), $t_{(113)} = 5.04$, $p < .001$, $d = 0.49$. However, in Blocks 3 and 4, this pattern reversed, with significantly higher accuracy on study questions following immediate feedback ($M = 0.82$, $SD = 0.20$) than delayed feedback ($M = 0.74$, $SD = 0.23$), $t_{(113)} = 3.70$, $p < .001$, $d = 0.35$.

Accuracy as a Function of Feedback Timing Preference. Group means are reported in Table 7. For Blocks 1 and 2, study question accuracy was significantly lower when participants' preference for immediate feedback was honored compared to when it was dishonored, $t_{(102)} = 4.53$, $p < .001$, $d = 0.45$. For Blocks 3 and 4, accuracy was significantly higher when participants' preference for immediate feedback was honored compared to when it was dishonored, $t_{(101)} = 3.36$, $p = .001$, $d = 0.33$.

Table 7. Experiment 2: Study Question Accuracy as a Function of Feedback Preference Across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M* (*SD*)	Blocks 3 and 4 *M* (*SD*)
Immediate	Honored	0.77 (0.22)	0.83 (0.20)
Immediate	Dishonored	0.87 (0.18)	0.75 (0.23)
Delayed	*Honored*	*0.95 (0.10)*	*0.68 (0.26)*
Delayed	*Dishonored*	*0.73 (0.28)*	*0.82 (0.22)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred immediate feedback due to the small number of participants indicating a delayed feedback preference.

Initial Posttest Accuracy

Overall Accuracy by Feedback Timing. In Blocks 1 and 2, initial posttest accuracy was significantly higher when participants had experienced delayed feedback (M = 0.95, SD = 0.13) than when they had experienced immediate feedback (M = 0.90, SD = 0.18), $t_{(113)} = 2.90$, p = .004, d = 0.27, although the near-ceiling levels of performance suggest caution is necessary when interpreting the findings. In addition, this difference did not hold for Blocks 3 and 4, where posttest accuracy was similar between the delayed (M = 0.90, SD = 0.14) and immediate (M = 0.89, SD = 0.16) feedback conditions, $t_{(113)} = 0.66$, p = .513, d = 0.06.

Accuracy as a Function of Feedback Timing Preference. Group means are reported in Table 8. During Blocks 1 and 2, initial posttest accuracy was significantly lower when participants' preference for immediate feedback was honored compared to when it was dishonored, $t_{(102)} = 2.68$, p = .009. For those who indicated a preference for delayed feedback, descriptive comparisons indicated higher accuracy on the initial posttest when preferences for delayed feedback were honored than when they were dishonored. In Blocks 3 and 4, initial posttest accuracy did not differ between when participants' preference for immediate feedback was honored and dishonored, $t_{(101)} = 0.52$, $p = .607$, $d = 0.05$.

Table 8. Experiment 2: Initial Posttest Accuracy as a Function of Feedback Preference Across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M* (*SD*)	Blocks 3 and 4 *M* (*SD*)
Immediate	Honored	0.90 (0.1)	0.90 (0.16)
Immediate	Dishonored	0.95 (0.13)	0.91 (0.14)
Delayed	*Honored*	*1.00 (0.00)*	*0.88 (0.14)*
Delayed	*Dishonored*	*0.95 (0.10)*	*0.85 (0.17)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred immediate feedback due to the small number of participants indicating a delayed feedback preference.

Initial Transfer Test Accuracy

Overall Accuracy by Feedback Timing. For Blocks 1 and 2, initial transfer test accuracy was significantly higher in the delayed feedback condition ($M = 0.88$, $SD = 0.19$) than in the immediate feedback condition ($M = 0.70$, $SD = 0.27$), $t_{(113)} = 6.18$, $p < .001$, $d = 0.58$. For Blocks 3 and 4, transfer test accuracy was again higher in the delayed feedback condition ($M = 0.86$, $SD = 0.17$) compared to the immediate feedback condition, $M = 0.73$, $SD = 0.21$), $t_{(113)} = 15.06$, $p < .001$, $d = 1.41$.

Accuracy as a Function of Feedback Timing Preference. Group means are reported in Table 9. For Blocks 1 and 2, accuracy in the immediate feedback condition was lower when participants' preference for immediate feedback was honored compared to when it was dishonored, $t_{(102)} = 4.57$, $p < .001$, $d = 0.45$. For Blocks 3 and 4, accuracy in the immediate feedback condition did not differ between when participants' preference for immediate feedback was honored and dishonored, $t_{(101)} = 0.47$, $p = .639$, $d = 0.02$.

Table 9. Experiment 2: Initial Transfer test Accuracy as a Function of Feedback Preference Across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M*(*SD*)	Blocks 3 and 4 *M*(*SD*)
Immediate	Honored	0.73 (0.29)	0.75 (0.23)
Immediate	Dishonored	0.81 (0.24)	0.76 (0.21)
Delayed	*Honored*	*0.84 (0.20)*	*0.76 (0.22)*
Delayed	*Dishonored*	*0.64 (0.26)*	*0.74 (0.27)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred immediate feedback due to the small number of participants indicating a delayed feedback preference.

Final Posttest Accuracy

Overall Accuracy by Feedback Timing. For Blocks 1 and 2, accuracy did not differ between the immediate ($M = 0.90$, $SD = 0.16$) and delayed ($M = 0.89$, $SD = 0.18$) feedback conditions, $t_{(113)} = 0.73$, $p = .466$, $d = 0.07$. For Blocks 3 and 4, accuracy similarly did not differ between the immediate ($M = 0.85$, $SD = 0.19$) and delayed ($M = 0.86$, $SD = 0.17$) feedback conditions, $t_{(113)} = 0.48$, $p = .631$, $d = 0.05$.

Accuracy as a Function of Feedback Timing Preference. Group descriptives are reported in Table 10. For Blocks 1 and 2, accuracy in the immediate feedback condition did not differ between when participants' preference for immediate feedback was honored or dishonored, $t_{(102)} = 0.77$, $p = .445$, $d = 0.08$. For Blocks 3 and 4, accuracy in the immediate feedback condition also did not differ between when participants' preference for immediate feedback was honored and dishonored, $t_{(101)} = 0.32$, $p = .752$, $d = 0.03$.

Table 10. Experiment 2: Final Posttest Accuracy as a Function of Feedback Preference Across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M*(*SD*)	Blocks 3 and 4 *M*(*SD*)
Immediate	Honored	0.90 (0.16)	0.86 (0.18)
Immediate	Dishonored	0.88 (0.18)	0.86 (0.17)
Delayed	*Honored*	*0.82 (0.15)*	*0.91 (0.13)*
Delayed	*Dishonored*	*0.79 (0.28)*	*0.91 (0.17)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred immediate feedback due to the small number of participants indicating a delayed feedback preference.

Final Transfer Test Accuracy

Overall Accuracy by Feedback Timing. For Blocks 1 and 2, accuracy was higher in the delayed feedback condition ($M = 0.81$, $SD = 0.24$) than in the immediate feedback condition ($M = 0.63$, $SD = 0.29$), $t_{(113)} = 5.01$, $p < .001$, $d = 0.47$. For Blocks 3 and 4, accuracy did not differ between the delayed ($M = 0.76$, $SD = 0.21$) and immediate ($M = 0.75$, $SD = 0.23$) feedback conditions, $t_{(113)} = 0.56$, $p = .575$, $d = 0.05$.

Accuracy as a Function of Feedback Timing Preference. Group Descriptives are reported in Table 11. For Blocks 1 and 2, final transfer accuracy was significantly lower when participants' preference for immediate feedback was honored compared to when it was dishonored, $t_{(102)} = 4.57$, $p < .001$, $d = 0.45$. However, no significant difference in final transfer accuracy was observed for Blocks 3 and 4, $t_{(101)} = 0.47$, $p = .638$, $d = 0.05$.

Table 11. Experiment 2: Final Transfer Test Accuracy as a Function of Feedback Preference Across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M*(*SD*)	Blocks 3 and 4 *M*(*SD*)
Immediate	Honored	0.63 (0.29)	0.75 (0.23)
Immediate	Dishonored	0.81 (0.24)	0.76 (0.21)
Delayed	*Honored*	*0.84 (0.20)*	*0.76 (0.22)*
Delayed	*Dishonored*	*0.64 (0.26)*	*0.74 (0.27)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred immediate feedback due to the small number of participants indicating a delayed feedback preference.

Judgements of Learning (JOLs)

Overall Posttest JOLs by Feedback Timing. For Blocks 1 and 2, posttest JOLs were higher in the delayed feedback condition (M = 0.93, SD = 0.13) than in the immediate feedback condition (M = 0.88, SD = 0.17), $t_{(113)} = 3.83$, $p < .001$, $d = 0.36$. For Blocks 3 and 4, posttest JOLs did not differ between the delayed (M = 0.79, SD = 0.20) and immediate (M = 0.81, SD = 0.19) feedback conditions, $t_{(113)} = 1.33$, $p = .186$, d = 0.12.

Posttest JOL as a Function of Feedback Timing Preference. Group descriptives are reported in Table 12. For Blocks 1 and 2, posttest JOLs in the immediate feedback condition were lower when participants' preference for immediate feedback was honored compared to when it was dishonored, $t_{(102)} = 3.58$, $p < .001$, $d = 0.35$. For Blocks 3 and 4, posttest JOLs in the immediate feedback condition did not differ between when participants' preference for immediate feedback was honored and dishonored, $t_{(101)} = 1.29$, $p = .199$, $d = 0.13$. Figure 8 provides a visual comparison of participants' initial and final posttest performance compared to their JOLs.

Table 12. Experiment 2: Posttest JOL as a Function of Feedback Preference across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M* (*SD*)	Blocks 3 and 4 *M* (*SD*)
Immediate	Honored	0.88 (0.17)	0.80 (0.19)
Immediate	Dishonored	0.93 (0.13)	0.79 (0.20)
Delayed	*Honored*	*0.93 (0.12)*	*0.85 (0.21)*
Delayed	*Dishonored*	*0.89 (0.17)*	*0.86 (0.19)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred immediate feedback due to the small number of participants indicating a delayed feedback preference.

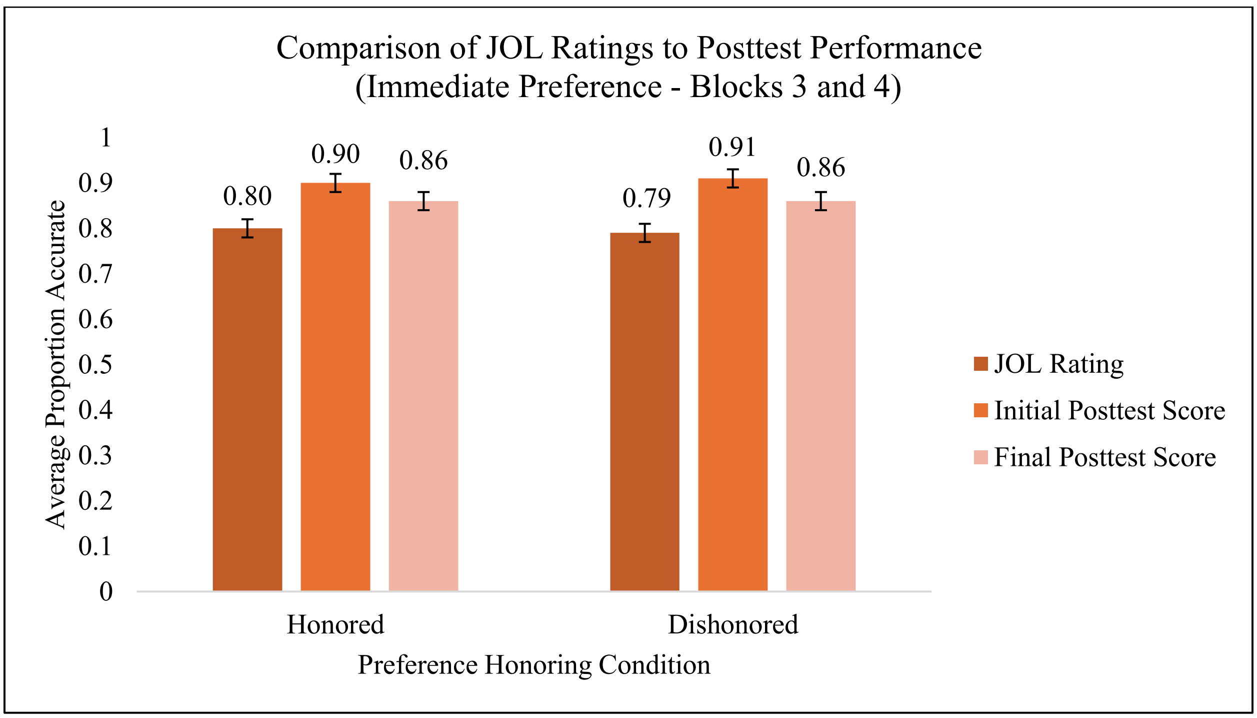

Figure 8. Visual comparison of participants' who indicated a preference for immediate feedback judgments of learning (JOLs), initial posttest, and final posttest performance. JOLs reflect predicted proportion correct at posttest, whereas posttest scores reflect actual performance. Honored data reflects immediate feedback whereas dishonored data reflects delayed feedback conditions.

Overall Transfer test JOLs by Feedback Timing. For Blocks 1 and 2, transfer test JOLs were higher in the delayed feedback condition ($M = 0.84$, $SD = 0.17$) than in the immediate feedback condition (M = 0.80, SD = 0.19), $t_{(113)}$ = 2.70, p = .008, d = 0.26. For Blocks 3 and 4, transfer-test JOLs did not significantly differ between the delayed (M = 0.75, SD = 0.20) and immediate (M = 0.78, SD = 0.19) feedback conditions, $t_{(113)}$ = 1.77, p = .07, d = 0.18.

Transfer test JOL as a Function of Feedback Timing Preference. Group Descriptives are reported in Table 13. For Blocks 1 and 2, posttest JOLs in the immediate feedback condition were lower when participants' preference for immediate feedback was honored compared to when it was dishonored, $t_{(102)}$ = 2.36, p = .020, d = 0.23. For Blocks 3 and 4, transfer test JOLs in the immediate feedback condition did not differ between when participants' preference for immediate feedback was honored and dishonored, $t_{(101)}$ = 2.15, p = .034, d = 0.21. Figure 9 provides a visual comparison of participants' initial and final posttest performance compared to their JOLs.

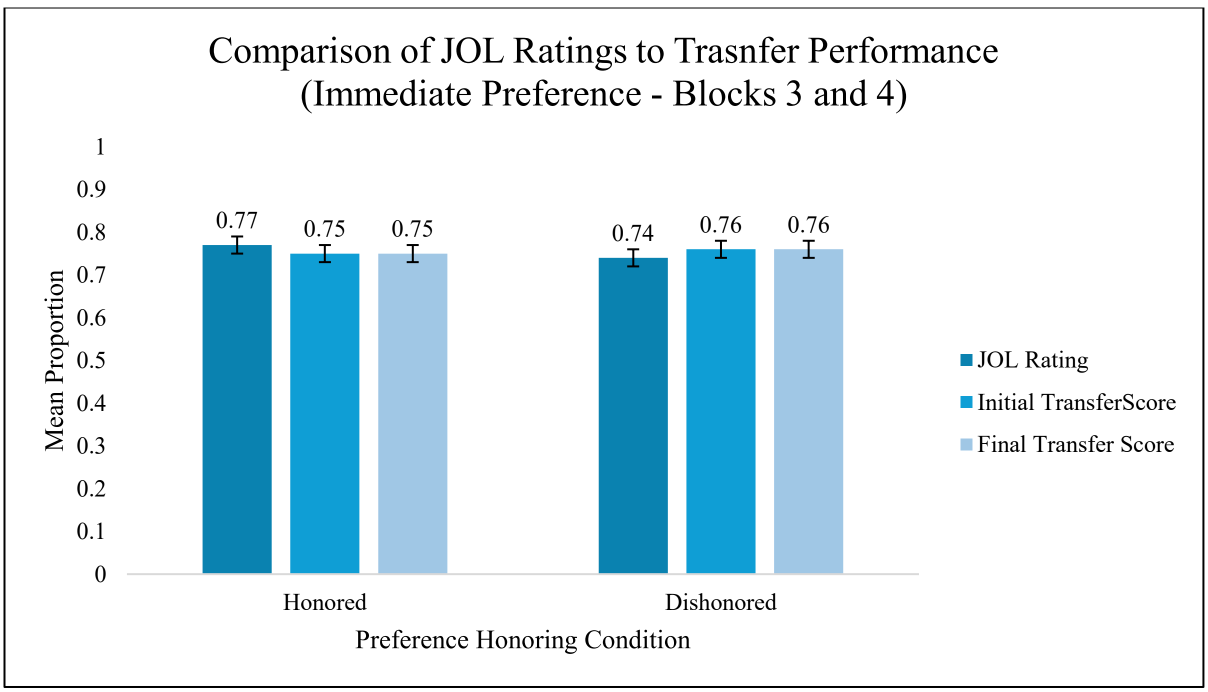

Figure 9. Visual comparison of participants' who indicated a preference for immediate feedback judgments of learning (JOLs), initial transfer test, and final transfer test performance. JOLs reflect predicted proportion correctly at transfer test, whereas transfer test scores reflect actual performance. Honored data reflects immediate feedback whereas dishonored data reflects delayed feedback conditions.

Table 13. Experiment 2: Transfer Test JOLs as a Function of Feedback Preference across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M* (*SD*)	Blocks 3 and 4 *M* (*SD*)
Immediate	Honored	0.80 (0.19)	0.77 (0.19)
Immediate	Dishonored	0.84 (0.17)	0.74 (0.20)
Delayed	*Honored*	*0.89 (0.13)*	*0.86 (0.22)*
Delayed	*Dishonored*	*0.80 (0.22)*	*0.83 (0.20)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred immediate feedback due to the small number of participants indicating a delayed feedback preference.

Perceived Workload (NASA-TLX)

Overall Workload. Perceived workload was assessed using the NASA-TLX (Hart & Staveland, 1988) and was first split between overall ratings associated with immediate and delayed feedback for Blocks 1 and 2 as well as Blocks 3 and 4. Descriptive statistics and inferential comparisons between feedback timing conditions for all NASA-TLX subscales are presented in Appendix H (Tables H2–H3). To assess the impact of feedback timing on perceptions of workload during training, ratings for each subscale were compared by feedback condition via paired-samples t-tests.

In Blocks 1 and 2, participants rated immediate feedback ($M = 10.54$, $SD = 5.15$) as significantly more mentally demanding than delayed feedback ($M = 9.02$, $SD = 5.08$), $t_{(113)} = 3.53$, $p < .001$, $d = 0.33$. They also rated the study phase as significantly more frustrating when given immediate feedback ($M = 5.25$, $SD = 5.30$) over delayed feedback ($M = 3.98$, $SD = 3.89$), $t_{(113)} = 3.03$, $p = .003$, $d = 0.36$. Conversely, they indicated that their subjective success in their performance was significantly higher when receiving delayed feedback ($M = 17.02$, $SD = 3.99$) relative to immediate feedback ($M = 15.19$, $SD = 4.92$), $t_{(113)} = 3.84$, $p < .001$, $d = 0.36$. However, in Blocks 3 and 4, this pattern of results changed. Participants no longer indicated any significant differences in their perceptions of mental workload or frustration, but did indicate higher perceptions of success in their performance when given immediate (M = 15.53, SD = 4.33) over delayed feedback (M = 14.34, SD = 4.72), $t_{(113)} = 3.26$, $p = .001$, $d = 0.31$.

Workload as a Function of Feedback Timing Preference. Descriptive statistics for both the immediate and delayed preference groups, as well as inferential comparisons between preference-honored and preference-dishonored conditions for the immediate preference group, are reported in Appendix I (Tables I3–I6). During Blocks 1 and 2, participants who preferred immediate feedback reported significantly higher mental workload when their preference was honored (M = 10.49, SD = 5.16) compared to when it was dishonored (M = 9.02, SD = 5.07), $t_{(102)} = 3.26$, $p = .002$, $d = 0.32$. Participants also reported feeling significantly more frustrated when their preference was honored (M = 5.33, SD = 5.45) rather than dishonored (M = 4.17, SD = 3.96), $t_{(102)} = 2.62$, $p = .010$, $d = 0.26$. In contrast, they reported that they felt they were more successful in their performance during study when their preferences were dishonored (M = 16.87, SD = 4.08) rather than honored (M = 15.31, SD = 4.85), $t_{(102)} = 3.21$, $p = .002$, $d = 0.32$.

In Blocks 3 and 4, these patterns essentially reversed. Participants in this immediate preference group reported significantly higher mental workload when their preference was dishonored (M = 12.80, SD = 5.48) rather than honored (M = 12.33, SD = 5.36), $t_{(101)} = 2.20$, p = .030, $d = 0.22$, and felt their performance was significantly higher when given feedback that honored their preference (M = 15.41, SD = 4.41) than dishonored their preference (M = 14.28, SD = 4.70), $t_{(101)} = 2.86$, $p = .005$, $d = 0.28$. Moreover, they did not report any significant difference in frustration between preference honoring conditions in these later blocks.

Perceptions of Training.

Descriptive statistics for all perceptions of training measures are provided in Appendix J (Table J2).

Discussion

Experiment 2 assessed how providing learners input into their feedback timing influenced both their learning outcomes and subjective perceptions of learning, both through their initial judgements and after experience with the task. Across both Experiments 1 and 2, most participants indicated a preference for immediate feedback, replicating prior literature regarding feedback preferences and supporting H2 (Lefreve & Cox, 2016; 2017; Rowe & Wood, 2008). Interestingly, the finding from Experiment 1 that participants were more likely to shift their preference from immediate to delayed feedback after experiencing the task was not replicated in Experiment 2. Participants in Experiment 2 were similarly likely to shift their preference to immediate feedback. This inconsistency suggests that experience with the learning materials may not consistently alter learners' feedback timing preferences to the degree initially expected. Instead, preferences for immediate feedback appear to remain relatively stable, even when learners experience conditions that may ultimately be less beneficial for learning outcomes.

It was also hypothesized that delayed feedback would produce superior learning outcomes when compared to immediate feedback (H3). Overall, findings provided little support for this hypothesis. While delayed feedback produced some benefits for learning performance during the initial posttest and transfer test, these effects were limited primarily to earlier study blocks and were no longer observed at the final posttest or final transfer assessment. The pattern of findings across the initial and delayed tests may also provide support for Metcalfe et al.'s (2009) argument that benefits associated with delayed feedback may partially reflect recency effects at test rather. Because delayed feedback advantages were strongest during the earlier assessments and diminished over time, the present findings suggest that the mechanisms underlying delayed feedback benefits may be more complex than accounts based exclusively on desirable difficulty or increased cognitive effort during encoding (Butler et al., 2007; Guzman-Muñoz & Johnson, 2007; Kluger & DeNisi, 1996).

Experiment 2 also specifically tested the hypothesis that matching learners' preferred feedback timing would enhance learning outcomes, reduce subjective workload, and improve perceptions of performance (H4). Although inferential comparisons involving participants who preferred delayed feedback were again limited by the small sample size of this subgroup, findings from the immediate-preference group remain theoretically informative. Importantly, participants who preferred immediate feedback did not reliably benefit when this preference was honored. In some cases (e.g., immediate transfer performance), dishonoring learners' immediate feedback preference was associated with superior accuracy. These findings suggest that learners' preferred feedback timing does not necessarily correspond to the feedback timing that best supports learning, mirroring other findings in the self-regulated learning literature (Bjork et al., 2014; Soderstrom & Bjork, 2015).

With regard to how feedback timing and user preferences influence subjective perceptions of learning, findings suggest that feedback timing shaped learners' perceptions of the learning experience, although these effects again differed across phases of the study. During Blocks 1 and 2, participants rated immediate feedback as more mentally demanding and more frustrating than delayed feedback, while also perceiving themselves as less successful during learning. However, these differences in perceived mental demand and frustration were no longer observed during Blocks 3 and 4 and participants instead reported greater perceptions of success when receiving immediate feedback. Similar patterns emerged within the immediate-preference group when comparing preference-honored and preference-dishonored conditions. Early in training, honoring participants' preference for immediate feedback was associated with greater perceived mental demand and frustration, whereas dishonoring this preference was associated with greater perceptions of performance success. However, in later blocks, participants reported greater perceived workload when their immediate feedback preference was dishonored and greater perceptions of success when their preference was honored.

Collectively, these findings suggest that learners' subjective perceptions of workload and performance are dynamic and may shift as learners gain experience with the task and feedback conditions (Vandewaeter et al, 2011; Johnson, et al., 2017; Johnson & Priest, 2014)). More importantly, subjective perceptions did not consistently align with objective learning outcomes. Across both experiments, learners generally preferred immediate feedback and often perceived it as beneficial for their performance, despite evidence that delayed feedback sometimes produced equivalent or superior transfer performance. These findings suggest that learners may not have full metacognitive awareness of the instructional conditions that best support their long-term learning (Dunlosky & Rawson, 2012; Roediger & Karpicke, 2006).

CHAPTER 4 – EXPERIMENT 3

Just as previous literature indicates that learners demonstrate preferences regarding feedback timing, research also suggests that learners hold preferences for the content of feedback (Rowe & Wood, 2008). Two broad categories of feedback content are commonly examined: explanatory feedback and outcome feedback. Although these terms encompass a range of implementations, explanatory feedback is generally defined as providing reasoning or elaboration beyond whether an answer was correct, whereas outcome feedback simply indicates correctness (Johnson et al., 2017; Johnson & Priest, 2014). Qualitative work by Lefevre and Cox (2016; 2017) suggests that learners often report preferring explanatory feedback, particularly because it provides additional information about why a response was correct or incorrect. However, despite this stated subjective preference, learners report a preference for feedback to be succinct and efficient and often favor simpler outcome feedback when given control over what they view (Lefreve & Cox, 2016).

Research has also demonstrated that preferences and effectiveness of feedback content may differ as a function of learner expertise (Johnson & Priest, 2014). A review of feedback design by Johnson et al. (2017) notes that expert learners tend to prefer and perform best with concise outcome feedback, whereas novice learners benefit more from explanatory feedback that supports development and conceptual understanding. Therefore, learners may not always accurately identify the type of feedback that will optimize learning, particularly when metacognitive judgments are inaccurate or influenced by convenience or perceived cognitive effort (Dunlosky & Rawson, 2012; Roediger & Karpicke, 2006; Soderstrom & Bjork, 2015). This warrants further investigation of what may drive learners' preference for feedback content, and how those preferences impact learning.

Although most work regarding preferences for feedback content focuses on learners' subjective preferences or behavior, Pridemore and Klein (1991) examined the impact of providing learner control over feedback content on learning outcomes. Learners either selected explanatory or outcome feedback for each item or had feedback type assigned by the system. Although they found no overall difference between learner-controlled and system-controlled feedback delivery, they did observe a substantial benefit of explanatory feedback over outcome feedback. However, the between-subjects nature of this study does not allow for comparisons between learner preferences for explanatory or outcome feedback.

Prior work on learner preferences and metacognitive judgments has argued that evaluating whether learner control is beneficial requires comparing situations in which learner preferences are both honored and dishonored (Pashler et al, 2008). This approach has been useful in identifying discrepancies between what learners believe will support their learning and what actually enhances retention and performance (Kornell & Metcalfe, 2006). While Experiments 1 and 2 found an overwhelming preference for immediate feedback over delayed feedback, it remains necessary to investigate whether preferences for feedback content similarly influence learning outcomes, and whether honoring those preferences improves performance. This was undertaken in Experiment 3.

Method

Participants

The final sample included in the analysis was approximately 106 participants (Cis-men = 44, Cis-women = 60, Trans-women = 1, Nonbinary = 1). The sample ranged between 18 to 71 years old (*M* = 36.76, *SD* = 11.79). All participants were located in the United States, all were fluent English speakers, and all reported normal or corrected to normal vision. Participants were recruited through Prolific and compensated \$15 for their participation in session 1, and an additional \$7.50 for their participation in session 2. The first session of the experiment took an average of 56.68 (*SD* = 14.83) minutes to complete, and the second session of the experiment took an average of 22.63 (*SD* = 9.42) minutes to complete. On average, participants completed the second session 60.44 (*SD* = 9.12) hours after completing the first session.

An a priori power analysis for a repeated-measures ANOVA indicated that a total sample of 100 participants would be required to detect an effect size of .15 with 95% power, assuming an alpha level of .05 and a two-tailed test. To ensure sufficient power across conditions and to account for possible attrition, 120 participants were recruited. To ensure data quality, participants whose total completion time fell 3 standard deviations above or below the sample mean were removed prior to analysis (*n* = 10). Additionally, participants who failed to return for the second session were also removed from the analysis (*n* = 4).

Task

Experiment 3 was nearly identical to Experiment 2 with the exception that, instead of examining the honoring preferences for timing of feedback, the goal was to investigate the impact of honoring user preferences for the content of feedback. Feedback timing was constant in this experiment, with feedback being delivered immediately after each response during study.

Study Design

Experiment 3 utilized producing a 2 (Feedback content: outcome or explanatory) × 2 (Preference Honoring: honored or dishonored) within-subjects design. In addition, performance was assessed across multiple study blocks (Blocks 1-2 vs. Blocks 3-4) and test timepoints (e.g., initial and final assessments).

Materials

Feedback Instructions and Examples. Participants were instructed on the difference between outcome and explanatory feedback and given a visual depiction representing the difference between the two, before indicating their initial preference for feedback content (see Figure 10). They were instructed that they would experience both types of feedback during the study and that their selection would not impact on the way feedback was presented to them.

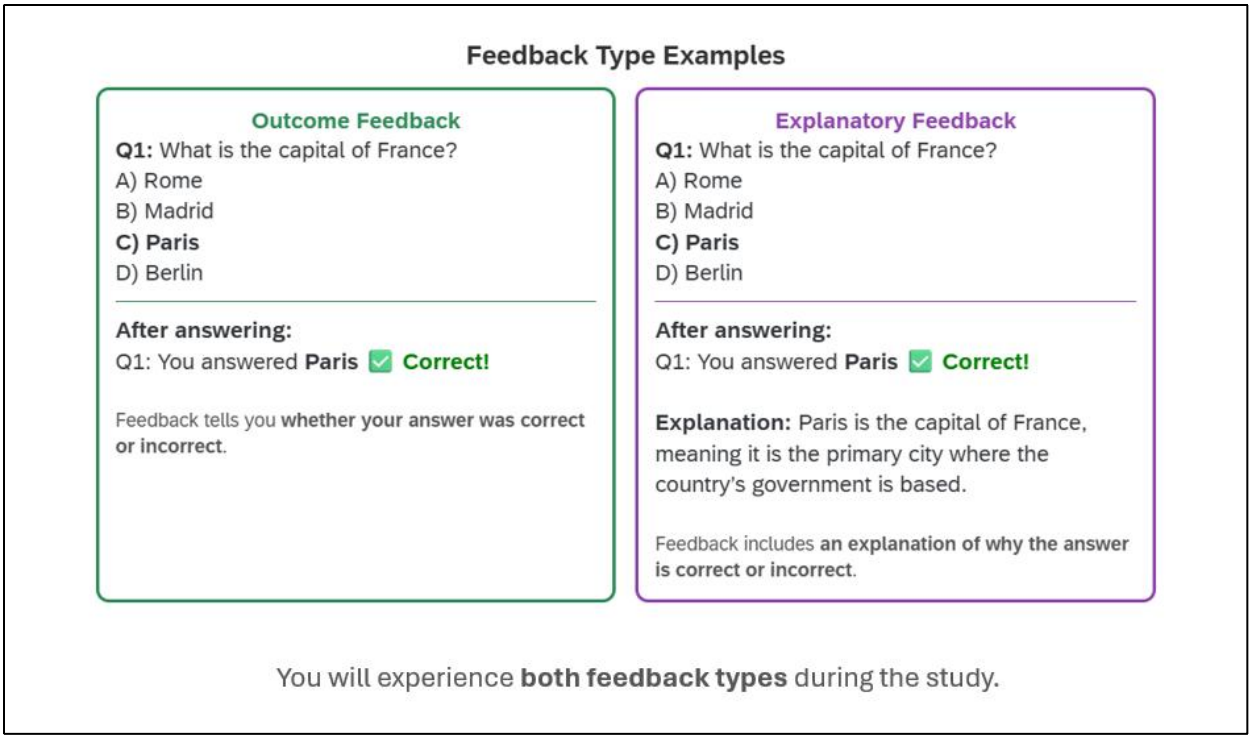

Figure 10. Visual Depiction representing outcome (left panel) and explanatory (right panel) feedback shown to participants. This content is unrelated to the content participants are instructed on.

Feedback messages. Participants were given feedback regarding the accuracy of their response to each question. To hold the timing of feedback constant, feedback was displayed immediately following each response to a study question for 15 seconds. Feedback messages were formatted consistently across items. In the outcome feedback condition, feedback messages restated the question, indicated whether their input was correct or incorrect, and provided the correct answer if incorrect. In the explanatory feedback condition, the messages contained the same information with the addition of a 1-sentence explanation of the concept. Visual examples of the differences between feedback messages between content conditions are provided in Figure 11.

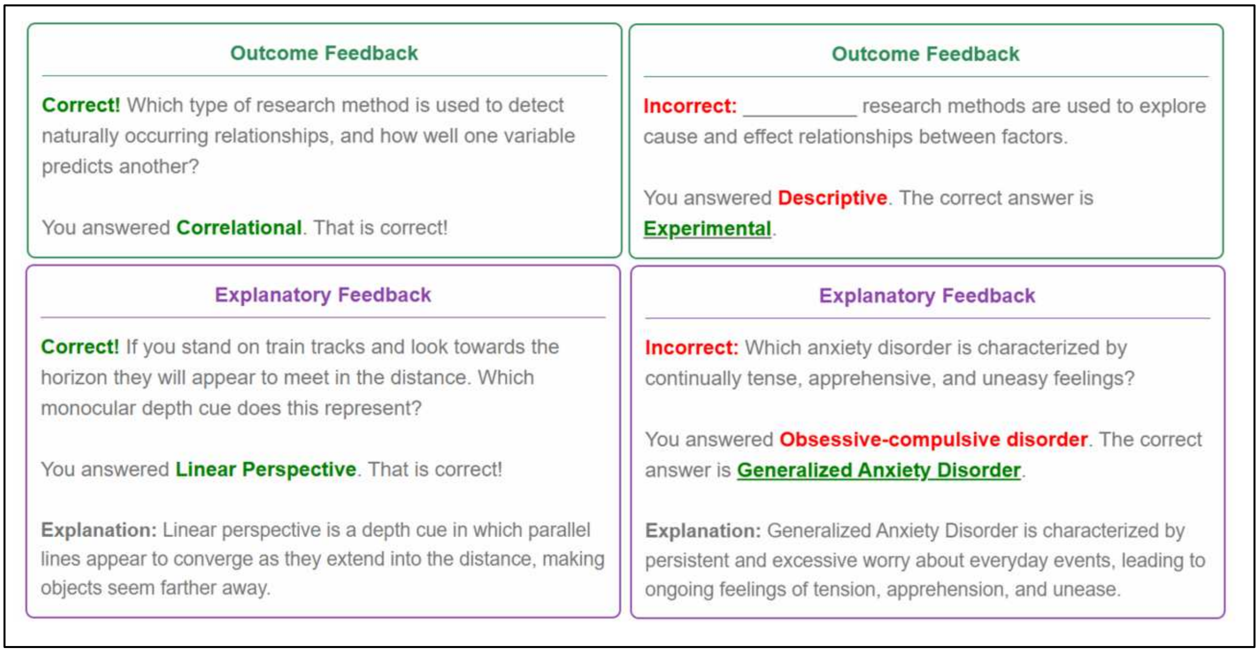

Figure 11. Examples of feedback messages for both Outcome and Explanatory conditions, for both correct and incorrect responses made by participants.

Procedure

Participants were first instructed on both types of feedback content and given a figure representing the difference in explanatory and correct answer feedback (see Figure 10). Participants were instructed that they would experience each type of feedback and would later be asked to select which type of feedback they prefer. Otherwise, with the exception of the nature of the feedback, the materials and procedure were identical to Experiment 2 (see Figure 6).

Results

Feedback Preference Selection

For the initial selection, 87.74% (n = 93) of participants indicated a preference for explanatory feedback. In their final selection, 83.96% (n = 89) of participants indicated that they still preferred explanatory feedback. A binomial test indicated that significantly more participants preferred explanatory feedback to outcome feedback at both timepoints ($p < .001$)

Figure 12 provides a visual representation of the percentage of participants who indicated a preference for either type of feedback across both timepoints. Eight participants changed their preference from explanatory feedback at the initial selection to outcome feedback at the final timepoint, whereas 4 participants who initially preferred outcome feedback switched to explanatory feedback by the final timepoint. An exact McNemar's test indicated no significant change in preference ($p > .05$), such that there were not significantly more people who changed their preference from explanatory to outcome or vice versa (OR = 2, 95% CI [0.54, 9.08]).

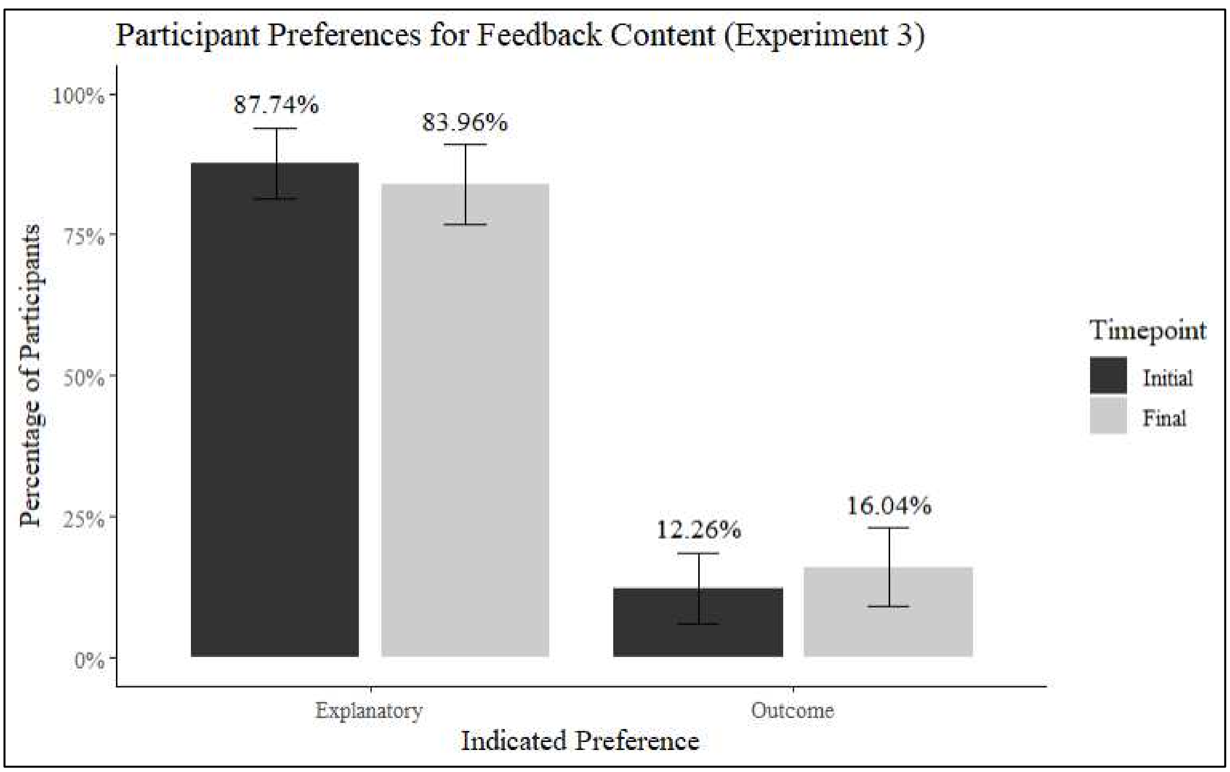

Figure 12. Percentages of participants who indicated a preference for immediate versus delayed feedback across time points for Experiment 2. Error bars represent 95% confidence intervals for associated percentages.

Similarly to Experiments 1 and 2, inferential analyses were restricted to participants who indicated a preference for explanatory feedback and paired-samples *t*-tests were used to compare performance across preference-honoring conditions within this group. Performance data for the outcome-preference group are reported descriptively due to the small sample size in this subgroup.

Study Question Accuracy

Overall Accuracy by Feedback Content. For Blocks 1 and 2, study question accuracy was significantly higher when participants were given explanatory feedback ($M = 0.89$, $SD = 0.17$) than outcome feedback ($M = 0.78$, $SD = 0.24$), $t_{(105)} = 4.19$, $p < .001$, $d = 0.41$. Blocks 3 and 4 showed the opposite pattern, with outcome feedback ($M = 0.81$, $SD = 0.20$) showing higher study question accuracy than explanatory ($M = 0.77$, $SD = 0.22$), $t_{(105)} = 2.22$, $p = .029$, $d = 0.22$.

Accuracy as a Function of Feedback Content Preference. Group descriptives are reported in Table 15. In Blocks 1 and 2, participants who indicated a preference for explanatory feedback showed significantly higher accuracy on study questions when that preference was honored ($M = 0.90$, $SD = 0.17$) than when it was dishonored ($M = 0.80$, $SD = 0.23$), $t_{(92)} = 3.72$, $p < .001$, $d = 0.39$. In Blocks 3 and 4, accuracy on study questions was similar when preference for explanatory feedback was honored and dishonored, $t_{(88)} = 1.61$, $p = .111$, $d = 0.17$.

Table 15. Experiment 3: Study Question Accuracy as a Function of Feedback Preference Across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M*(*SD*)	Blocks 3 and 4 *M*(*SD*)
Explanatory	Honored	0.90 (0.17)	0.78 (0.22)
Explanatory	Dishonored	0.80 (0.23)	0.81 (0.20)
Outcome	*Honored*	*0.67 (0.28)*	*0.81 (0.22)*
Outcome	*Dishonored*	*0.85 (0.19)*	*0.72 (0.20)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred explanatory feedback due to the small number of participants indicating an outcome feedback preference.

Initial Posttest Accuracy

Overall Accuracy by Feedback Content. In Blocks 1 and 2, explanatory feedback produced higher accuracy ($M = 0.94$, $SD = 0.15$) than outcome feedback ($M = 0.90$, $SD = 0.19$) at initial posttest, but this difference did not reach the threshold for statistical significance, $t_{(105)} = 1.88$, $p = .063$, $d = 0.18$, and was indicative of performance at ceiling, suggesting caution in interpreting the findings. Participants showed similar initial posttest accuracy in Blocks 3 and 4, regardless of whether they received outcome ($M = 0.91$, $SD = 0.12$) or explanatory ($M = 0.91$, $SD = 0.17$) feedback during study, $t_{(105)} = 0.27$, $p = .790$, $d = 0.03$.

Accuracy by Feedback Content Preference. Group descriptives are reported in Table 16. In Blocks 1 and 2, participants who indicated a preference for explanatory feedback were slightly more accurate on the initial posttest when their preference was honored than dishonored, but this difference was not statistically significant, $t_{(92)} = 1.77$, $p = .079$, $d = 0.18$. For Blocks 3 and 4, participants demonstrated nearly equivalent accuracy at posttest when their preference for explanatory feedback was honored or dishonored, $t_{(88)} = 0.37$, $p = .710$, $d = 0.04$.

Table 16. Experiment 3: Initial Posttest Accuracy as a Function of Feedback Preference Across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M*(*SD*)	Blocks 3 and 4 *M*(*SD*)
Explanatory	Honored	0.95 (0.13)	0.91 (0.13)
Explanatory	Dishonored	0.92 (0.16)	0.91 (0.17)
Outcome	*Honored*	*0.81 (0.33)*	*0.91 (0.12)*
Outcome	*Dishonored*	*0.87 (0.24)*	*0.92 (0.13)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred explanatory feedback due to the small number of participants indicating an outcome feedback preference.

Initial Transfer Test Accuracy

Overall Accuracy by Feedback Content. For Blocks 1 and 2, transfer test accuracy was higher in the explanatory feedback condition ($M = 0.88$, $SD = 0.19$) than in the outcome feedback condition ($M = 0.77$, $SD = 0.26$) $t_{(105)} = 4.07$, $p < .001$, $d = 0.40$. For Blocks 3 and 4, accuracy was also higher in the explanatory feedback condition ($M = 0.84$, $SD = 0.18$) relative to the outcome feedback condition ($M = 0.57$, $SD = 0.17$), $t_{(105)} = 13.05$, $p < .001$, $d = 1.27$.

Accuracy by Feedback Content Preference. Group descriptives are reported in Table 17. Within the explanatory feedback condition, accuracy for Blocks 1 and 2 was higher when participants' preference for explanatory feedback was honored, $t_{(92)} = 3.73$, $p < .001$, $d = 0.39$. For Blocks 3 and 4, accuracy in the explanatory feedback conditions similarly was higher when participants' preference for explanatory feedback was honored, $t_{(88)} = 11.52$, $p < .001$, $d = 0.42$)

Table 17. Experiment 3: Initial Transfer Test Accuracy as a Function of Feedback Preference Across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M*(*SD*)	Blocks 3 and 4 *M*(*SD*)
Explanatory	Honored	0.89 (0.19)	0.85 (0.18)
Explanatory	Dishonored	0.79 (0.24)	0.58 (0.17)
Outcome	*Honored*	*0.67 (0.36)*	*0.53 (0.20)*
Outcome	*Dishonored*	*0.85 (0.19)*	*0.82 (0.22)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred explanatory feedback due to the small number of participants indicating an outcome feedback preference.

Final Posttest Accuracy

Overall Accuracy by Feedback Content. For Blocks 1 and 2, accuracy was higher in the outcome feedback condition ($M = 0.91$, $SD = 0.18$) than in the explanatory feedback condition ($M = 0.79$, $SD = 0.21$), $t_{(105)} = 3.98$, $p < .001$, $d = 0.39$. For Blocks 3 and 4, accuracy was again higher in the outcome feedback condition ($M = 0.84$, $SD = 0.21$) compared to the explanatory feedback condition ($M = 0.76$, $SD = 0.19$), $t_{(105)} = 3.45$, $p < .001$, $d = 0.33$.

Accuracy by Feedback Content Preference. Group Descriptives are reported in Table 18. Within the explanatory feedback condition, accuracy for Blocks 1 and 2 was higher when participants' preference for explanatory feedback was dishonored compared to when it was honored, $t_{(92)} = 4.94$, $p < .001$, $d = 0.51$. For Blocks 3 and 4, accuracy in the explanatory feedback condition also was higher when participants' preference for explanatory feedback was dishonored relative to when it was honored, $t_{(88)} = 2.97$, $p = .004$, $d = 0.32$.

Table 18. Experiment 3: Final Posttest Accuracy as a Function of Feedback Preference Across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M* (*SD*)	Blocks 3 and 4 *M* (*SD*)
Explanatory	Honored	0.80 (0.20)	0.84 (0.20)
Explanatory	Dishonored	0.93 (0.15)	0.77 (0.18)
Outcome	*Honored*	*0.73 (0.28)*	*0.73 (0.28)*
Outcome	*Dishonored*	*0.75 (0.29)*	*0.71 (0.27)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred explanatory feedback due to the small number of participants indicating an outcome feedback preference.

Final Transfer Test Accuracy

Overall Accuracy by Feedback Content. For Blocks 1 and 2, accuracy did not differ between the outcome feedback condition (M = 0.67, SD = 0.30) and the explanatory feedback condition (M = 0.67, SD = 0.24), $t_{(105)} = 0.17$, $p = .868$, $d = 0.02$. For Blocks 3 and 4, accuracy similarly did not differ between the outcome (M = 0.75, SD = 0.21) and explanatory (M = 0.74, SD = 0.23) feedback conditions, $t_{(105)} = 0.52$, $p = .605$, $d = 0.05$.

Accuracy by Feedback Content Preference. Group descriptives are reported in Table 19. Within the explanatory feedback condition, accuracy for Blocks 1 and 2 did not differ between when participants' preference for explanatory feedback was honored and dishonored, $t_{(92)} = 0.12$, $p = .906$, $d = 0.01$. For Blocks 3 and 4, accuracy in the explanatory feedback condition similarly did not differ between when participants' preference for explanatory feedback was honored and dishonored, $t_{(88)} = 0.94$, $p = .350$, $d = 0.01$.

Table 19. Experiment 3: Final Transfer Test Accuracy as a Function of Feedback Preference Across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M* (*SD*)	Blocks 3 and 4 *M* (*SD*)
Explanatory	Honored	0.67 (0.24)	0.74 (0.22)
Explanatory	Dishonored	0.67 (0.30)	0.76 (0.21)
Outcome	*Honored*	*0.54 (0.25)*	*0.71 (0.22)*
Outcome	*Dishonored*	*0.63 (0.24)*	*0.75 (0.24)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred explanatory feedback due to the small number of participants indicating an outcome feedback preference.

Judgements of Learning (JOLs)

Overall Posttest JOL by Feedback Content. For Blocks 1 and 2, posttest JOLs were higher in the explanatory feedback condition (M = 0.91, SD = 0.15) than in the outcome feedback condition (M = 0.86, SD = 0.19), $t_{(105)} = 3.17$, $p = .002$, $d = 0.31$. For Blocks 3 and 4, posttest JOLs were higher in the outcome feedback condition (M = 0.79, SD = 0.21) compared to the explanatory feedback condition (M = 0.74, SD = 0.23), $t_{(105)} = 2.85$, $p = .005$, $d = 0.28$.

Posttest JOL by Feedback Content Preference. Group descriptives are reported in Table 20. For Blocks 1 and 2, posttest JOLs differed significantly as a function of feedback content preference, $t_{(92)} = 3.40$, $p = .001$, $d = 0.35$. For Blocks 3 and 4, posttest JOLs also differed significantly as a function of feedback content preference, $t_{(88)} = 2.88$, $p = .005$, $d = 0.31$. Figure 13 provides a visual comparison between JOLs and actual test performance as a function of preference honoring.

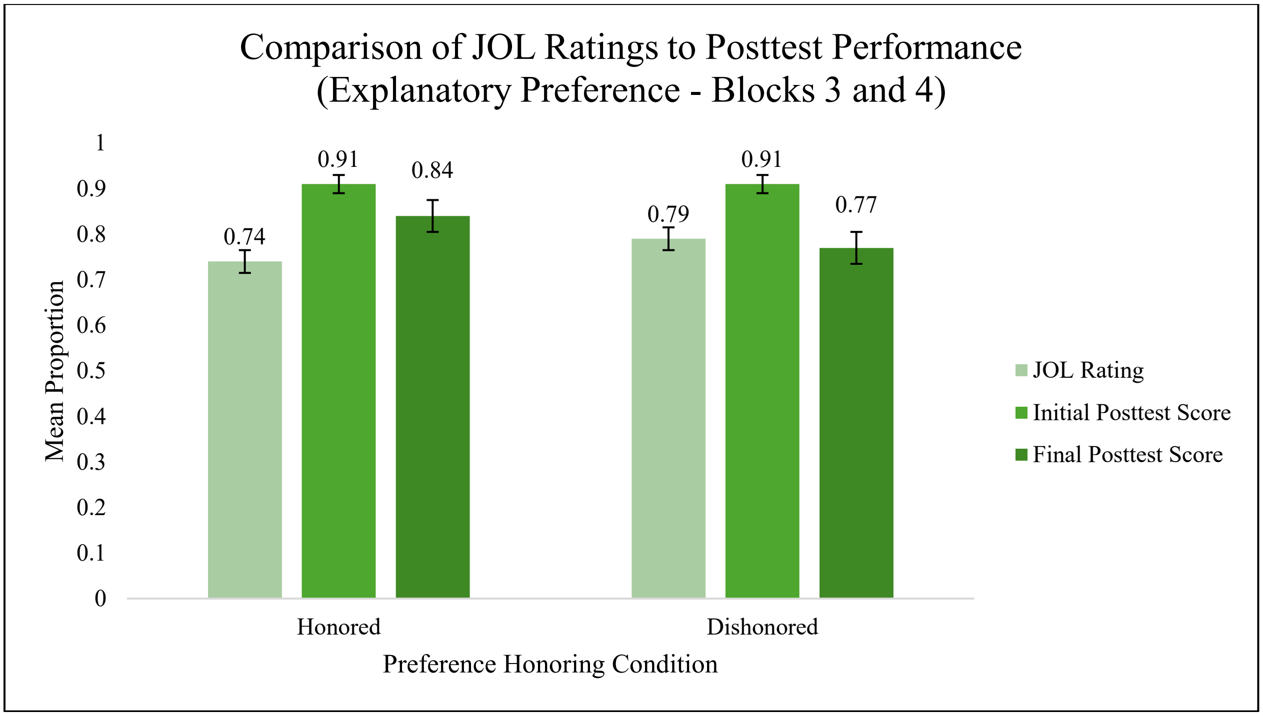

Figure 13. Visual comparison of participants who indicated a preference for explanatory feedback judgments of learning (JOLs), initial posttest, and final posttest performance. JOLs reflect predicted proportion correct at posttest, whereas posttest scores reflect actual performance. Honored data reflects explanatory feedback whereas dishonored data reflects outcome feedback conditions.

Table 20. Experiment 3: Posttest JOLs as a Function of Feedback Preference Across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M*(*SD*)	Blocks 3 and 4 *M*(*SD*)
Explanatory	Honored	0.91 (0.15)	0.74 (0.23)
Explanatory	Dishonored	0.85 (0.20)	0.79 (0.22)
Outcome	*Honored*	*0.88 (0.13)*	*0.75 (0.19)*
Outcome	*Dishonored*	*0.88 (0.17)*	*0.71 (0.22)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred explanatory feedback due to the small number of participants indicating an outcome feedback preference.

Overall Transfer Test JOLs by Feedback Content. For Blocks 1 and 2, transfer test JOLs were higher in the explanatory feedback condition (*M* = 0.85, *SD* = 0.19) than in the outcome feedback condition (*M* = 0.79, SD = 0.20), $t_{(105)} = 3.29$, $p = .001$, $d = 0.32$. For Blocks 3 and 4, transfer test JOLs did not significantly differ between the outcome (*M* = 0.75, *SD* = 0.22) and explanatory (*M* = 0.72, *SD* = 0.23) feedback conditions, $t_{(105)} = 1.84$, $p = .068$, $d = 0.18$.

Transfer test JOL by Feedback Content Preference. Group descriptives are reported in Table 21. For Blocks 1 and 2, transfer test JOLs were significantly higher when participants who preferred explanatory feedback had their preference honored rather than dishonored, $t_{(92)} = 3.96$, $p < .001$, $d = 0.41$. For blocks 3 and 4, transfer test JOLs were not significantly different based on whether the preference for explanatory feedback was honored or dishonored $t_{(88)} = 1.51$, $p = .134$, $d = 0.16$. Figure 14 provides a visual comparison between JOLs and actual test performance as a function of preference honoring.

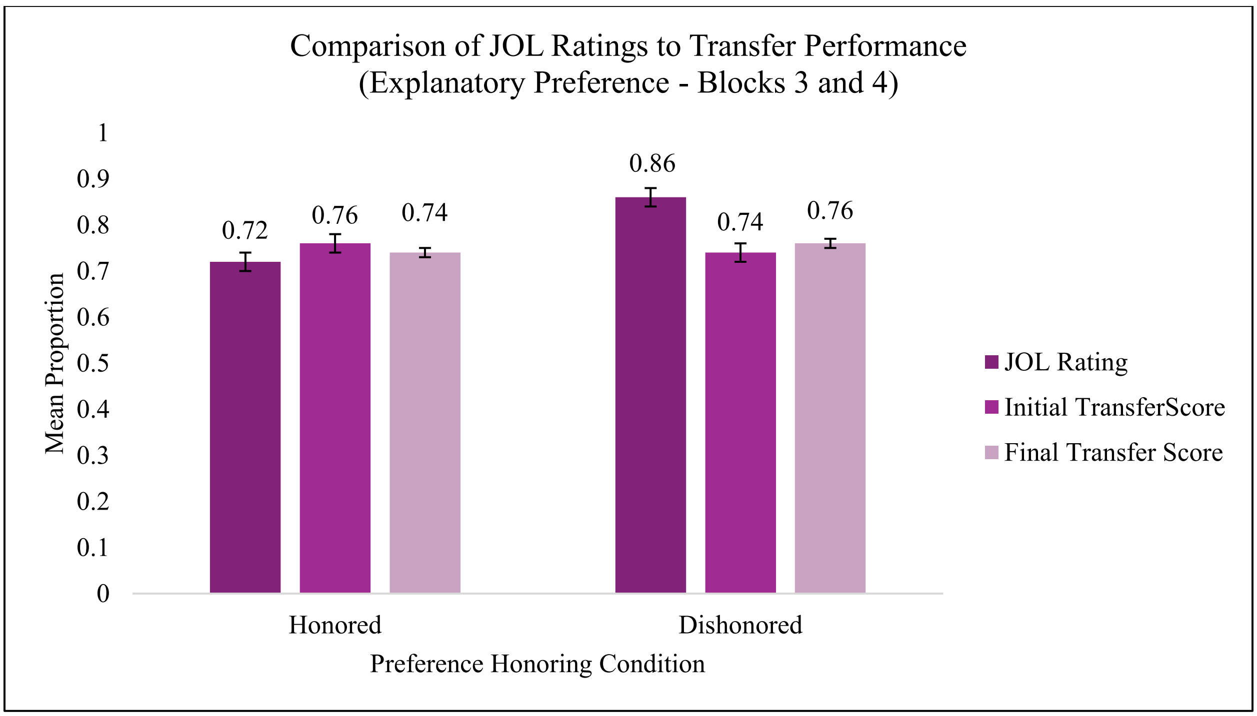

Figure 14. Visual comparison of participants who indicated a preference for explanatory feedback judgments of learning (JOLs), initial transfer test, and final transfer test performance. JOLs reflect predicted proportion correctly at transfer test, whereas transfer test scores reflect actual performance. Honored data reflects explanatory feedback whereas dishonored data reflects outcome feedback conditions.

Table 21. Experiment 3: Transfer Test JOLs as a Function of Feedback Preference Across Study Blocks

Feedback Condition	Preference Condition	Blocks 1 and 2 *M*(*SD*)	Blocks 3 and 4 *M*(*SD*)
Explanatory	Honored	0.87 (0.19)	0.72 (0.23)
Explanatory	Dishonored	0.80 (0.21)	0.86 (0.22)
Outcome	*Honored*	*0.79 (0.17)*	*0.71 (0.20)*
Outcome	*Dishonored*	*0.77 (0.24)*	*0.67 (0.20)*

Note. Values represent proportion correct. Inferential comparisons were conducted only for participants who preferred explanatory feedback due to the small number of participants indicating an outcome feedback preference.

Perceived Workload (NASA-TLX)

Overall Workload. Descriptive statistics and inferential comparisons between feedback content conditions for all NASA-TLX subscales are presented in Appendix H (Tables H4–H5). To assess the impact of feedback content on perceptions of workload during training, ratings for each subscale were compared by feedback content condition via paired-samples t-tests.

In Blocks 1 and 2, participants rated outcome feedback (M = 10.83, SD = 4.96) as significantly more mentally demanding than explanatory feedback (M = 9.58, SD = 5.17), $t_{(105)}$ = 3.00, p = .003, d = 0.29. They also indicated that the study phase was significantly more frustrating when given outcome feedback (M = 5.51, SD = 5.04) over explanatory feedback (M = 4.25, SD = 4.62), $t_{(105)}$ = 3.19, p = .002, d = 0.31. Accordingly, participants indicated that their subjective success was significantly higher when receiving explanatory feedback (M = 16.90, SD = 4.09) relative to outcome feedback (M = 15.40, SD = 4.64), $t_{(105)}$ = 3.30, p = .001, d = 0.32. In Blocks 3 and 4, no significant effect of feedback content was found on any of the five subscales.

Workload as a Function of Feedback Timing Preference. Descriptive statistics for both the immediate and delayed preference groups, as well as inferential comparisons between preference-honored and preference-dishonored conditions for the immediate preference group, are reported in Appendix I (Tables I7–I10). During Blocks 1 and 2, participants who preferred explanatory feedback reported significantly higher mental workload when their preference was dishonored (M = 10.69, SD = 4.91) compared to when it was honored (M = 9.43, SD = 5.09), $t_{(92)}$ = 2.72, p = .008, d = 0.28. Participants also reported feeling significantly more frustrated when their preference was dishonored (M = 5.23, SD = 4.90) rather than honored (M = 3.94, SD = 4.56), $t_{(92)}$ = 3.42, p = .001, d = 0.35. They also reported that they felt they were more successful in their performance during study when their preferences were honored (M = 16.91, SD = 4.21) rather than dishonored (M = 15.57, SD = 4.62), $t_{(92)}$ = 2.77, p = .007, d = 0.29).

Again, in Blocks 3 and 4, none of these significant differences in feedback content emerged, and no significant effect of feedback content was found on any of the five subscales.

Perceptions of Training.

Descriptive statistics for all perceptions of training measures are provided in Appendix J (Table J3).

Discussion

Across the first two experiments, honoring preferences for feedback timing did not reliably improve learning outcomes, although it did influence learners' subjective perceptions of their learning experience, including perceived workload and performance. Experiment 3 extended this work by shifting the focus from when feedback is provided to what type of feedback is provided, comparing explanatory and outcome feedback. As hypothesized, learners demonstrated a strong and consistent preference for explanatory feedback over outcome feedback (H5), replicating prior findings indicating that learners favor more detailed, elaborative forms of instructional support (Mayer, 1997; Johnson & Priest, 2014; Johnson et al., 2017).

Thus, like feedback timing, learners hold clear preferences regarding feedback content that may not necessarily reflect optimal learning, just as Experiments 1 and 2 demonstrated with feedback timing, and in alignment with the broader literature on learner preferences.

I also predicted that explanatory feedback would produce superior learning outcomes, overall, compared to outcome feedback (H6). This hypothesis received mixed support. For posttest performance, explanatory feedback led to higher accuracy than outcome feedback during earlier study blocks. However, this advantage did not persist at later blocks, where no reliable differences emerged. In contrast, outcome feedback was associated with better performance on the delayed posttest across both early and later blocks, suggesting a potential shift in the relative benefits of feedback type over time. This could in part be due to familiarity with the task over time and learners adopting a strategy for integrating smaller amounts of feedback (Vandewaeter et al, 2011; Kalyuga, 2009; Kalyuga, 2021; Lee et al., 2014). For transfer performance, explanatory feedback generally produced better outcomes than outcome feedback during both early and later blocks, although this advantage did not extend to the final transfer test, where no significant differences were observed. Together, these findings suggest that the benefits of explanatory versus outcome feedback may depend on both the timing of assessment and the phase of learning, rather than reflecting a consistently superior feedback type as posed in the feedback literature (Bangert-Drowns et al., 1991; Schute, 2008).

Experiment 3 also investigated whether aligning feedback content with learner preferences would enhance learning outcomes. This hypothesis also received mixed support (H7). At the immediate posttest, there were no effects of preference honoring on performance. At the delayed posttest, however, performance was higher when participants' preference for explanatory feedback was dishonored (i.e., when they received outcome feedback) than when it was honored. Preference honoring significantly improved performance on initial transfer tests, but this effect did not hold for the delayed transfer test. Additionally, accuracy during the learning phase itself showed inconsistent differences across preference honoring conditions. In sum, preference alignment did not systematically influence learning performance.

In all, Experiment 3 demonstrated that preferences for feedback content, like preferences for feedback timing, do not reliably align with optimal learning outcomes. While learners consistently favored explanatory feedback, this preference did not translate into consistent performance benefits when it was honored. Instead, learning outcomes appeared to depend more on feedback type and assessment timing than whether instruction matched learners' stated preferences. Thus, learner preferences may be poor indicators of the instructional conditions that most effectively support learning.

CHAPTER 5 – GENERAL DISCUSSION

Adaptive systems are increasingly capable of personalizing instruction, yet personalization often occurs without meaningful learner input regarding how feedback is delivered (Durlach, 2019). While self-regulated learning research has shown a discrepancy between learner's preferences and optimal learning strategies, it has also shown that these preferences drive the ways learners interact with instructional content (Bjork, et al., 2013; Metcalfe et al., 2025; Putnam, et al., 2016; Rhodes & Castel, 2009). Although the primary goal of adaptive training systems is to provide individualized learning experiences that optimize performance, instructional design features may also influence learners' willingness to persist with the system (Bjork, et al., 2013; Durlach, 2019; Durlach & Spain, 2014; Metcalfe et al., 2025; Putnam, et al., 2016). Consider, for example, a language-learning application that delivers feedback in a rigid and non-personalized manner, without allowing learners any control over how or when that feedback is presented. Although such feedback may be instructionally accurate, continuously receiving corrective information in a format misaligned with learner preferences may reduce motivation, increase frustration, and negatively influence perceptions of competence and autonomy. As a result, learners may disengage from the system altogether, despite the potential instructional benefits of the feedback itself.

Prior literature has produced inconsistent findings regarding the most effective feedback timing and content strategies (Johnson et al., 2017). Moreover, relatively little research has examined the cognitive and metacognitive mechanisms that shape learners' feedback preferences or the extent to which these preferences influence learning outcomes (Durlach, 2019; Lefevre & Cox, 2016; Rowe & Wood, 2008). In the three present experiments, learners demonstrated strong and consistent preferences regarding both the timing and content of feedback, favoring feedback that was delivered immediately and that provided detailed explanatory information. In Experiments 1 and 2, nearly 90% of participants preferred immediate rather than delayed feedback, while a similarly substantial proportion of participants in Experiment 3 preferred explanatory feedback over less informative outcome-based feedback. Importantly, learners generally maintained their initial preferences even after direct experience with the learning task and feedback conditions.

Although learners expressed clear preferences for the timing and content of their feedback, these preferences did not reliably correspond to superior learning performance. Delayed feedback occasionally produced benefits for transfer performance and early posttest accuracy, but these advantages were diminished across later assessments. Furthermore, learners who preferred immediate feedback did not consistently benefit when this preference was honored, as might be expected if learners possessed accurate metacognitive insight into which feedback conditions best supported their learning (Pashler et al, 2008; Kornell & Metcalfe, 2006; Rhodes et al, 2013). Similar patterns emerged for feedback content preferences. Although explanatory feedback sometimes improved transfer performance, outcome feedback occasionally produced superior delayed posttest performance. Collectively, these findings suggest that preference alignment alone is insufficient to enhance learning outcomes. Learners may prefer feedback conditions that feel more supportive, intuitive, or effective during learning, even when those conditions do not consistently maximize long-term retention or transfer. These findings thus align with broader theories of metacognitive miscalibration, suggesting that learners may require support in identifying and selecting instructional strategies that optimally benefit learning (Bjork, et al., 2014; Dunlosky & Rawson, 2012; Roediger & Karpicke, 2006).

While honoring learner preferences did not reliably improve objective performance, it meaningfully influenced subjective perceptions of the learning experience. During earlier phases of training, immediate feedback was associated with greater perceived mental workload, increased frustration, and lower perceptions of success. During later phases of training immediate feedback became associated with greater perceived success and lower workload when learner preferences were honored. These findings suggest that subjective perceptions of feedback may evolve dynamically as learners gain familiarity with both the task and the feedback structure itself. Consequently, adaptive systems may benefit from incorporating greater learner control later in training once learners have acquired sufficient experience to make more informed judgments regarding their preferred feedback conditions.

Theoretical implications

The present findings also contribute to theoretical discussions regarding the mechanisms underlying feedback effectiveness. Specifically, the findings provide support for theories proposing that some delayed feedback advantages may partially reflect recency effects rather than exclusively desirable difficulty mechanisms (Bjork, 2020; Metcalfe et al., 2009; Mayer, 2024). Additionally, the present findings suggest that immediate feedback may impose greater cognitive demands during early learning phases (Fiorella et al., 2012; Johnson et al., 2017; Kluger & DeNisi, 1996; Mayer, 1997). Regarding feedback content, the relative benefits of explanatory versus outcome feedback depended on both the timing of assessment and the learner's phase of training rather than reflecting a universally superior feedback type, as has sometimes been implied within the feedback literature (Bangert-Drowns et al., 1991; Shute, 2008). Together, these findings reinforce the broader conclusion that feedback effectiveness is likely context-dependent and shaped by dynamic interactions among learner preferences, task demands, and instructional design.

This research also directly informs how introducing more elements of self-regulated learning into computerized training impacts learning outcomes. Across all three experiments, it became clear that learner preferences for feedback timing and content reflected subjective mental workload and perceptions of one's own performance within a task rather than accurate assessments of their current learning state. As honoring learner preferences did not consistently produce better learning outcomes, preferences for feedback timing and content mirror preferences for other instructional interventions, particularly methods that feel less mentally demanding and produce an inflated sense of knowledge (Bjork et al., 2014; Dunlosky & Rawson, 2012; Karpicke & Roediger, 2006; Kirk-Johnson et al., 2019; Soderstrom & Bjork, 2015). Additionally, the present experiment provides support for the notion that feedback serves as cognitive mechanism of calibrating mental models of knowledge and an influence on learners' motivation and subjective self-perceptions (Shute, 2008). Therefore, feedback design should aim to accommodate both accurate learning as well as promote a sense of capability in the learner.

Practical implications

The present findings can inform principles of design in adaptive training systems that have the potential to apply to more commercially relevant domains. By refining our understanding of the impact certain feedback interventions have on the learner, we can increase training system efficiency and effectiveness. Further, by refining our understanding of how learner preferences for feedback play a role in their perceptions of and learning with a system we can improve training retention. With most learners indicating a preference for immediate, explanatory feedback, this can be implemented to promote learners' preferences into adaptive training systems, especially if the system is intended to be more self-directed or incentivize continued use of the system. As honoring learner preferences improved more subjective perception of workload and learning than objective learning outcomes, systems may benefit from partial learner control whereby learner preferences are weighted alongside performance metrics and cognitive load indicatives, allowing for strategic override of preferences when these preferences are found to be suboptimal.

Limitations

The findings should be interpreted with several experimental limitations in mind. First, the global judgement of feedback preference produced a small sample size for those preferring delayed and outcome-based feedback. While these groups represent a clear minority, the distribution of preferences limits conclusions that can be drawn about honoring or dishonoring preferences in these groups, especially in comparison to those who prefer immediate or explanatory feedback. In addition, preferences and performance with feedback measured in a controlled context may vary greatly from educational environments. Although this study took place online in a computerized format, it is possible that these findings do not apply well to classroom instruction or adaptive systems intended to teach domain specific complex tasks. Moreover, these studies included feedback presentation that was based on accuracy alone. More sophisticated adaptive instructional systems that take more objective workload measures like response time or dual-task performance into feedback presentation may find different results around subjective perceptions differing by feedback type.

Future Directions

Future research should examine how learner feedback preferences can be incorporated into adaptive algorithms that simultaneously weigh factors such as cognitive workload, performance accuracy, and learner preference to create feedback interventions that both support effective learning and promote positive subjective learning experiences. Additionally, the present studies relied on global judgments of feedback preference, which limited the scope of comparisons that could be made regarding learner control. Future work should therefore investigate whether an item-by-item approach to feedback selection (cf. Hays et al., 2010), similar to those commonly used in the self-regulated learning literature, provides additional insight into how preference honoring and dishonoring influence learning outcomes. Finally, future work could examine whether individuals value an explanation as to why their preferences have been honored or dishonored and the subsequent effect of this on their learning and subjective experience.

Conclusion

Collectively, learner preferences for feedback are meaningful components of the learning experience, but their instructional value depends on the goal of implementing feedback: whether that is to motivate, inform, or produce greater learning outcomes. Learners are active participants in feedback interventions. Therefore, effective adaptive instruction may require not only determining what feedback is most beneficial for learning but also understanding how learners experience and respond to that feedback over time.

REFERENCES

Astwood, R. S., Van Buskirk, W. L., Cornejo, J. M., & Dalton, J. (2008). The impact of different feedback types on decision-making in simulation-based training environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 52*(26), 2062–2066.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The Instructional Effect of Feedback in Test-Like Events. *Review of Educational Research, 61*(2), 213–238. https://doi.org/10.3102/00346543061002213

Bates, L. E., Myers, S. J., DeLosh, E. L., & Rhodes, M. G. (2025). The learning benefits of mastery quizzing versus other quizzing methods. *Psychology Learning & Teaching, 24*(1), 74–91. https://doi.org/10.1177/14757257241301733

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. *From learning processes to cognitive processes: Essays in honor of William K. Estes, 2*, 35-67.

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. Journal of *Applied Research in Memory and Cognition, 9*(4), 475.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology, 64*(1), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Butler, A.C., Karpicke, J.D., & Roediger, H.L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied, 13,* 273–281.

Carolan, T. F., Hutchins, S. D., Wickens, C. D., & Cumming, J. M. (2014). Costs and benefits of more learner freedom: Meta-analyses of exploratory and learner control training methods. *Human Factors*, *56*(5), 999-1014.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction*, *8*(4), 293-332.

Clegg, B. A., Karduna, A., Holen, E., Garcia, J., Rhodes, M. G., & Ortega, F. R. (2024). Multimedia and immersive training materials influence impressions of learning but not learning outcomes. *2022 Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).*

Cowan, N. (1999). An embedded-processes model of working memory. Models of working memory: Mechanisms of active maintenance and executive control, *20*(506), 1013-1019.

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences, 116*, 19251-19257.

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self-evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271-280.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58.

Durlach, P. J. (2019). Fundamentals, Flavors, and Foibles of Adaptive Instructional Systems. In R. A. Sottilare & J. Schwarz (Eds.), Adaptive Instructional Systems (Vol. 11597, pp. 76–95). Springer International Publishing. https://doi.org/10.1007/978-3-030-22341-0_7

Durlach, P. J., & Ray, J. M. (2011). Designing Adaptive Instructional Environments: Insights from Empirical Evidence: Defense Technical Information Center. https://doi.org/10.21236/ADA552677

Durlach, P., Spain, R. (2014). Framework for instructional technology: methods of implementing adaptive training and education (No. ARIRR1335). CRC Press, Boca Raton. www.dtic.mil/docs/citations/ADA597411

Fiorella, L., Vogel-Walcutt, J. J., & Schatz, S. (2012). Applying the modality principle to real-time feedback and the acquisition of higher-order cognitive skills. *Educational Technology Research and Development, 60* , 223–238

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist, 34*(10), 906.

Fyfe, E. R., de Leeuw, J. R., Carvalho, P. F., Goldstone, R. L., Sherman, J., Admiraal, D., ... & Motz, B. A. (2021). ManyClasses 1: Assessing the generalizable effect of explanatory feedback versus delayed feedback across many college classes. *Advances in Methods and Practices in Psychological Science*, *4*(3), https://doi.org/10.1177/25152459211027575

Guzman-Muñoz, F. J., & Johnson, A. (2007). Error feedback and the acquisition of geographical representations. Applied Cognitive Psychology, 22, 979-995. doi:10.1002/acp.141

Hays, M. J., Kornell, N., & Bjork, R. A. (2010). The costs and benefits of providing feedback during learning. *Psychonomic bulletin & review*, *17*(6), 797-801.

Jason, M.P., Wissel, S., Franziska, I., & Undorf, M. (2026). Judgements of learning in the wild: Establishing ecological validity with an intelligent tutoring system in a field of study. *Learning and Instruction*, *103*, 102293.

Johnson, C. I., Bailey, S. K. T., & Van Buskirk, W. L. (2017). Designing Effective Feedback Messages in Serious Games and Simulations: A Research Review. In P. Wouters & H. Van Oostendorp (Eds.), Instructional Techniques to Facilitate Learning and Motivation of Serious Games (pp. 119–140). Springer International Publishing. https://doi.org/10.1007/978-3-319-39298-1_7

Johnson, C. I., & Priest, H. A. (2014). The feedback principle in multimedia learning. In R. E. Mayer (Ed.), The Cambridge handbook of multimedia learning (pp. 449–463). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.023

Kalyuga, S. (2009). The expertise reversal effect. *Educational Psychologist*, 38(1), 25-31

Kalyuga, S. (2021). The expertise reversal principle in multimedia learning. Cambridge University Press.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, *17*(3), 249-255.

Kimball, D. R., Smith, T. A., & Muntean, W. J. (2012). Does delaying judgments of learning really improve the efficacy of study decisions? Not so much. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(4), 923–954.

Kirk-Johnson, A., Galla, B. M., & Fraundorf, S. H. (2019). Perceiving Effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cognitive Psychology*, *115*, 101–237.

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

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of experimental psychology: Learning, Memory, and Cognition*, *32*(3), 609-622.

Kornell, N., & Rhodes, M. G. (2013). Feedback reduces the metacognitive benefit of tests. *Journal of Experimental Psychology: Applied*, *19*(1), 1-13.

Kulik, J. A., & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. Review of educational research, 58(1), 79-97.

Landsberg, C. R., Van Buskirk, W. L., & Astwood Jr, R. S. (2010). Does feedback type matter? Investigating the effectiveness of feedback content on performance outcomes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54(27), 2304-2308

Lee, C. H., Kalyuga, S., & Wales, S. (2014). Expertise reversal effect and its instructional implications. *Applying the science of learning in education: Infusing psychological science into the curriculum*, 31-44.

Lefevre, D., & Cox, B. (2016). Feedback in technology-based instruction: Learner preferences. British Journal of Educational Technology, 47(2), 248-256.

Lefevre, D., & Cox, B. (2017). Delayed instructional feedback may be more effective, but is this contrary to learners' preferences?. British Journal of Educational Technology, 48(6), 1357-1367.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. Journal of Educational Psychology, 106(4), 901–918. https://doi.org/10.1037/a0037123

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? Educational psychologist, 32(1), 1-19.

Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review*, *36*(1), 8.

Mayer, R. E., & Moreno, R. (1998a). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. Journal of Educational Psychology, 90(2), 312–320. https://doi.org/10.1037/0022-0663.90.2.312

Mayer, R. E., & Moreno, R. (1998b). A cognitive theory of multimedia learning: Implications for design principles. Journal of Educational Psychology, 91(2), 358-368.

Mayer, R. E., & Johnson, C. I. (2010). Adding instructional features that promote learning in a game-like environment. *Journal of Educational Computing Research, 42*, 241–265.

Metcalfe, J. (2011). Desirable difficulties and studying in the region of proximal learning. *Successful remembering and successful forgetting: A Festschrift in honor of Robert A. Bjork*, 259-276.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. Psychonomic Bulletin & Review, 15(1), 174-179.

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

Metcalfe, J., Xu, J., Vuorre, M., Siegler, R., Wiliam, D., & Bjork, R. A. (2025). Learning from errors versus explicit instruction in preparation for a test that counts. British Journal of Educational Psychology, 95(1), 11-25.

Myers, D. G., & DeWall, C. N. (2018). Psychology (Twelfth edition). Worth Publishers.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, *51*(2), 102-116.

Nelson T.O. & Narens L. (1990). Metamemory: a theoretical framework and new findings. In The Psychology of Learning and Motivation, vol. 26, ed. GH Bower, pp. 125–73. New York: Academic

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. The Cambridge handbook of multimedia learning, 27, 27-42.

Park, O., Lee, J.: Adaptive instructional systems. In: Jonassen, D.H. (ed.) Handbook of Research on Educational Communications and Technology, pp. 651–684. Lawrence Erlbaum Associates Publishers, Mahwah (2004).

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning Styles: Concepts and Evidence. *Psychological Science in the Public Interest, 9*(3), 105-118)

Putnam, A. L., Sungkhasettee, V. W., & Roediger, H. L. (2016). Optimizing learning in college: Tips from cognitive psychology. Perspectives on Psychological Science, 11, 652-660.

Pridemore, D. R., & Klein, J. D. (1991). Control of feedback in computer-assisted instruction. *Educational Technology Research and Development*, *39*(4), 27-32.

The NCES Fast Facts Tool provides quick answers to many education questions (National Center for Education Statistics). (n.d.). National Center for Education Statistics. Retrieved September 8, 2025, from https://nces.ed.gov/fastfacts/display.asp?id=80

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. *The Oxford handbook of metamemory*, *1*, 65-80.

Rhodes, M. G. (2019). Metacognition. *Teaching of Psychology*, *46*(2), 168-175.

Rhodes M.G. & Castel A.D. (2008). Memory predictions are influenced by perceptual information: evidence for metacognitive illusions. *Journal of Experimental Psychology*, *137*. 615–25

Rhodes, M. G. & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, *16*(3), 550-554.

Rhodes, M. G., Sitzman, D. M., & Rowland, C. A. (2013). *Monitoring and control of learning own-race and other-race faces*. *Applied Cognitive Psychology, 27*(5), 553–563.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological science, 17(3), 249-255.

Rowe, A. D., & Wood, L. N. (2008). Student perceptions and preferences for feedback. *Asian social science*, *4*(3), 78-88.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological science*, *3*(4), 207-218.

Sitzman, D. M., Rhodes, M. G., & Kornell, N. (2016). The influence of feedback on predictions of future memory performance. *Memory & Cognition*, *44*(7), 1102-1113.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78* , 153–189.

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science, 10(2)*, 176-199.

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. Computers in Human Behavior, 27(1), 118-130.

VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist, 46(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Vygotsky, L. S. (l987) The collected works of L. S. Vygotsky, Volume 1. Problems of general psychology including the volume thinking and speech. R. W. Rieber & A. S. Carton (Eds.), New York: Plenum Press.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449-455.

Wickens, C. D. (2021). Attention: theory, principles, models and applications. *International Journal of Human–Computer Interaction*, 37(5), 403-417.

Appendix A: Learning Phase Passages

P1 - The Cerebral Cortex:

The cerebral cortex is the body's ultimate control and information-processing center. It can be divided into four pairs of lobes: the frontal lobes, the parietal lobes, the occipital lobes, and the temporal lobes. The frontal lobes are involved in judgment and making plans. The parietal lobes are involved in receiving sensory input for touch and body position. The occipital lobes are involved in receiving information from our visual fields. The temporal lobes are involved in processing auditory information by receiving sensory input from the ears.

P2 - Operant Conditioning:

In operant conditioning, reinforcement is any event that strengthens the behavior it follows. Positive reinforcement increases behaviors by presenting something desirable. Negative reinforcement increases behaviors by stopping or reducing something undesirable. Punishment is an event that tends to decrease the behavior that it follows. Positive punishment decreases behaviors by presenting something undesirable, while negative punishment decreases behaviors by stopping or reducing something desirable.

P3 - Group Behavior:

Group behavior can impact our own behavior in a variety of circumstances. The bystander effect demonstrates that when many other people are present, the likelihood that an individual will help someone in need decreases. Social loafing is the tendency for people in a group to exert less effort when pooling their efforts toward attaining a common goal than when individually accountable. Group polarization is the enhancement of a group's prevailing inclinations through discussion within the group. Groupthink is the mode of thinking that occurs when the desire for harmony in a decision-making group overrides a realistic appraisal of alternatives.

<u>P4 - The Modal Model of Memory:</u>

Atkinson and Shiffrin (1968) proposed a modal model of memory with three stages. We first record to-be-remembered information as a fleeting sensory memory. Sensory memory is the explanatory, very brief recording of sensory information in the memory system. From there, we process information into short-term memory, where we encode it through rehearsal. Short-term memory is activated memory that holds a few items briefly, such as the seven digits of a phone number while calling, before the information is stored or forgotten. Finally, information moves into long-term memory for later retrieval. Long-term memory is the relatively permanent and limitless storehouse of the memory system, including knowledge, skills, and experiences.

<u>P5 - The Big 5 Personality Factors:</u>

The Big Five personality factors are conscientiousness, agreeableness, neuroticism, openness, and extraversion. These dimensions can be objectively measured, are relatively stable across the lifespan, and apply across cultures. A very conscientious person would be organized, careful, and disciplined. A very agreeable person would be soft-hearted, trusting, and helpful. A very neurotic person would be anxious, insecure, and self-pitying. A very open person would be imaginative, independent, and prefer variety. A very extraverted person would be sociable, fun-loving, and affectionate.

P6 - Piaget's' Stages of Cognitive Development:

Jean Piaget proposed that children develop cognitively through a series of stages: *Sensorimotor stage* (birth–2 years): infants know the world mostly in terms of their sensory impressions and motor activities. *Preoperational stage* (2–6/7 years): children learn to use language but do not yet comprehend the mental operations of concrete logic. *Concrete operational stage* (7–11 years): children gain the mental operations that allow them to think logically about concrete events. *Formal operational stage* (12 years–adulthood): people begin to think logically about abstract concepts.

P7 - Research Methods:

Descriptive research methods are used to observe and record behavior using case studies, naturalistic observations, or surveys. Correlational research methods are used to detect naturally occurring relationships and to assess how well one variable predicts another. Experimental research methods are used to explore cause-and-effect relationships between factors by manipulating an independent variable (i.e., the factor whose effect is being studied).

P8 - Neurons:

A neuron, or nerve cell, is the basic building block of the nervous system. A neuron contains dendrites, which are branching extensions that receive messages. Dendrites conduct neural impulses toward the cell body of the neuron. From there, the neural impulse travels down the axon, which passes messages away from the cell body through its terminal branches. The terminal branches of an axon form junctions with other cells and pass the neural impulse on to other neurons or to muscles or glands.

P9 - Depth Perception:

Depth perception enables us to estimate an object's distance from us. We often depend on certain monocular cues (i.e., depth cues that only require one eye) when judging distance. *Relative height*: objects higher in our field of vision are perceived as farther away. *Relative size*: smaller objects are perceived as being farther away. *Interposition*: if one object partially blocks our view of another, the blocking object is perceived as closer. *Linear perspective*: parallel lines are perceived as converging in the distance.

P10 - Anxiety:

Anxiety disorders are marked by distressing, persistent anxiety and often dysfunctional anxiety-reducing behaviors. Generalized anxiety disorder is a condition in which a person is continually tense, apprehensive, and uneasy. Panic disorder is marked by unpredictable, minutes-long episodes of intense dread in which a person experiences terror and accompanying chest pain, choking, or other frightening sensations. A phobia is characterized by a persistent, irrational fear and avoidance of a specific object, activity, or situation. Obsessive-compulsive disorder (OCD) is a condition characterized by unwanted repetitive thoughts (obsessions), actions (compulsions), or both.

Appendix B: Study Quizzes / Posttest Questions

<u>SQ1 - The Cerebral Cortex:</u>

1. Which lobes in the cerebral cortex are responsible for processing auditory information?
    a. Frontal
    b. Temporal
    c. Parietal
    d. Occipital

2. Which lobes of the cerebral cortex receive sensory input for receiving information from our visual fields?
    a. Temporal
    b. Parietal
    c. Frontal
    d. Occipital

<u>SQ2 - Operant Conditioning:</u>

1. _______________increases behaviors by presenting something desirable.
    a. Negative reinforcement
    b. Positive reinforcement
    c. Negative punishment
    d. Positive punishment

2. Which operant conditioning term refers to decreasing behaviors by stopping or reducing something desirable?
    a. Negative reinforcement
    b. Positive reinforcement
    c. Negative punishment
    d. Positive punishment

SQ3 - Group Behavior:

1. What is it called when there is an enhancement of a group's prevailing inclinations through discussion within the group?
    a. Group polarization
    b. Groupthink
    c. The bystander effect
    d. Social loafing
2. The phenomenon in which individuals exert less effort when working in a group toward a common goal compared to working alone is known as:
    a. Groupthink
    b. Group polarization
    c. Social loafing
    d. The bystander effect

SQ4 - The Modal Model of Memory:

1. We would have difficulty remembering a series of twelve digits for more than

several seconds, but we can remember seven-digit phone numbers when we are about to call someone thanks to the capacity of our:

a. Sensory memory

b. Short-term memory

c. Long-term memory

d. Rehearsal

2. Which type of memory is the explanatory and very brief recording of sensory information in the memory system?

a. Sensory memory

b. Short-term memory

c. Long-term memory

d. Working memory

SQ5 - The Big 5 Personality Factors:

1. ________________is a personality dimension related to being soft-hearted, trusted, and helpful.

a. Conscientiousness

b. Agreeableness

c. Neuroticism

d. Openness

2. Which personality dimension is related to being sociable, fun-loving, and affectionate?

a. Conscientiousness

b. Agreeableness
c. Neuroticism
d. Extraversion

SQ6 - Piaget's' Stages of Cognitive Development:

1. Lee can logically understand abstract concepts, such as freedom. Lee is likely in the____ stage of cognitive development.
    a. Sensorimotor
    b. Preoperational
    c. Concrete operational
    d. Formal operational
2. Alyssa's daughter can think logically about things she can see, touch, or experience directly. According to Piaget, what stage of development is Alyssa's daughter likely in?
    a. Sensorimotor
    b. Preoperational
    c. Concrete operational
    d. Formal operational

SQ7 - Research Methods:

1. Which type of research method is used to detect naturally occurring relationships, and how well one variable predicts another?
    a. Descriptive
    b. Correlational

c. Experimental

d. Independent variable

2. ________________research methods are used to explore cause and effect relationships between factors.

a. Descriptive

b. Correlational

c. Experimental

d. Independent variable

SQ8 - Neurons:

1. Which part of a neuron receives messages from other cells?

a. Dendrites

b. Cell body

c. Axon

d. Terminal branches

2. The _________________of a neuron form junctions with other cells and pass the neural impulse on to other neurons or to muscle glands.

a. Dendrites

b. Cell body

c. Axon

d. Terminal branches

SQ9 - Depth Perception:

1. Which monocular depth cue refers to our tendency to perceive objects higher in our file of vision as farther away?
    a. Linear perspective
    b. Interposition
    c. Relative height
    d. Relative size
2. If you stand on train tracks and look towards the horizon they will appear to meet in the distance. Which monocular depth cue does this represent?
    a. Linear perspective
    b. Interposition
    C. Relative height
    d. Relative size

<u>SQ10 - Anxiety:</u>

1. Which anxiety disorder is characterized by continually tense, apprehensive, and uneasy feelings?
    a. Generalized anxiety disorder
    b. Panic disorder
    c. A phobia
    d. Obsessive-compulsive disorder
2. ________________ is an anxiety disorder marked by unpredictable, minutes-long episodes of intense dread in which a person experiences terror and accompanying chest pain, choking, or other frightening sensations.

a. Generalized anxiety disorder

b. Panic disorder

c. A phobia

d. Obsessive-compulsive disorder

Appendix C: Transfer Test Questions

Transfer Test 1 – Cerebral Cortex:

1. Damage to which lobes of the cerebral cortex would most likely result in difficulty understanding spoken language?
    a. Frontal
    b. Temporal
    c. Occipital
    d. Parietal
2. After a head injury, a patient begins having trouble recognizing objects that they see, even though their eyes are healthy. Damage to which lobes is most likely responsible for this difficulty?
    a. Frontal
    b. Temporal
    c. Occipital
    d. Parietal

Transfer Test 2 – Operant Conditioning:

1. When you start your car, it beeps at you loudly until you fasten your seat belt. By fastening your seat belt, you end the loud beeping. Which operant conditioning term does this demonstrate?
    a. Negative reinforcement
    b. Positive reinforcement
    c. Negative punishment

    d. Positive punishment

2. A parent gives their teenager extra chores after they break curfew, making the teen less likely to break curfew in the future. This is an example of:
    a. Negative reinforcement
    b. Positive reinforcement
    c. Positive punishment
    d. Negative punishment

Transfer Test 3 – Group Behavior:

1. Students who oppose the death penalty meet on their university's campus for a weekly discussion with like-minded peers. After several meetings the group's stance on the death penalty grows more extreme, such that they are all now vehemently opposed to the death penalty rather than just being slightly against it. The change in their behavior is an example of
    a. Group polarization
    b. Groupthink
    c. The bystander effect
    d. Social loafing
2. During a group project, several members contribute very little effort because they assume others will do the work. This behavior is known as:
    a. Groupthink
    b. The bystander effect
    c. Group polarization

d. Social loafing

Transfer Test 4 – Modal Model of Memory:

1. You park your car in a public parking lot and are required to enter your license plate information at the meter when paying. While walking to the meter you repeat your license plate number to yourself in your head, so that you can remember what it is when typing the digits in several seconds later. In this example you are storing your license plate information in
   a. Sensory memory
   b. Short-term memory
   c. Long-term memory
   d. Rehearsal
2. The brief moment when a flash of lightning leaves a quick visual impression in your mind before fading is an example of:
   a. Sensory memory
   b. Short-term memory
   c. Long-term memory
   d. Rehearsal

Transfer Test 5 – Big Five Personality Factors:

1. Someone who is anxious, insecure, and tends to feel self-pity would score high on which Big Five personality trait?

   a. Neuroticism

b. Agreeableness

c. Extraversion

d. Conscientiousness

2. Someone who is highly sociable, energetic, and enjoys being around others would likely score high on which Big Five personality trait?

a. Neuroticism

b. Agreeableness

c. Extraversion

d. Conscientiousness

Transfer Test 6 – Piaget's Stages of Development:

1. A child can now mentally perform logical operations about real-world objects, such as understanding that the amount of water stays the same when poured into a differently shaped glass. This ability is typical of the ___ stage of cognitive development.

a. Sensorimotor

b. Preoperational

c. Concrete operational

d. Formal operational

2. Teenagers begin to reason about hypothetical situations and abstract concepts such as justice or morality during the ___ stage of cognitive development.

a. Sensorimotor

b. Preoperational

c. Concrete operational

d. Formal operational

<u>Transfer Test 7 – Research Methods:</u>

1. Dr. Ramos is interested in whether packaging color affects buyers' likelihood of purchasing different flavors of ice cream. She runs a study in which some ice cream brands are packaged in red containers and others are packaged in blue containers to see if one color leads to higher levels of purchasing. What kind of study is Dr. Ramos doing?
    a. Descriptive
    b. Correlational
    c. Experimental
    d. Independent variable
2. Which type of research method would best allow a psychologist to determine whether two naturally occurring variables, like hours of sleep and test scores, are related?
    a. Descriptive
    b. Correlational
    c. Experimental
    d. Survey-only

<u>Transfer Test 8 – Neurons:</u>

1. You accidentally place your hand on a hot surface and immediately pull it away

quickly, thanks to the neurons in your hand that alerted you to the pain. Which part of a neuron would have gotten the information that the surface was hot and then passed it on to the rest of your nervous system?

   a. **Dendrites**
   b. Cell body
   c. Axon
   d. Terminal branches

2. After a neural impulse travels down the axon, which part of the neuron passes the signal on to other cells?
   a. Dendrites
   b. Cell body
   c. Axon
   d. **Terminal branches**

Transfer Test 9 – Depth Perception:

1. A farmer stares at her parallel rows of corn and sees that they appear to meet in the distance. Which monocular depth cue does this represent?
   a. **Linear perspective**
   b. Interposition
   c. Relative height
   d. Relative size
2. When two similar objects appear at different heights in your visual field and the higher one appears farther away, this demonstrates:

a. Relative size

b. Linear perspective

c. Relative height

d. Interposition

<u>Transfer Test 10 – Anxiety:</u>

1. While driving home from work, Pam suddenly felt intense dread, accompanied by trembling, dizziness, chest pains, and choking sensations. Pam may have:

    a. Generalized anxiety disorder

    b. Panic disorder

    c. Phobia

    d. Obsessive-compulsive disorder

2. A person constantly worries about many different everyday events, feeling tense and uneasy most of the time. This pattern is characteristic of:

    a. Panic disorder

    b. Generalized anxiety disorder

    c. Obsessive-compulsive disorder

    d. Phobia

Appendix D: Judgement of Learning Questions

Experiment 1:

1. You will later receive 6 questions related to these topics. How many do you predict you will answer correctly? (Range: 1-6)

Experiments 2 and 3:

1. You will later be tested on the same set of [4 or 6] questions you previously studied. How many do you predict you will answer correctly? (Range 1-4 or 6)

2. You will later be tested on [4 or 6] questions related to the content you just learned. The questions may be phrased differently than the ones you saw earlier, but they relate to the same topics. How many do you predict you will answer correctly? (Range 1-4 or 6)

Appendix E: NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988)

Please rate your experience on the following scales **based on the block you just completed**, during which you received **Explanatory/Delayed/Outcome/Explanatory Feedback**.

1. Mental Demand: How mentally demanding was the task?
2. Temporal Demand: How hurried or rushed was the pace of the task?
3. Performance: How successful were you in accomplishing what you were asked to do?
4. Effort: How hard did you have to work to accomplish your level of performance?
5. Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?

Note: The Physical Demand subscale of the NASA-TLX (Hart & Staveland, 1988) is excluded for the present studies because the computer-based task requires only minimal physical effort. Removing this subscale ensures workload ratings focused on the cognitive and perceptual demands most relevant to the experimental manipulations.

Appendix F: Distractor Task

On the following screens, you will see some math problems. Please try to solve them as quickly as possible.

You will have 20 seconds to answer each question, and then the program will advance to the next question.

Set 1:
(2 + 7) × 9 - 4
(9 + 3) + 12 × 2
(10 + 22 - 2) × 2
(13 + 5) × (7 - 5)
(3 + 8) × 9 - 3
5 × 5 + (4 + 6)
(15 + 20 - 3) × 1
(3 + 7) × (13 + 5)
7 × 10 + (10 - 10)

Set 2:
(4 + 6) × 8 - 5
(12 + 8) + 6 × 3
(9 + 15 - 4) × 2
(11 + 9) × (6 - 3)
(5 + 7) × 7 - 8
6 × 4 + (5 + 9)
(20 + 18 - 6) × 1
(4 + 8) × (10 + 2)
8 × 9 + (6 - 6)

Set 3:
(7 + 5) × 6 - 9
(14 + 3) + 10 × 2
(8 + 22 - 5) × 2
(9 + 4) × (5 - 2)
(6 + 9) × 8 - 7
7 × 7 + (3 + 5)
(18 + 25 - 7) × 1
(5 + 6) × (9 + 8)
10 × 6 + (9 - 9)

Set 4:
(3 + 9) × 5 - 6
(10 + 4) + 11 × 3
(12 + 20 - 8) × 2
(7 + 3) × (8 - 5)
(9 + 6) × 7 - 2
8 × 6 + (2 + 4)
(25 + 15 - 5) × 1
(6 + 8) × (12 + 3)
9 × 9 + (8 - 8)

Appendix G: Training Perceptions Exit Survey

Please respond to your level of agreement for each statement with 1 being “Strongly Disagree” and 7 being “Strongly Agree.”

1. Overall, I liked the content of this training
2. I believe that the feedback I received focused my attention on learning strategies to perform this task better.
3. The difficulty of questions was appropriate for my skill level.
4. The system customized my training based on my answers.
5. The system was able to accurately assess my performance.
6. I believe the system accurately accommodated my feedback preferences.

Appendix H: Perceived Workload by Feedback Condition

Experiment 1:

Table H1. Experiment 1: NASA-TLX Ratings Across Feedback Timing Conditions

Subscale	Immediate Feedback	Delayed Feedback			
	M (*SD*)	*M* (*SD*)	$t_{(80)}$	*p*	Cohen's *d*
Mental	11.15 (5.44)	11.17(5.63)	-0.05	.959	-0.01
Temporal	6.44 (5.19)	6.59 (5.40)	-0.34	.736	-0.04
Performance	13.94 (5.28)	14.37 (5.22)	-1.42	.158	-0.16
Effort	11.33 (5.50)	11.32 (5.84)	0.03	.977	0.00
Frustration	5.25 (5.33)	5.04 (5.10)	0.48	.634	0.05

Note. Mean parameter values for each of the analyses are shown for the immediate feedback and delayed feedback conditions, as well as the results of paired samples t-tests comparing the parameter estimates between the feedback conditions.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Experiment 2:

Table H2. Experiment 2: NASA-TLX Ratings Across Feedback Timing Conditions for Blocks 1-2

Subscale	Immediate Feedback	Delayed Feedback			
	M (*SD*)	*M* (*SD*)	$t_{(111)}$	*p*	Cohen's *d*
Mental	10.54 (5.15)	9.02 (5.08)	3.53	$< .001^{***}$	0.33
Temporal	5.54 (4.81)	5.07 (4.50)	1.34	.171	0.13
Performance	15.19 (4.92)	17.02 (3.99)	-3.84	$< .001^{***}$	0.36
Effort	11.41 (4.84)	10.54 (5.16)	1.91	.059	0.18
Frustration	5.25 (5.30)	3.98 (3.89)	3.03	$.003^{**}$	0.29

Note. Mean parameter values for each of the analyses are shown for the immediate feedback and delayed feedback conditions, as well as the results of paired samples t-tests comparing the parameter estimates between the feedback conditions.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table H3. Experiment 2: NASA-TLX Ratings Across Feedback Timing Conditions for Blocks 3-4

Subscale	Immediate Feedback	Delayed Feedback			
	M (SD)	M (SD)	$t_{(111)}$	p	Cohen's d
Mental	12.09 (5.57)	12.47 (5.57)	-1.83	.070	0.17
Temporal	6.20 (5.40)	6.53 (5.42)	-1.47	.143	0.14
Performance	15.53 (4.33)	14.34 (4.72)	3.26	.001**	0.31
Effort	12.60 (5.36)	12.86 (5.42)	-1.01	.316	0.10
Frustration	5.97 (5.33)	6.46 (5.50)	-1.50	.135	0.14

Note. Mean parameter values for each of the analyses are shown for the immediate feedback and delayed feedback conditions, as well as the results of paired samples t-tests comparing the parameter estimates between the feedback conditions.
*$p < .05$, **$p < .01$, ***$p < .001$

Experiment 3:

Table H4. Experiment 3: NASA-TLX Ratings Across Feedback Content Conditions for Blocks 1-2

Subscale	Explanatory Feedback	Outcome Feedback			
	M (SD)	M (SD)	$t_{(105)}$	p	Cohen's d
Mental	9.58 (5.17)	10.83 (4.96)	-3.00	.003**	0.29
Temporal	6.61 (5.21)	6.46 (5.26)	0.39	.698	0.04
Performance	16.90 (4.09)	15.40 (4.64)	3.30	.001**	0.32
Effort	10.62 (5.14)	11.13 (5.19)	-1.19	.236	0.12
Frustration	4.25 (4.62)	5.51 (5.04)	-3.19	.002**	0.31

Note. Mean parameter values for each of the analyses are shown for the immediate feedback and delayed feedback conditions, as well as the results of paired samples t-tests comparing the parameter estimates between the feedback conditions.
*$p < .05$, **$p < .01$, ***$p < .001$

Table H5. Experiment 3: NASA-TLX Ratings Across Feedback Content Conditions for Blocks 3-4

Subscale	Explanatory Feedback	Outcome Feedback			
	M (*SD*)	*M* (*SD*)	$t_{(105)}$	*p*	Cohen's *d*
Mental	13.22 (4.90)	12.79 (5.07)	1.52	.132	0.15
Temporal	7.35 (5.81)	7.56 (5.94)	-0.90	.371	0.09
Performance	14.42 (4.78)	15.19 (4.49)	-1.84	.068	0.18
Effort	12.91 (5.05)	12.54 (5.31)	1.47	.144	0.14
Frustration	6.56 (5.68)	6.06 (5.26)	1.36	.178	0.13

Note. Mean parameter values for each of the analyses are shown for the immediate feedback and delayed feedback conditions, as well as the results of paired samples t-tests comparing the parameter estimates between the feedback conditions.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Appendix I: Perceived Workload by Preference Honoring Condition

Experiment 1:

Table I1. Experiment 1: NASA-TLX Ratings for Immediate Preference Group

Subscale	Immediate Preference (n = 76)				
	Honored	Dishonored			
	M (SD)	M (SD)	$t_{(75)}$	p	Cohen's d
Mental	11.29 (5.42)	11.32 (5.55)	-0.05	.959	0.00
Temporal	6.57 (5.18)	6.86 (5.46)	-0.644	.512	0.07
Performance	13.64 (5.30)	14.20 (5.30)	-1.75	.085	0.02
Effort	11.49 (5.38)	11.46 (5.71)	0.05	.955	0.00
Frustration	5.29 (5.13)	5.04 (4.89)	0.53	.595	0.06

Note. Mean parameter values for each of the analyses are shown for participants who indicated a preference for immediate feedback only. Results of paired samples t-tests comparing the parameter estimates between the feedback honoring conditions.
*p < .05, **p < .01, ***p < .001

Table I2. Experiment 1: NASA-TLX Ratings for Delayed Preference Group

Subscale	Delayed Preference (n = 5)	
	Honored	Dishonored
	M (SD)	M (SD)
Mental	9.00 (7.07)	9.00 (5.92)
Temporal	2.60 (1.82)	4.60 (5.46)
Performance	17.00 (3.08)	18.40 (2.07)
Effort	9.20 (8.04)	9.00 (7.42)
Frustration	5.00 (8.49)	4.60 (8.65)

Note. Descriptive values are shown for those who indicated a preference for delayed feedback only. Due to the small sample size included in this group, inferential comparisons were not conducted.

Experiment 2:

Table I3. Experiment 2: NASA-TLX Ratings for Immediate Preference Group Across Blocks 1-2

Subscale	Immediate Preference (n = 103)				
	Honored	Dishonored			
	M (*SD*)	*M* (*SD*)	$t_{(102)}$	*p*	Cohen's *d*
Mental	10.49 (5.16)	9.02 (5.07)	3.21	.002**	0.32
Temporal	5.24 (4.75)	4.86 (4.42)	1.08	.283	0.11
Performance	15.31 (4.85)	16.87 (4.08)	-3.21	.002**	0.32
Effort	11.52 (4.76)	10.50 (5.11)	-2.13	.036*	0.21
Frustration	5.33 (5.45)	4.17 (3.96)	2.62	.010*	0.26

Note. Mean parameter values for each of the analyses are shown for participants who indicated a preference for immediate feedback only. Results of paired samples t-tests comparing the parameter estimates between the feedback honoring conditions.
*p < .05, **p < .01, ***p < .001

Table I4. Experiment 2: NASA-TLX Ratings for Delayed Preference Group Across Blocks 1-2

Subscale	Delayed Preference (n = 11)	
	Honored	Dishonored
	M (*SD*)	*M* (*SD*)
Mental	9.27 (5.33)	11.09 (4.76)
Temporal	6.55 (4.93)	8.00 (4.54)
Performance	18.73 (2.00)	14.00 (5.20)
Effort	11.45 (5.75)	10.73 (5.50)
Frustration	1.82 (1.78)	4.00 (2.90)

Note. Descriptive values are shown for those who indicated a preference for delayed feedback only. Due to the small sample size included in this group, inferential comparisons were not conducted.

Table I5. Experiment 2: NASA-TLX Ratings for Immediate Preference Group Across Blocks 3-4

Subscale	Immediate Preference (n = 102)				
	Honored	Dishonored			
	M (*SD*)	*M* (*SD*)	$t_{(100)}$	*p*	Cohen's *d*
Mental	12.33 (5.48)	12.80 (5.36)	-2.20	.030*	0.22
Temporal	6.28 (5.39)	6.66 (5.41)	-1.53	.130	0.15
Performance	15.41 (4.41)	14.28 (4.70)	2.86	.005**	0.28
Effort	12.91 (5.16)	13.08 (5.23)	-0.62	.537	0.06
Frustration	6.01 (5.21)	6.55 (5.47)	-1.59	.116	0.16

Note. Mean parameter values for each of the analyses are shown for participants who indicated a preference for immediate feedback only. Results of paired samples t-tests comparing the parameter estimates between the feedback honoring conditions.
*p < .05, **p < .01, ***p < .001

Table I6. Experiment 2: NASA-TLX Ratings for Delayed Preference Group Across Blocks 3-4

Subscale	Delayed Preference (n = 12)	
	Honored	Dishonored
	M (*SD*)	*M* (*SD*)
Mental	10.00 (6.47)	10.42 (6.05)
Temporal	5.00 (5.15)	5.08 (5.20)
Performance	14.42 (5.04)	16.67 (3.14)
Effort	11.25 (6.59)	10.17 (6.35)
Frustration	5.00 (5.64)	5.08 (6.27)

Note. Descriptive values are shown for those who indicated a preference for delayed feedback only. Due to the small sample size included in this group, inferential comparisons were not conducted.

Experiment 3:

Table I7. Experiment 3: NASA-TLX Ratings for Explanatory Preference Group Across Blocks 1-2

Subscale	Explanatory Preference (n = 93) Honored	Explanatory Preference (n = 93) Dishonored			
	M (*SD*)	*M* (*SD*)	$t_{(92)}$	*p*	Cohen's *d*
Mental	9.43 (5.09)	10.69 (4.91)	-2.72	.008**	0.28
Temporal	6.46 (5.17)	6.24 (5.15)	0.56	.553	0.06
Performance	16.91 (4.21)	15.57 (4.62)	2.77	.007**	0.29
Effort	10.46 (5.11)	11.19 (5.23)	-1.62	.108	0.17
Frustration	3.94 (4.56)	5.23 (4.90)	-3.42	.001**	0.35

Note. Mean parameter values for each of the analyses are shown for participants who indicated a preference for explanatory feedback only. Results of paired samples t-tests comparing the parameter estimates between the feedback honoring conditions.
*p < .05, **p < .01, ***p < .001

Table I8. Experiment 3: NASA-TLX Ratings for Outcome Preference Group Across Blocks 1-2

Subscale	Outcome Preference (n = 11) Honored	Outcome Preference (n = 11) Dishonored
	M (*SD*)	*M* (*SD*)
Mental	11.85 (5.37)	10.62 (5.87)
Temporal	8.08 (5.92)	7.69 (5.59)
Performance	14.15 (4.85)	16.77 (3.19)
Effort	10.69 (5.09)	11.77 (5.43)
Frustration	7.54 (5.77)	6.54 (4.54)

Note. Descriptive values are shown for those who indicated a preference for outcome feedback only. Due to the small sample size included in this group, inferential comparisons were not conducted.

Table I9. Experiment 3: NASA-TLX Ratings for Explanatory Preference Group Across Blocks 3-4

Subscale	Explanatory Preference (n = 89)				
	Honored	Dishonored			
	M (SD)	M (SD)	$t_{(88)}$	p	Cohen's d
Mental	13.00 (5.01)	12.67 (5.22)	1.03	.305	0.11
Temporal	7.29 (5.89)	7.56 (6.06)	-1.04	.301	0.11
Performance	14.52 (4.93)	15.03 (4.68)	-1.14	.258	0.12
Effort	12.81 (5.08)	12.43 (5.37)	1.35	.180	0.14
Frustration	6.33 (5.80)	5.94 (5.56)	1.03	.306	0.11

Note. Mean parameter values for each of the analyses are shown for participants who indicated a preference for explanatory feedback only. Results of paired samples t-tests comparing the parameter estimates between the feedback honoring conditions.
*$p < .05$, **$p < .01$, ***$p < .001$

Table I10. Experiment 3: NASA-TLX Ratings for Outcome Preference Group Across Blocks 3-4

Subscale	Outcome Preference (n = 12)	
	Honored	Dishonored
	M (SD)	M (SD)
Mental	13.41 (4.27)	14.35 (4.23)
Temporal	7.53 (5.42)	7.65 (5.51)
Performance	16.00 (3.32)	13.88 (3.98)
Effort	13.12 (5.12)	13.41 (4.98)
Frustration	6.65 (3.37)	7.76 (4.99)

Note. Descriptive values are shown for those who indicated a preference for outcome feedback only. Due to the small sample size included in this group, inferential comparisons were not conducted.

Appendix J: Perceptions of Training

Table J1. Experiment 1: Descriptive Statistics for Perceptions of Training Questionnaire

	M	*SD*	*Min*	*Max*
Overall, I liked the content of this training	6.02	0.97	3	7
I believe that the feedback I received focused my attention on learning strategies to perform this task better.	5.79	1.33	1	7
The difficulty of questions was appropriate for my skill level.	5.67	1.34	2	7
The system customized my training based on my answers.	4.36	1.59	1	7
The system was able to accurately assess my performance.	5.63	1.12	2	7
I believe the system accurately accommodated my feedback preferences.	4.91	1.35	2	7

Table J2. Experiment 2: Descriptive Statistics for Perceptions of Training Questionnaire

	M	*SD*	*Min*	*Max*
Overall, I liked the content of this training	5.74	1.06	1	7
I believe that the feedback I received focused my attention on learning strategies to perform this task better.	5.63	1.09	2	7
The difficulty of questions was appropriate for my skill level.	5.78	1.01	2	7
The system customized my training based on my answers.	3.97	1.69	1	7
The system was able to accurately assess my performance.	5.85	1.06	2	7
I believe the system accurately accommodated my feedback preferences.	5.08	1.52	1	7

Table J3. Experiment 3: Descriptive Statistics for Perceptions of Training Questionnaire

	M	*SD*	*Min*	*Max*
Overall, I liked the content of this training	5.65	1.14	1	7
I believe that the feedback I received focused my attention on learning strategies to perform this task better.	5.41	1.39	1	7
The difficulty of questions was appropriate for my skill level.	5.72	1.15	1	7
The system customized my training based on my answers.	4.34	1.44	1	7
The system was able to accurately assess my performance.	5.42	1.14	1	7
I believe the system accurately accommodated my feedback preferences.	5.13	1.34	2	7