IMPACT OF CONTENT DOMAIN AND PREFERENCE ON PEDAGOGICAL AGENT EFFECTIVENESS

A Thesis
Presented to
The Academic Faculty

by

Reganne McKenna Miller

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
College of Sciences
School of Psychology

Georgia Institute of Technology
December 2025

IMPACT OF CONTENT DOMAIN AND PREFERENCE ON PEDAGOGICAL AGENT EFFECTIVENESS

Approved by:

Dr. Richard Catrambone, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Bruce Walker
School of Psychology
Georgia Institute of Technology

Dr. Mengyao Li
School of Psychology
Georgia Institute of Technology

Date Approved: [September 5, 2025]

Table Of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial intelligence
ANOVA	Analysis of variance
ANCOVA	Analysis of covariance
APA	Animated pedagogical agent
API	Agent Persona Instrument (Ryu & Baylor, 2005)
API-R	Agent Persona Instrument Revised (Schroeder et al., 2018)
CLT	Cognitive Load Theory (Sweller, 1988; Sweller et al., 1998)
IMMS	Instructional Materials Motivation Survey (Keller, 2010)
NASA-TLX	NASA Task Load Index (Hart & Staveland, 1988)
PA	Pedagogical agent
RIMMS	Reduced Instructional Materials Motivation Survey (Loorbach et al., 2015)
TTS	Text-to-speech
WPM	Words per minute
VIF	Variance inflation factor

SUMMARY

Pedagogical agents (PAs) are virtual, typically anthropomorphic, characters used in online learning environments to improve users' learning outcomes. They have been shown to increase learning and motivation and are also suggested to reduce cognitive load. Despite the widespread use and investigation of PAs, questions remain. One is: does a learner who prefers not to learn with a PA reap the same pedagogical benefits as one who does? To explore this, the effects of matching versus mismatching PA usage with learners' preferences were analyzed. Additionally, whether PA effectiveness varies across content domains (examined through mathematics and art) was investigated. This study comprises two experiments. Experiment 1 used a 2 (PA Assignment: With-PA vs. No-PA) by 2 (Lesson: Math vs. Art) mixed design to assess immediate learning, retention, motivation, subjective workload, PA persona ratings, and opinions. Experiment 1 was designed to (a) evaluate the PA's effectiveness prior to Experiment 2, and (b) compare its impact in a math lesson versus an art lesson. Experiment 2 used a 2 (PA Preference: Prefers vs. Does Not Prefer) by 2 (PA Assignment: With-PA vs. No-PA) between-subjects design. The same dependent variables (excluding retention) were used to analyze the effects of preference-assignment congruence. No hypotheses were supported, and, notably, the main effect of PA Assignment was nonsignificant across both experiments. Significant main effects of Lesson (Experiment 1) and PA Preference (Experiment 2) were found. Implications, limitations, and avenues for future research are discussed.

Keywords: Pedagogical agent, preference, math, humanities, learning, motivation, cognitive load, workload, PA persona

CHAPTER 1. INTRODUCTION

The rapid growth of digital technologies throughout the last few decades has led to the emergence of various tools designed to improve educational outcomes, often referred to as educational technologies. These technologies, ranging from online courses and educational apps to augmented and virtual reality experiences, include pedagogical agents (PAs).

PAs are digital characters that assist learners in online education with the purpose of enhancing the user's experience and success (Dai et al., 2022; Schroeder & Adesope, 2015; Schroeder & Adesope, 2014). Their appearance is often that of a human cartoon avatar , although they can also be designed in a variety of nonhuman or abstract forms, and they can be utilized in various settings, such as university lessons and workplace training. Examples of PAs include Microsoft's now-retired Office Assistant, Clippy (Figure 1), an example of a nonhuman PA, and Morehouse College's new humanlike AI teaching assistants (Figure 2; Nobles, 2024).

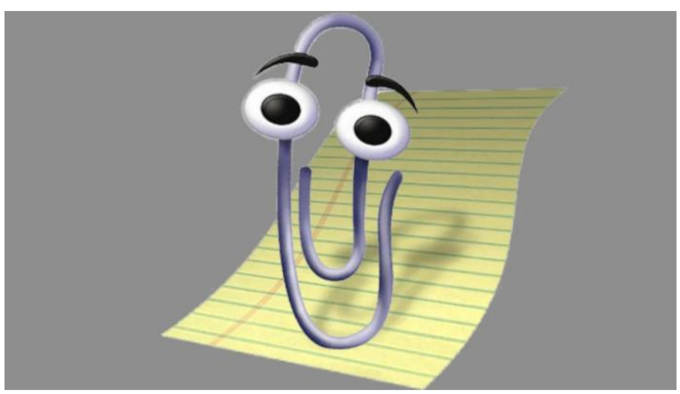


Figure 1 – Microsoft's Clippy.

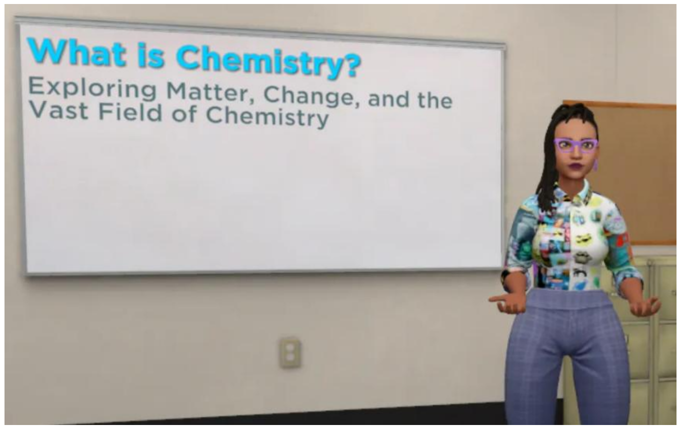


Figure 2 – One of Morehouse College's new AI teaching assistants (Nobles, 2024).
Credit: Muhsinah Morris. Image retrieved from Axios Atlanta.

Proponents of PAs highlight their benefits, such as increased learning and motivation, and decreased cognitive load (Davis et al., 2022; Dincer & Doganay, 2017; Schroeder & Adesope, 2014; Yung, 2009; Yung & Paas, 2015). However, opponents of PAs argue that the effectiveness of PAs on learning and cognitive load is overestimated or potentially negative (Clark & Choi, 2005).

Research on PAs has been extensive, spanning over twenty years of investigation (Dai et al., 2022), covering topics such as learning outcomes (Li et al., 2019; Schroeder et al., 2013), individual difference effects (e.g., based on gender, race; Ozogul et al., 2013; van der Meij et al., 2015), and PA design (Kim et al., 2007; Liew et al., 2013; Ryu & Ke, 2018). Despite the broad scope of PA research thus far, at least one important question has yet to be investigated: If users do not want to learn with pedagogical agents, what are the consequences of making them do so anyway? This is the key question that was investigated in the present study. Additionally, the differences in pedagogical agent effectiveness between the content domains of math and art were explored.

1.1 Benefits of Pedagogical Agents

Research supporting the use of PAs has suggested that they have positive effects on many variables, including learning, cognitive load, and motivation. The following sections will discuss each of these three variables in more detail.

1.1.1 Learning

PAs are widely utilized for their purported benefits in enhancing learning. However, the evidence regarding their effectiveness remains ambiguous; some studies question the educational benefits of PAs, with a few authors suggesting that they could harm learning, while others support their positive impact on learning.

Opponents of PAs argue that they could create an "information rich display" that increases cognitive load and decreases learning (Clark & Choi, 2007, p. 380). Others report no significant difference in learning outcomes between PA and no-PA conditions (Moundridou & Virvou, 2002; van Mulken et al., 1998). For example, in a 2011 meta-analysis, Heidig and Clarebout found that, out of 15 studies comparing a PA condition versus a no-PA control condition, 9 reported no differences in learning outcomes between the two groups.

As noted by Heidig and Clarebout (2011), there could be an explanation for the lack of PA learning benefits in some studies. These results could stem from the vast array of design decisions possible for a PA; these decisions include, but are certainly not limited to, the amount of visible body (e.g., full body vs. "talking head"), type of being (e.g., paperclip, parrot, robot, human), and realism (e.g., 2-dimensional cartoon, 3-dimensional cartoon, real human). Additionally, the content domain taught by the PA can vary greatly, with examples including nanotechnology (Dirkin et al., 2005) and electric motors (Mayer et al., 2003a). Thus, it is reasonable to speculate that the lack of results in some studies is at least partially a product of the design and subject matter choices made in those studies.

Proponents of PAs argue for their positive effect on user learning, as demonstrated in Moreno et al.'s 2001 study. This study used a science lesson on plant survival. College students taught by a speaking, animated PA showed improved performance on transfer problems compared to those without a PA, although no significant difference was found on retention problems. A second experiment replicated these findings with 7th graders (Moreno et al., 2001). Another example of PAs enhancing learning comes from Kizilkaya and Askar's 2008 study in which they compared a group of 6th graders that received instruction from a PA-assisted lesson to a group that received an identical lesson without the PA. Using a science achievement test to measure learning, participants in the PA condition scored higher than students in the no-PA condition.

Reflecting the above-mentioned findings, Schroeder et al. (2013) conducted a meta-analysis on 43 PA studies with a total of 3,088 participants. The studies included in the meta-analysis measured learning as recall, retention, transfer, or some combination of these types of learning. Although the authors did not report individual findings for each of these learning types, they did report a significant effect of PAs on overall learning ($g = 0.19$).

1.1.1.1 Social Agency Theory

As outlined by Schroeder et al. (2022), one explanation for the increased learning observed with PAs is Social Agency Theory (Mayer et al., 2003b). Based on the cooperation principle (Grice, 1975), this theory suggests that, due to social cues (e.g., eye gaze, voice) presented by PAs, when learners interact with PAs, social conversation schemas are elicited. These schemas prompt the use of social rules typical in human-to-human interaction (Table 1; Grice, 1975). As a result of these social rules being elicited, the listener (i.e., the learner) tries to better understand what the speaker (i.e., the PA) is communicating (Mayer et al., 2003b). This is accomplished by the learner via deep cognitive processing, which facilitates more meaningful learning.

Table 1 – Grice's (1975) Social Rules.

Rule	Description
Quantity	The speaker is expected to provide only the amount of information that is required by the conversation – no more and no less.
Quality	The speaker is expected to provide only information that they believe to be true.
Relevancy	The speaker is expected to provide only information that is relevant to the current conversation topic, which can change throughout the conversation.
Manner	The speaker is expected to be concise when communicating. This consists of being orderly, being brief, and avoiding obscurity of expression and ambiguity.

1.1.2 Cognitive Load

Cognitive Load Theory (CLT; Sweller, 1988; Sweller et al., 1998) explains how learning materials impose a cognitive load on learners and how this load influences the learning process. CLT is based on the idea that working memory is a limited resource with constraints on the number of elements it can process at a given time. If the load imposed by a learning task is too high, learning is impeded. CLT defines three types of cognitive load: intrinsic, extraneous, and germane.

Intrinsic cognitive load is related to the complexity of the learning materials and the level of element interactivity; this refers to the degree to which the elements in the material depend on each other (Chen et al., 2023; Sweller et al., 1998; Sweller et al., 2019). "Elements" refers to the pieces of information that a user is learning. If a lesson has a low level of element interactivity, this means that most or all of the elements can be learned independently from each other. If a lesson has a high level of element interactivity, this means that most or all of the elements must be learned together, as learning one element depends on understanding the other elements. Higher element interactivity imposes a higher level of intrinsic cognitive load on the learner. Aside from element interactivity, intrinsic cognitive load is also influenced by the knowledge level of the learner (Chen et al., 2023; Sweller et al., 2019); a person who is already knowledgeable on a topic will experience less load than a novice.

Extraneous cognitive load is related to how the instructional materials are presented to the learner and the difficulty of the tasks involved in learning. Thus, the level of extraneous cognitive load can be altered by changing the nature of the materials and how they are presented. Distractions in the learning environment (i.e., qualities of the learning environment that are unnecessary for learning) increase extraneous cognitive load. As advised by Sweller (2010), the goal with learning materials should always be to decrease the amount of extraneous cognitive load imposed on the user, as this should have a "beneficial or at worst, neutral effect" (p. 134). Although it was previously thought that element interactivity was related only to intrinsic cognitive load (Sweller et al., 1998), in recent updates to CLT, it was proposed that it is also related to extraneous cognitive load (Sweller et al., 2019). This is because the level of element interactivity can be reduced by presenting the learning materials in an efficient way.

The last type of cognitive load in CLT is germane cognitive load (Sweller et al., 1998; Sweller et al., 2019). Germane cognitive load refers to the cognitive resources dedicated to the processing and construction of knowledge, which are necessary steps for effective learning. The role of germane cognitive load has been updated since its description in Sweller et al.'s 1998 paper. It was previously characterized as being substituted for extraneous cognitive load, thus contributing to overall load levels. This definition has been changed to suggest that germane cognitive load actually serves a redistribution function, moving working memory resources from extraneous activities to intrinsic activities (Sweller et al., 2019).

PA-related literature typically focuses on extraneous cognitive load, as debates tend to surround whether PAs are a distraction in the learning environment, thereby increasing extraneous cognitive load. Critics of PAs, such as Clark & Choi (2005, 2007), argue that PAs increase extraneous cognitive load, potentially causing a split-attention effect where learners must divide their focus between the lesson and the PA. This "agents-as-distractors" viewpoint (Frechette & Moreno, 2010, p. 64) suggests that PAs become an unnecessary element in the learning environment. According to CLT (Sweller, 1988; Sweller et al., 1998), such non-essential elements require additional cognitive processing, possibly detracting from learning by limiting the resources available for processing intrinsic materials and overloading working memory.

Alternatively, some authors have argued that PAs are beneficial for cognitive load. While there are few articles that have definitively found that PAs lead to a reduction in cognitive load, arguments have been made in support of this idea. For example, some authors have operationalized cognitive load as perceived difficulty level. In a study by Atkinson (2002), undergraduate students watched a lesson on problem solving that either did or did not include audio and did or did not include a PA; a control condition with no audio and no PA was also included. The results showed that participants in the audio-PA condition reported significantly lower perceived example difficulty than participants in the control condition. Similar results were found in a study by van Mulken et al. (1998) where participants learned about the parts and kinematics of pulley systems. The researchers found that participants in the PA condition rated the difficulty of the lesson significantly more positively than participants in the no-PA condition.

Other authors who endorse the cognitive load benefits of PAs have argued specifically for PAs that enhance the learning environment via nonverbal cues such as deictic gestures; deictic gestures are specific pointing gestures, such as those made with an outstretched index finger (Li et al., 2019), that direct learners' attention to relevant on-screen content. This stance is sometimes called the "agents-as-complements" viewpoint (Frechette & Moreno, 2010, p. 63) and is also consistent with CLT (Sweller, 1988), as attention-focusing cues are an established way to lessen extraneous cognitive load (Baddeley, 2006; Clark et al., 2005). In their 2015 study, Yung and Paas investigated the effects of a gesturing PA on learning the circulatory system, utilizing PA and no-PA conditions. The study revealed no significant differences in cognitive load between the two groups. However, the authors also calculated a score of learning efficiency (Paas and van Merrienboer, 1993) that quantifies the relationship between learning performance and cognitive load. It was found that the PA condition resulted in enhanced learning efficiency compared to the no-PA condition, indicating a more favorable relationship between learning and cognitive load.

The proposed benefits of a gesturing PA on cognitive load raises questions about the source of this effectiveness. Specifically, it prompts investigation into whether these benefits are attributable to the PA or simply to the act of gesturing (i.e., signaling) itself. This distinction is crucial for understanding the comparative effectiveness of non-PA signaling mechanisms, such as arrows. Choi and Clark (2006) found support for equivalence between these two methods, with no significant differences in learning or mental effort found between an arrows condition and a gesturing PA condition. Alternatively, Moreno et al.'s 2010 study found results favoring a gesturing PA over arrows. In this study, a control condition (no visual signaling), gesturing PA condition, and arrows condition were compared. Participants in the PA condition reported significantly lower difficulty ratings compared to the control condition, while the arrows condition did not differ significantly from the control condition. However, no significant differences in perceived difficulty were found between the PA and arrows conditions.

Many researchers have adopted an intermediary viewpoint regarding the impact of PAs on cognitive load, supported by studies showing no significant effect. For instance, in a study by Schroeder (2017), participants' perceived mental effort was measured at two time points. The results showed no significant difference in mental effort between the PA and no-PA conditions either after the lesson or after the post-test. Further evidence comes from a systematic review of 17 studies by Yusoff et al. (2023). Twelve of these studies reported no effect of PAs on cognitive load, while five reported a positive effect (i.e., decreased cognitive load). Importantly, no studies were identified in which PAs led to increased cognitive load. Other studies and meta-analyses have echoed these neutral findings, with no differences in cognitive load or similar measures (e.g., perceived difficulty) being found between PA and no-PA conditions (Davis, 2018; Frechette & Moreno, 2010; Moreno et al., 2001).

In summary, while the contradictions surrounding the effect of PAs on cognitive load warrant continued investigation, recent research (Schroeder, 2017; Yusoff et al., 2023) supports a neutral stance, that PAs do not affect cognitive load, rather than the opposing stance that PAs increase cognitive load.

1.1.3 Motivation

In the present study, motivation was evaluated using Keller's (1987) ARCS Model of motivation, comprising four dimensions: attention, relevance, confidence, and satisfaction. Attention involves obtaining and sustaining a learner's focus. Keller (1987) notes that while obtaining a learner's attention is easy, the challenge comes with sustaining it. This is particularly true for PA research, where researchers must ensure that the PA is not just capturing attention due to novelty but is also engaging enough to maintain the learner's attention, allowing for continued benefits of the PA over time.

Relevance refers to how applicable a learner feels that the subject of a lesson is to them. Without this, a learner might not feel the need to learn and/or might feel alienated in the learning environment. Keller (1987) notes, “Relevance can come from the way something is taught; it does not have to come from the content itself” (p. 7). This clarification is important for education-based research as researchers cannot always ensure that each participant will want to learn about the selected content domain. However, with properly designed materials, this dimension of the ARCS Model can still be satisfied.

Confidence encompasses a learner's belief in their ability to succeed in the learning environment. Both too little and too much of this attribute could be harmful to a learner’s motivation. If the learner has no confidence in their ability to succeed, they will be unmotivated to learn. On the other hand, if they are overly confident, they might ignore parts of the lesson due to their erroneous belief that they already know the content.

Satisfaction pertains to reinforcing positive behavior using either intrinsic or extrinsic rewards. In the context of the motivation survey used in the present study (Reduced Instructional Materials Motivation Survey; RIMMS; Loorbach et al., 2015), discussed later, satisfaction refers to the intrinsic rewards of enjoyment and success.

The ARCS Model has been used to measure motivation in previous PA studies. Dincer and Doganay’s 2017 study utilized four conditions: two PA conditions and two no-PA conditions. They measured motivation with the Instructional Materials Motivation Survey (IMMS; Keller, 2010), which evaluates motivation using the four dimensions of the ARCS Model. They found that both PA conditions scored significantly higher than one or both no-PA conditions on all four subscales and on the composite IMMS score. This supports the motivational benefits of PAs and the use of ARCS Model-based surveys in PA research.

1.2 Gaps in Literature

Despite the decades of research on PAs, questions about the conditions in which they can be optimally used remain unanswered. The present study investigated two such questions. These were: (1) whether PA effectiveness differs between two dissimilar domains (e.g., math and art), and (2) how a user's preference for learning from a PA impacts its effectiveness.

1.2.1 Content Domain

In their 2013 meta-analysis, Schroeder et al. reviewed 43 studies involving over 3,000 participants and observed distinct effects of PAs in different subject areas. Specifically, they found a significant effect size for learning in PA studies that utilized mathematics lessons ($g = 0.28$), while PA studies that utilized humanities lessons yielded a small, insignificant effect size ($g = 0.06$). The authors suggested that this effectiveness discrepancy between math and humanities could be due to the abstract nature of math. They proposed that perhaps the PA's ability to explain and demonstrate complex mathematical topics engaged or motivated learners, leading to enhanced learning.

Despite Schroeder et al.'s (2013) findings, no systematic comparison of PA effects in math versus humanities lessons exists in the current literature. Instead, most PA studies have concentrated on a single domain, often math or science. This has left a significant gap in understanding the content domains where PAs are most effective, ultimately leading to unanswered questions regarding the circumstances in which a PA's implementation is most beneficial.

1.2.2 Math versus Humanities

Mathematics and the humanities serve as suitable domains for comparison regarding a PA's effectiveness due to their inherent differences. The primary difference frequently found between math and humanities lessons is the level of element interactivity (Sweller, 1988; Sweller et al., 1998). As described in the previous section *1.1.2 Cognitive Load*, element interactivity refers to the degree of interdependence between the elements of information in a lesson. In math, element interactivity tends to be high because the learner's understanding of each element is often contingent upon their understanding of the interconnected elements. For example, understanding the Bernoulli trials equation, $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$, requires the learner to first understand what each variable (x, n, and p) represents, the mathematical operations involved (combinations and exponents), and the meaning of the equation's result. In contrast, a humanities lesson typically has a lower level of element interactivity. For example, in a lesson on the principles of design, the learner would be able to understand each of the principles independently; mastery of one principle would not necessarily impact comprehension of the others. Given that increased element interactivity typically results in heightened cognitive load for the learner (Sweller, 2011), it logically follows that math lessons, characterized by their higher element interactivity, would demand more cognitive effort than humanities lessons.

The difference in element interactivity between math and humanities allows inferences to be made as to why PAs are potentially more effective in math lessons than in humanities lessons (Schroeder et al., 2013). For instance, math lessons tend to start with more foundational knowledge and advance to more complicated applications of that knowledge, with each successive step building off the last. In contrast, humanities lessons are often composed of many segments of equally difficult content; while some of these sections could be related, they do not need to build off of each other. It is possible that PAs are more effective at guiding learners through the sequence of a math lesson than they are at instructing on humanities lessons of a more modular nature. Another potential explanation is that PAs might offer more benefits in subjects that are cognitively demanding, such as math, especially if they are utilizing deictic gestures to signal relevant portions of on-screen content. In a high-load context, the gestures offered by PAs could potentially lower the load more effectively, thereby favoring their use in math over humanities lessons.

While these explanations are merely hypotheses, they offer an interesting starting point for investigating the differential effectiveness of PAs between two highly discriminable content domains, such as math and the humanities, and why such differences might occur.

1.2.3 Preference

In learning environments involving PAs, learners are typically not given the opportunity to choose between learning with or without a PA; it is built into the learning program and therefore cannot simply be removed at the learner's discretion. Whether a PA is involved is therefore solely up to the designer of the learning environment, meaning each learner's preference for learning with a PA is not considered. Although no systematic review has documented this phenomenon, it is supported by reported examples of PA-enhanced systems. For instance, AutoTutor, a PA-based tutoring system, relied on an on-screen agent as the primary medium for content delivery and offered no option to remove it (Graesser et al., 2005). Similarly, Max was a “permanently applied…information kiosk agent in a public computer museum” (Krämer et al., 2007, p. 239), implying constant presence. In another case, the James the Butler system allowed users to mute the agent’s voice, but its visual presence remained on-screen, with speech still appearing in text bubbles (Hone et al., 2005). Furthermore, the author has yet to encounter an educational system that provides a clear option to disable the PA entirely, and recent meta-analyses and reviews do not consider the issue (Apoki et al., 2022; Castro-Alonso et al., 2021; Davis, 2018; Davis et al., 2022; Siegle et al., 2023). Despite this apparent disregard for learner preference, no research has examined whether preference for learning with or without a PA affects learning outcomes in PA-enhanced environments.

1.2.3.1 Preference in No-PA Environments

Limited research in standard online learning environments (i.e., without PAs) has explored the impact of honoring user preferences on outcomes. Freitag and Sullivan (1995) asked participants to choose between a basic or comprehensive training program and then randomly assigned them to either their preferred or non-preferred program. The study found that participants whose program matched their preference scored higher on the post-test than those in the program opposite their preference. The matched group also spent less time on the program than the unmatched group and had more positive opinions about the program's likability, amount of information provided, desire to learn more about the program topic, and confidence in using what they learned.

In Cinkara and Bagceci’s (2013) study, 1,783 Turkish students taking an online English course were surveyed about their attitudes toward online language learning; each student was categorized as feeling very positive, positive, negative, or very negative towards online language learning. Their grades on two post-course language tests were used to measure learning. There was a significant positive correlation between online language learning opinions and grades, suggesting that more positive opinions of online language learning were associated with better grades.

Both of these studies support the notion of honoring learner preferences and attitudes in online learning, although neither study was conducted on a PA-enhanced environment.

1.2.3.2 Learner Choice & Personalization in PA-Enhanced Environments

While no research has specifically examined the effects of users' preferences for PAs on learning outcomes, studies on learner choice and personalization in PA-enhanced environments suggest its potential impact.

Ozogul et al. (2013) found that students who selected their PA's design (from four possible designs) performed better on far transfer questions than those randomly assigned a PA, indicating that learner choice can positively affect learning. This concept has been supported by Dincer and Doganay (2017), who stated that “Learners should be provided with [computer-assisted instruction] programs that can be personalized depending on learners’ needs and preferences” (p. 74). Gulz (2004) also noted that the benefits of PAs are not universal and are susceptible to individual differences.

These findings support the use of learner choice in PA-enhanced learning environments, implying that honoring learner preferences could significantly influence dependent measures. Ultimately though, no prior research has examined learning performance as a function of whether a learner's preference for learning with a PA is honored.

1.2.3.3 Preference-Performance Dissociation

While prior education research suggests that honoring user preferences can benefit learning, it is also important to note that user experience (UX) literature has suggested that preference is not always indicative of performance. Andre and Wickens (1995) highlighted several studies in which preference for design features failed to align with performance. For example, Blanchard et al. (1993) found no performance differences between two letter-to-number mapping methods, even though one was significantly more preferred. Similarly, participants in Jeffrey and Beck’s (1972) study preferred the use of color in displays, despite it having no effect on performance. In other cases, user preference was inversely related to performance. Tognazzini (1992) found that participants incorrectly rated a slower text replacement device as being faster, and Keyson and Parsons (1990) found that among five computer system prototypes, participants gave the highest preference ratings to the one with the worst performance and the lowest preference ratings to the one with the best performance. These findings highlight the importance of empirically testing whether honoring preferences improves learning outcomes, as was done in the present study.

1.2.3.4 Dislike for PAs

The present study sought to understand the impact of user preference for PAs, a topic whose importance is highlighted by documented instances where participants express a dislike for learning with PAs. This illustrates that not all people react equally to PAs, highlighting the need for understanding personal choice when it comes to the presence versus absence of a PA in the learning environment.

A study by Hook et al. (2000) provides a case where participants expressed negative reactions to PAs. In this study, participants explored web pages either in the presence or absence of two virtual agents; note that it is more accurate to call the agents in this study “virtual” agents rather than “pedagogical” agents because participants were engaging in exploration rather than learning. The results showed that out of the 18 participants who interacted with the agents, 4 felt "often disturbed" and 6 "sometimes disturbed" by them (p. 13). Additionally, 6 participants reported that the agents often distracted from the web content. There was also a positive correlation between users' perceptions of the agents and their willingness to reuse the system, indicating that negative opinions of the agents tended to reduce inclination to use the system again.

Furthermore, Hook et al.’s (2000) study found that negative reactions to the agents were influenced by users' prior web and computer experience. More experienced users tended to be more disturbed by the agents. The authors hypothesized that irritation with the agents might have stemmed from frustration with them interfering in the web exploration process. This was likely more pronounced in participants with extensive experience due to the agents disrupting their already established browsing preferences.

The existence of negative attitudes towards PAs emphasizes the need to investigate if such attitudes could negatively impact learning. Hook et al.'s (2000) study also supports providing learner choice regarding the presence of PAs, given that individual differences were found to impact participants' attitudes towards the agents.

1.3 Present Study: Overview

The present study comprises two experiments. Experiment 1 (1) examined differences between a condition with a PA and a condition without a PA, (2) compared a PA's effectiveness in a math lesson versus art lesson, and (3) evaluated the effectiveness of a PA and two lessons designed by the author prior to their use in Experiment 2. Experiment 2 investigated the effect of congruence and incongruence between a learner's preference for a PA and their actual PA assignment. The outcomes examined in both experiments were learning (including retention in Experiment 1), motivation, subjective workload, PA persona, and opinions on the lesson and learning environment.

PA persona has not yet been defined in the present paper. In the context of PAs, "persona" refers to learner perceptions of a PA, often influenced by the PA's personality, characteristics, and humanlikeness (Davis, 2018; Ryu & Baylor, 2005; Schroeder et al., 2017). A common measure of PA persona, the Agent Persona Instrument (API; Ryu & Baylor, 2005), evaluates persona based on the learner's perceptions of the PA's (1) ability to facilitate learning, (2) credibility, (3) humanlikeness, and (4) engagement.

CHAPTER 2. PEDAGOGICAL AGENT DESIGN

In addition to the benefits of PAs and the unanswered questions surrounding PAs in the current literature, it is important to discuss the design decisions that were made for the PA in the current study. There are countless characteristics that can be manipulated for a PA, including its physical characteristics, auditory characteristics, and animations. Each one of these features has the potential to impact the efficacy of the PA on one or multiple dependent measures. Thus, all choices must be carefully considered during the design process. The present section will describe the design decisions that were made for the PA in this study and the rationale behind them. See Figure 3 for an overview of the features that will be discussed.

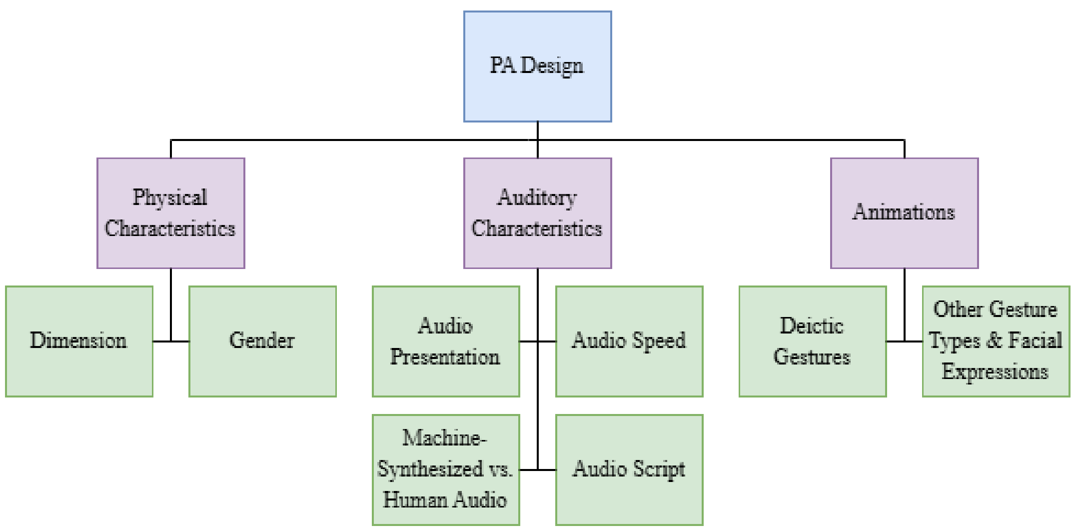


Figure 3 – Design choices that will be discussed in the current section.

2.1 Physical Characteristics

There are virtually limitless design choices that can be made for a PA's physical features. For the purposes of the current paper, two features will be discussed: the PA's dimensionality and the PA's gender.

2.1.1 Dimensionality

One feature to consider when designing a PA is whether it is 2-dimensional or 3-dimensional. Although limited research exists comparing these attributes, Castro-Alonso et al. (2021) did address this design decision in their meta-analysis, with results favoring a 2-dimensional PA. Specifically, 2-dimensional PAs showed a significant small to medium effect size ($g+ = 0.38$) compared to no-PA conditions, while 3-dimensional PAs did not demonstrate a significant effect ($g+ = 0.11$). The authors concluded by recommending a simpler 2-dimensional PA over a more visually complex 3-dimensional PA, attributing the 2-dimensional PA's advantages to it providing less nonessential information, thereby reducing its likelihood of overloading working memory. Thus, in the present study, a 2-dimensional PA was used.

2.1.2 Gender

The impact of a PA's gender on learning outcomes and user perception has been mixed, with no clear consensus in the literature. Some studies have found no significant effect of PA gender on various learning outcomes (Castro-Alonso et al., 2021; Schroeder & Adesope, 2015), while others report conflicting results, with some favoring male PAs (Baylor & Kim, 2004) and others favoring female PAs (Novick et al., 2019; Plant et al., 2009). Given this ambiguity in the literature, the choice of PA gender in the present study was guided by demographics. Reflecting data that indicates a higher prevalence of female teachers in the U.S. (77% female, 23% male; National Center for Education Statistics, 2023), it was decided that a female PA would be used. This decision, while not directly supported by experimental evidence, was based on learners' likely familiarity with female instructors.

2.2 Auditory Characteristics

The auditory characteristics that were considered for this study's PA included whether it would present audio, the similarity between the audio and on-screen text, the audio's speed, whether it was machine-synthesized or human, and the characteristics of the audio script.

2.2.1 Audio Presentation

It is important to consider whether to include audio in a PA-enhanced learning environment or to rely on a text-based presentation. Atkinson (2002) found that an audio-based lesson led to better performance than a text-based lesson on near transfer learning and was rated as less difficult by participants, aligning with other studies favoring audio (Craig et al., 2002; Moreno et al., 2000; Moreno et al., 2001). This is consistent with the modality effect, which states that audiovisual formats are more effective than purely visual formats (Sweller, 2011). Sweller (2011) notes that this dual-modality presentation should "increase effective working memory and decrease cognitive load" (p. 67). Thus, audio was used in the present study's lessons.

2.2.2 Audio-Text Redundancy

Once an audiovisual format is selected for a PA-assisted lesson, the next consideration is how closely the audio aligns with the on-screen text. Some researchers have suggested that simultaneously presenting identical audio narration and on-screen text in a PA condition can prompt the redundancy effect (Mayer et al., 2001; Sweller, 2011), impairing learning (Craig et al., 2002; Siegle et al., 2023). The redundancy effect is a multimedia learning construct that suggests that the presentation of the same information through auditory and visual input channels reduces learning.

In 2008, Mayer and Johnson reexamined the redundancy effect. They compared a group that was presented no redundant information (audio without on-screen text) to a group presented reduced redundant information (audio with only essential on-screen text). The group exposed to partial redundancy outperformed the nonredundant group on a retention test. A follow-up study replicated these findings. Thus, Mayer and Johnson (2008) suggested reevaluating the redundancy effect. They argued that, rather than eliminating all redundant on-screen text accompanying audio, using concise, key text can direct learners' attention to important words without adding extraneous processing. As a result of the nuanced understanding provided by Mayer and Johnson (2008), the current study incorporated audio along with succinct on-screen text.

2.2.3 Audio Speed

Determining the optimal "speaking" speed for a PA involves balancing comprehensibility and listener preference. A normal speaking rate is typically defined as being between 150 and 180 words per minute (WPM; Pastore & Ritzhaupt, 2015; Ritzhaupt & Barron, 2008; Ritzhaupt et al., 2015), with research showing that rates can be increased up to 225–275 WPM before learning outcomes and comprehension are affected (Barabasz, 1968; Reid, 1968; Ritzhaupt & Barron, 2008). Despite the understandability of such high speeds, they are typically not preferred by learners.

Ritzhaupt et al. (2015) compared video speeds of 167, 219, and 261 WPM and found that, although there was no difference in test performance between the three speeds, participant satisfaction was highest at the "normal" 167 WPM speed. This finding is mirrored in Pastore's (2015) study where participants preferred the audio speeds 164 WPM and 180 WPM over rates ranging from 197 to 246 WPM for the purpose of instruction/learning. Experiment 1 of the present study used an audio speed of 180 WPM, which falls within the listener-preferred range of 164–180 WPM suggested by Ritzhaupt et al. (2015) and Pastore (2015), aiming to balance comprehension and satisfaction.

2.2.4 Machine-Synthesized vs. Human Audio

The debate between using a human voice versus machine-synthesized voice for a PA has evolved with technology. Previously, researchers argued for the voice effect, which stated that a human voice is more beneficial for learning than a machine-synthesized voice (Mayer, 2014). This effect was consistent with Social Agency Theory (Mayer et al., 2003b), as a human voice would likely serve as a more salient social cue than a machine voice (Craig & Schroeder, 2019). However, modern advancements in text-to-speech engines challenge this effect, with studies by Craig and Schroeder (2017, 2019) demonstrating that modern voice engines can match human voices in credibility and facilitating learning. The authors concluded that modern voice-to-speech technologies negate the voice effect. Therefore, the current study used audio generated by a modern engine, NaturalReader (https://www.naturalreaders.com/online/), using the AI voice "Ava" at a rate of 180 WPM.

2.2.5 Audio Script

The "audio script" refers to the auditory portion of the lesson, meaning the information provided by the PA's "voice." In this study, the PA's audio script conveyed educational content while also incorporating encouragement and a conversational tone, aiming to enhance perceptions of the PA, user engagement, and learning.

The idea of incorporating encouragement comes from Baylor and Kim's 2005 study, in which they analyzed three PA roles: expert, motivator, and mentor. The expert PA provided only information, the motivator PA provided only encouragement, and the mentor PA provided both information and encouragement. The study found that the PAs providing encouragement (motivator and mentor) were perceived as more humanlike and engaging than the purely informational PA (expert). Furthermore, the mentor PA (provided both information and encouragement) resulted in significantly higher transfer scores than the other roles. Thus, in the present study, the PA held the role of a mentor by providing both information and encouragement to the learner. Encouragement was achieved via the incorporation of phrases such as "Great job!" and "You did awesome!"

As recommended by Mayer's (2023) personalization principle, the script also used a conversational tone (as opposed to a formal tone). This strategy was empirically supported by Mayer et al. (2004), who found that a conversational style, characterized by personalized language (e.g., "during inhaling, *your* diaphragm moves down" instead of "during inhaling, *the* diaphragm moves down"), significantly enhanced learners' transfer scores. A conversational tone was incorporated into the audio script via the use of personalized language, guiding/conversational questions, and casual phrases/language when appropriate.

2.3 Animations

When designing a PA, it is important to consider the types of animation the PA will use, such as gestures and facial expressions. According to the embodiment principle (Mayer & DaPra, 2012), humanlike animations enhance a PA's social presence and deepen user learning. This aligns with Social Agency Theory (Mayer et al., 2003b), which states that social cues from a PA, such as those presented via animations, encourage learners to process content more deeply due to the elicitation of conversational schemas (Schroeder et al., 2022). Findings consistent with the embodiment principle were found in a 2013 meta-analysis by Schroeder et al., where the results showed a significant effect size for animated PAs on learning (g = 0.15) but no effect for static PAs (g = 0.00).

2.3.1 Deictic Gestures

One type of gesture that can be used by PAs is the deictic gesture; this is a specific pointing gesture (Li et al., 2019) typically created by motioning with an extended index finger. The use of deictic gestures by PAs has been shown to impact learning and PA persona.

Studies by Baylor and Kim (2009) and Li et al. (2019) demonstrated that the use of deictic gestures by a PA significantly enhances learning (including immediate and delayed transfer and retention) compared to a non-gesturing PA. Such findings were confirmed in a 2018 meta-analysis by Davis which found that gesturing PAs have a significant effect on near transfer ($g = 0.39$) and retention ($g = 0.28$). Another study by Moreno et al. (2010) compared a deictically gesturing PA, pointing animated arrows, and a non-gesturing PA. The results showed that the gesturing PA led to increased post-test performance compared to the other two groups. Moreno et al.'s (2010) findings demonstrate the benefits of deictic gestures by a PA, even when compared to other visual signaling methods (i.e., arrows).

The APA (animated PA) signaling hypothesis provides a rationale for why deictic gestures improve learning outcomes (Moreno et al., 2010); it posits that the visual presence of a PA enhances learning by facilitating the cognitive processes essential for learning, such as by guiding learners' focus to important on-screen content through deictic gestures. This signaling aids learners with discriminating between relevant and irrelevant visual material, which has been suggested to be particularly important when instructing novices who likely lack the experience needed to appropriately identify the relevant materials themselves (Moreno et al., 2010).

Deictically gesturing PAs have also been shown to impact PA persona. To reiterate the earlier provided definition, PA persona refers to how a PA is perceived by learners and is affected by features such as the PA's personality, characteristics, and humanlikeness (Davis, 2018; Ryu & Baylor, 2005; Schroeder et al., 2017). This concept aligns with Mayer and DaPra's (2012) embodiment principle and Social Agency Theory (Mayer et al., 2003b): the PA's persona is related to how humanlike it is perceived to be; as the PA is perceived to be more humanlike, the learner tends to view the PA as more of a social partner, and deeper learning is facilitated. The previously discussed meta-analysis by Davis (2018) found that the use of deictic gestures by PAs had a significant positive effect on PA persona ($g = 0.44$).

Based on the findings that (1) users learn better from a deictically gesturing PA than a non-gesturing PA, and (2) deictic gestures enhance users' perceptions of the PA (measured as PA persona), the PA in the present study performed deictic gestures by pointing to relevant on-screen text with an outstretched finger. These gestures were timed to coincide with the words spoken in the audio, ensuring that the visual cue of the gesture was synchronized with the auditory cue of the relevant instruction.

2.3.2 Other Gesture Types & Facial Expressions

Research indicates that the use of non-deictic gestures and facial expressions by PAs enhances learning and PA persona. The benefits of such animations likely stem from their role in improving the PA's humanlikeness, its persona, and the social relationship between the PA and the learner. This reasoning is consistent with the embodiment principle (Mayer & DaPra, 2012) and Social Agency Theory (Mayer et al., 2003b).

A 2022 study by Schneider et al. demonstrated the learning benefits of non-deictic, humanlike gestures and facial expressions. Significant main effects were found for gestures and facial expressions on retention. Similar results were found on a transfer test. Baylor and Kim (2009) also examined facial expressions, with results showing that a PA with facial expressions enhanced learning relative to a PA without facial expressions.

Additionally, Baylor and Kim (2009) found advantages for facial expressions in terms of PA persona. The authors used the Agent Persona Instrument (API; Ryu & Baylor,

2005) to measure PA persona. This instrument includes four subscales: ability to facilitate learning, credibility, humanlikeness, and level of engagement. They found that the PA with facial expressions led to significantly higher scores on all four API subscales compared to the PA without facial expressions.

Given the documented benefits of non-deictic gestures and facial expressions, the current study incorporated both in the PA design. Non-deictic gestures included general gestures, such as moving the hands in time with speaking. Facial expressions included neutral, happy, excited, and inquisitive.

CHAPTER 3. EXPERIMENT 1

As discussed in section *1.3 Present Study: Overview*, Experiment 2 investigated the effects of PA preference. Prior to doing this, it was important to assess the effectiveness of the PA that was designed by the researcher. Therefore, Experiment 1 assessed whether, relative to a No-PA condition, the PA increased learning and motivation, and decreased subjective workload. Participant perceptions of the PA were also assessed through the measurement of PA persona and use of an opinion-based survey. Additionally, Experiment 1 compared the effectiveness of the PA in a math lesson and an art lesson.

3.1 Design

A 2×2 mixed design was used with the dependent variables of learning, motivation, subjective workload, PA persona, opinions on the lesson, and opinions on the assigned learning aid. The first independent variable was PA Assignment, which was a between-subjects variable with two levels: assigned a PA (With-PA) versus not assigned a PA (No-PA). The second independent variable was Lesson, which was a within-subjects variable with two levels: Math versus Art.

3.2 Hypotheses

1. Hypothesis 1.1: Participants in the With-PA condition will have better outcomes than those in the No-PA condition, such that:
   A. They will exhibit higher immediate learning (recall, recognition, and transfer).
   B. They will report higher motivation.

C. They will report lower workload.

D. They will report higher PA persona.

E. They will exhibit higher retention (recall, recognition, and transfer).

2. Hypothesis 1.2: Participants in the With-PA condition will have better outcomes in the Math condition than in the Art condition, such that:

   A. They will exhibit higher immediate learning (recall, recognition, and transfer).

   B. They will report higher motivation.

   C. They will report lower workload.

   D. They will report higher PA persona.

   E. They will exhibit higher retention (recall, recognition, and transfer).

3.3 Participant Recruitment

A priori analysis for a repeated measures ANOVA with within-between interaction was conducted in G*Power (Faul et al., 2007) with a medium effect size of $f = .25$. This test indicated that a sample size of 34 total participants would be required to achieve a power of .8 using the described design.

Participants were recruited from a large technical university in the southeastern United States. Participants received partial course credit through the university's SONA System, a research participation management platform commonly used in academic settings. Students who participated in an earlier PA study by the present author were ineligible to participate. Inclusion criteria comprised normal or corrected-to-normal hearing, vision, and color vision.

3.4 Materials

3.4.1 Pedagogical Agent

The PA used in the With-PA condition of Experiment 1 (Figure 4) was an audio-narrating 2-dimensional female PA. The PA displayed deictic gestures (pointing to relevant on-screen content with an outstretched index finger), non-deictic/general gestures (natural movements such as moving hands in time with speaking), and facial expressions (neutral, happy, excited, inquisitive; Figure 5). The PA also displayed lip syncing ("talking") and blinking. The justifications for these design decisions were discussed in Chapter 2: Pedagogical Agent Design.

In the No-PA condition, arrows were used in place of the PA (Figure 6). These arrows replicated the PA's pointing movements (i.e., deictic gestures) to ensure equivalent visual attention guidance across both the With-PA and No-PA conditions. The audio remained identical to that used in the With-PA condition.

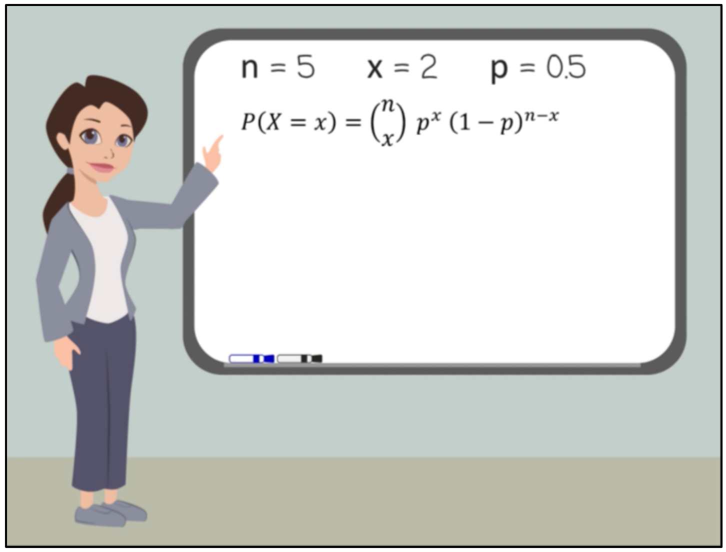


Figure 4 – PA and learning environment used in this study.

Note. The screen capture is from the math lesson. The PA is performing a deictic gesture.

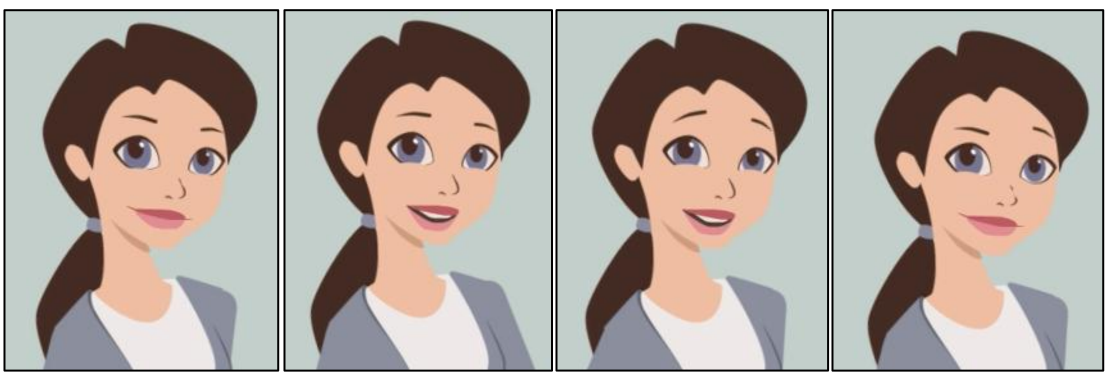


Figure 5 – Facial expressions of the PA in this study, from left to right: neutral, happy, excited, inquisitive.

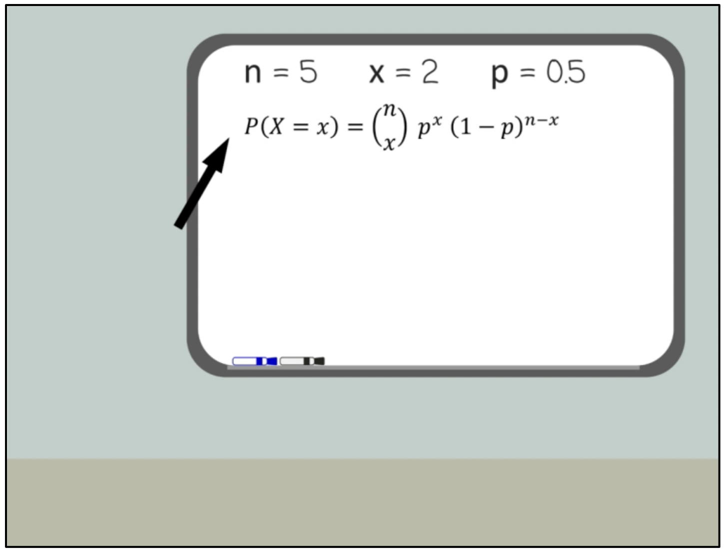


Figure 6 – Arrow in place of PA in No-PA condition

Note. The screen capture is from the math lesson. The arrow is mimicking the deictic gesture performed by the PA in Figure 4.

3.4.2 Lessons

Over two sessions, participants watched two video lessons, one per session: a math lesson (approximately 18 minutes in length) and an art lesson (approximately 19 minutes in length). Both videos featured audio and on-screen text, either with the PA or without the PA (arrows in place of PA). The topics of the video lessons were Bernoulli trials (math; Appendix A) and the principles of design (art; Appendix B). The order of the math and art lessons was counterbalanced across participants, ensuring that half received the math lesson in Session 1 and half received the art lesson in Session 1.

Consistent with the discussion from previous section *1.2.2 Math versus Humanities*, the primary distinction between the math and art lessons in this study was their level of element interactivity (Chen et al., 2023; Sweller et al., 1998). The math lesson presented a higher degree of element interactivity than the art lesson, as demonstrated in the next subsection (*3.4.2.1 Measuring Element Interactivity*), which was expected to impose a greater cognitive load on participants (Sweller et al., 1998).

3.4.2.1 Measuring Element Interactivity

Sample segments from each lesson were examined to demonstrate the differences in element interactivity between the present study's math and art lessons. To form a representative sample, each lesson's shortest segment (examined in the present section) and longest segment (examined in Appendix C) were selected.

As described in section *1.1.2 Cognitive Load*, element interactivity refers to the complexity of the instructional content, determined by the number of concepts or procedures to be learned (elements) that must be simultaneously processed in working memory (interactivity) for comprehension to occur (Chen et al., 2023; Sweller et al., 1998; Sweller et al., 2019). Chen et al. (2023) note that a measure of element interactivity must be informed by both the structure of the information being processed and the prior knowledge of the learner. The nature of this measurement means that element interactivity varies by the individual, as each learner's level of prior knowledge will differ. In the present Experiment 1, participants' self-reported prior knowledge ratings (1–5 Likert scale; see section *3.5 Measures*) did not differ substantially ($p = .157$) between the math ($M = 1.48$, $SD = 0.51$) and art ($M = 1.70$, $SD = 0.68$) lessons. However, their performance on the objective pre-test questions did differ ($p < .001$), with lower scores on the math pre-test (0–1 scale; $M = 0.12$, $SD = 0.26$) compared to the art pre-test ($M = 0.32$, $SD = 0.16$). Thus, even before considering the informational elements in each lesson, the pre-test results suggest a higher level of element interactivity for the math material due to participants' lower prior knowledge.

Building on Chen et al.'s (2023) approach to measuring element interactivity, this section estimates the number of interacting elements in each lesson by analyzing lesson segments, identifying clusters of elements that must be processed simultaneously in working memory for comprehension, and counting the interacting elements within and between clusters. Note that, as stated by Chen et al., "element interactivity as a measure of complexity is always an estimate rather than an exact measure" (p. 15). To demonstrate the level of element interactivity in the math lesson, the segment in Figure 7 was examined, as shown in the diagram in Figure 8.

For our first example, let's go back to considering coin flips. Imagine that you flip a coin 5 times. On each flip, you have a 50% chance of getting a heads. The question is: What is the probability of the coin landing on heads exactly 2 times? Let's walk through how to solve this problem step-by-step. Ready?

First, let's figure out the components of the equation by determining the values of our variables.

In this example, you're flipping the coin 5 times. This means that there are 5 total trials, so n equals 5. We're defining a "success" as getting heads, and we're trying to figure out the probability of getting heads 2 times. So, x equals 2. The problem told us that the probability of getting heads is 50%. We need to convert the percentage to a decimal to use it in our equation. 50% becomes 0.5, making p equal to 0.5.

Great job! Now, let's plug these values into our equation for a Bernoulli trial. Remember, we have n=5, x=2, and p=0.5.

$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$	Here's our equation with no numbers plugged in.
$P(X = 2) = \binom{5}{2} 0.5^2 (1-0.5)^{5-2}$	Now, let's plug in our values. We have 5 for n, 2 for x, and 0.5 for p. Great! Now that we have our numbers substituted into the equation, we can work on calculating the answer.
$P(X = 2) = 10 \times 0.25 \times (0.5)^3$	Since you don't have a calculator, I went ahead and calculated the value of the combination 5 choose 2 for you. It equals 10. I also calculated 0.5 squared, and it equals 0.25. Now, let's simplify 1 minus 0.5 by solving: 1 minus 0.5 equals 0.5. And for the superscript, 5 minus 2 equals 3.
$P(X = 2) = 10 \times 0.25 \times 0.125$	Next, we can calculate 0.5 cubed, which is 0.125.
$P(X = 2) = 0.3125$	Our last step is to multiply these three numbers: 10, 0.25, and 0.125. This gives us 0.3125.

There you have it! The probability of flipping a coin 5 times and the coin landing on heads exactly 2 times is 0.3125, or 31.25%.

Figure 7 – Shortest segment of math lesson, examined for level of element interactivity (see Figure 8).

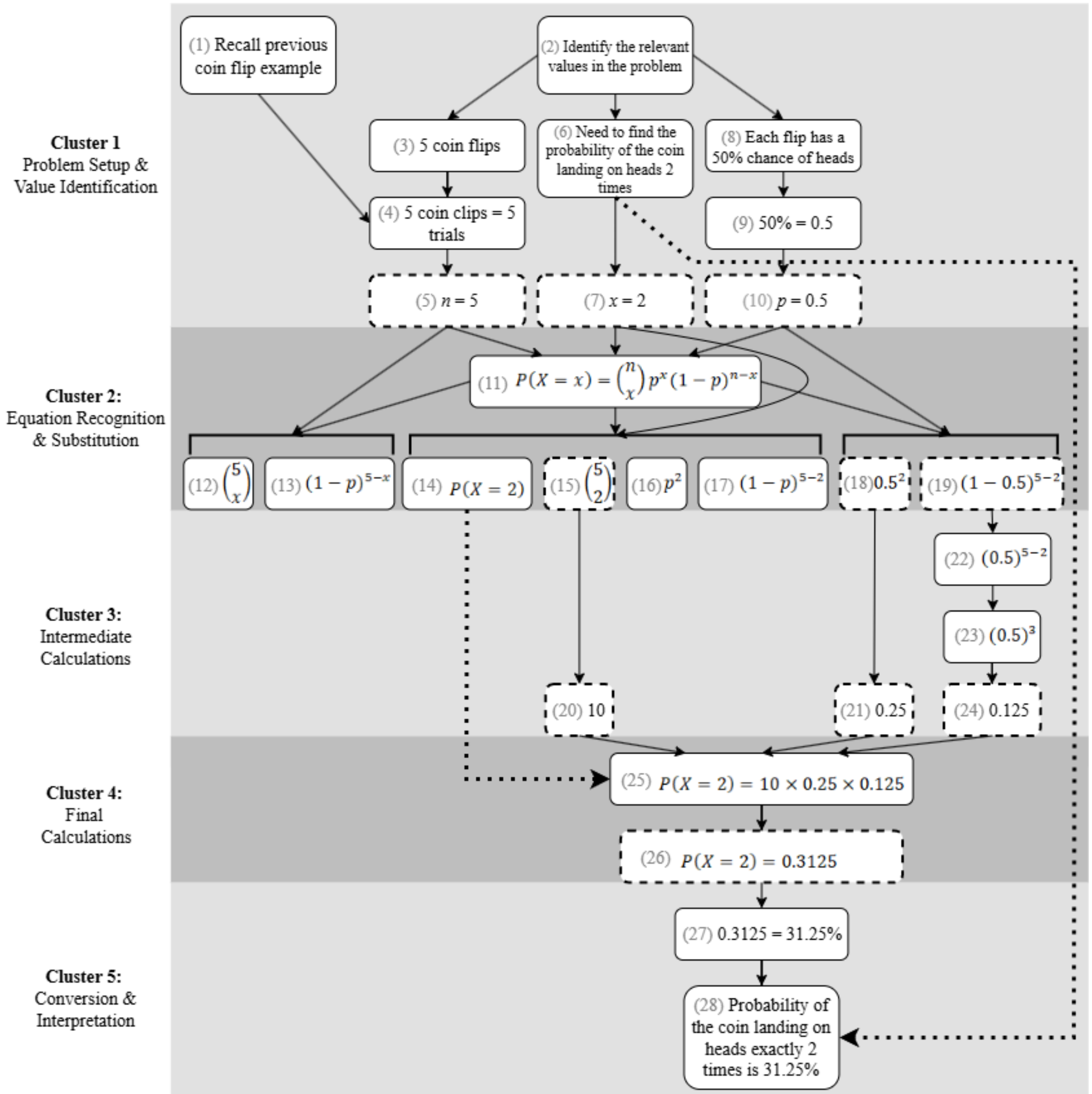
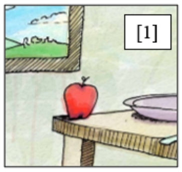


Figure 8 – Visual representation of the five element clusters in the math lesson segment shown in Figure 7.

Note. Alternating shaded regions indicate cluster boundaries. Elements within the same cluster must be processed simultaneously for comprehension of that cluster. Solid-box elements appear in only one cluster, while dashed-box elements are shared between adjacent clusters. Solid arrows show connections within a cluster and between adjacent clusters, while dotted arrows show connections across non-adjacent clusters. Brackets group elements to improve readability when one previous element connects to multiple subsequent elements.

Figure 8 shows a representative diagram of the element interactivity involved in the math lesson segment presented in Figure 7. The elements are divided into five clusters. Each cluster reflects a distinct stage of reasoning, during which simultaneous processing of those elements (i.e., element interactivity) is required for comprehension of that phase of problem solving. Cluster 1 (*Problem Setup & Value Identification*) contains elements 1–10; the learner must recall an earlier example and identify the relevant values in the scenario, such as the number of trials, success probability, and target outcome. Cluster 2 (*Equation Recognition & Substitution*) contains elements 5, 7, and 10–19; the learner must recognize the Bernoulli trial equation and substitute the appropriate values for *n, x*, and *p*. Cluster 3 (*Intermediate Calculations*) contains elements 15 and 18–24; the learner must observe the necessary computations on the substituted values, including solving the combination, squaring and cubing values, and simplifying expressions. Cluster 4 (*Final Calculations*) contains elements 14, 20–21, and 24–26; the learner must perform multiplication on the intermediate results to obtain the final answer. Cluster 5 (*Conversion & Interpretation*) contains elements 6 and 26–28; the learner must convert the final answer from a decimal to percentage and understand what the answer represents. Elements 5–7, 10, 15, 18–19, 20–21, 24, and 26 appear in multiple clusters to reflect their continued cognitive involvement across stages. The total number of interacting elements per cluster is 10, 12, 8, 6 and 4, respectively.

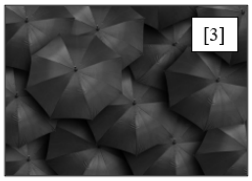
To allow for comparison of the levels of element interactivity in the math lesson and the art lesson, an art lesson segment of similar length (see Figure 9) was examined, as shown in the diagram in Figure 10.

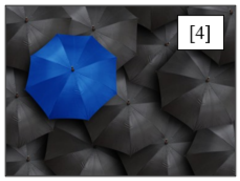


Let's start with the first principle: emphasis. Think of emphasis as the spotlight in an artwork. It's all about drawing your attention to the specific part of the piece that the artist really wants you to focus on. Artists can emphasize a part of their artwork in several ways. One common method is through placement. For example, in [1] this drawing, what was your attention drawn to first? It was probably the apple, right? This is because it's positioned in the center of the artwork, making it the focal point.

There are many other ways besides placement that artists can emphasize parts of their work. Let's talk about size. Take a look at [2] this painting. Notice how your eyes are immediately drawn to the tree? That's not a coincidence; it's a thoughtful decision by the artist to emphasize the tree by making it the largest object in the artwork.

Can you think of any other ways that emphasis can be created in art? If you thought of color, you're correct! Color is a common and powerful way that artists add emphasis to their work.





For example, in [3] this photo, do you notice how all the umbrellas blend together? There isn't a particular one that stands out, is there? But now, let's use color to create emphasis. Wow! [4] The blue umbrella really stands out now, doesn't it? By adding color to just one umbrella, we've created emphasis in this picture.

Color can also be used to emphasize part of the work without having to remove color from the rest of the piece. Instead, emphasis can be created by making one element of the work a different color than the rest. Let's take a look at [5] this drawing. Which tomato did you look at first? Even though all the tomatoes are equally vibrant, your attention was probably drawn to the green one because it's a different color than the rest.

Artists often use a combination of the methods we've covered to emphasize part of their work. Looking at this artwork, did you find your gaze drawn to the red eye almost immediately? This shows how effectively these techniques can be combined to create emphasis. In this image, the artist emphasized the red eye by 1) placing it in the center of the piece, 2) making it larger than any other objects, and 3) making it a brighter and different color than the rest of the piece.

Figure 9 – Shortest segment of art lesson, examined for level of element interactivity (see Figure 10).

Note. Bracketed numbers indicate the example image being referenced in the text.

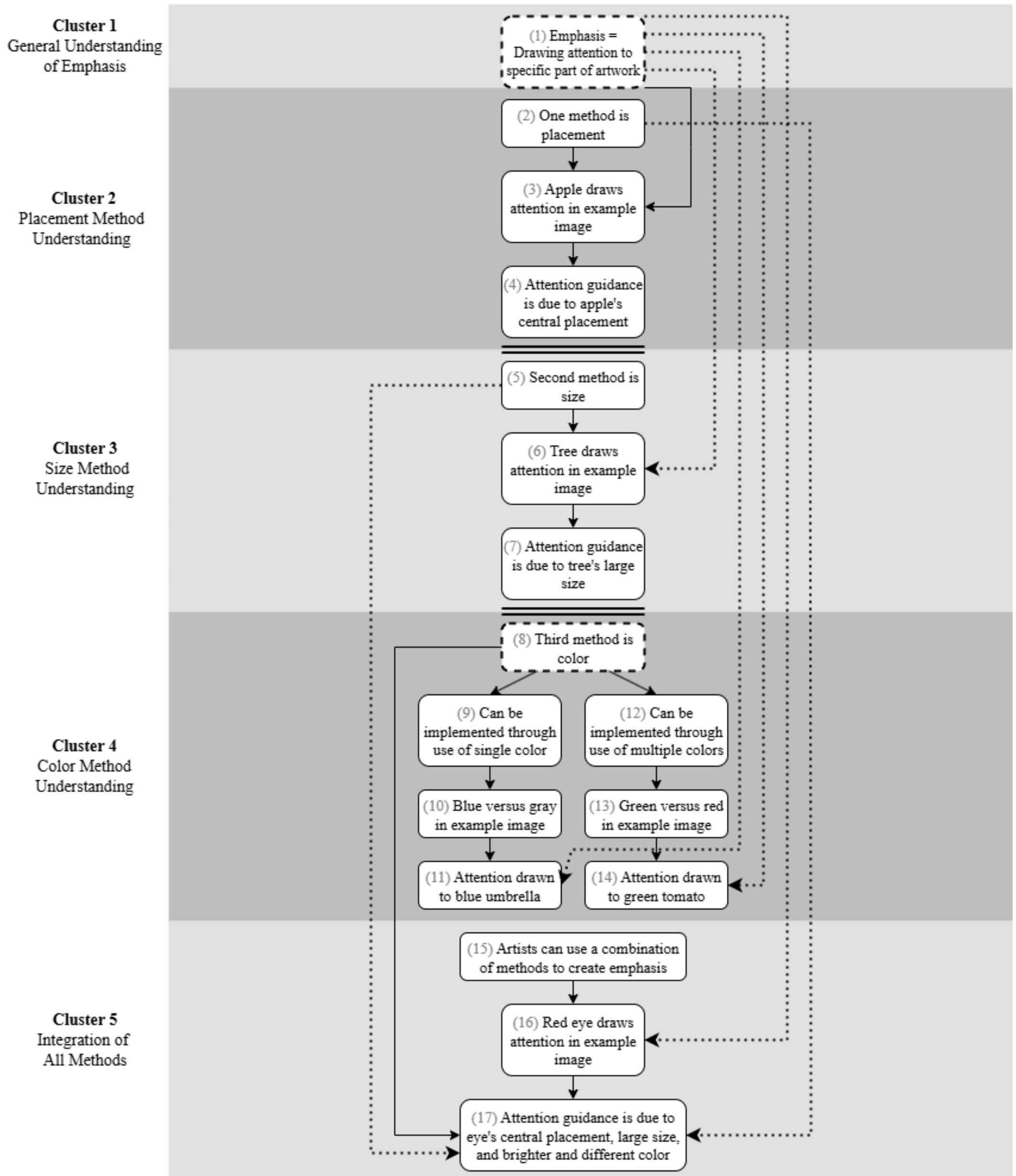


Figure 10 – Visual representation of the five element clusters in the art lesson segment shown in Figure 9.

Note. Alternating shaded regions indicate cluster boundaries. Elements within the same cluster must be processed simultaneously for comprehension of that cluster. Solid-box elements appear in only one cluster, while dashed-box elements are shared between adjacent clusters. Solid arrows show connections within a cluster and between adjacent clusters, while dotted arrows show connections across non-adjacent clusters. Brackets group elements to improve readability when one previous element connects to multiple subsequent elements. Double lines indicate that elements in a block are not connected to elements in the next block.

Figure 10 shows a representative diagram of the element interactivity involved in the art lesson segment presented in Figure 9. The elements are grouped into five clusters, each representing a distinct stage of reasoning. As in the previous diagram, understanding that stage of problem solving requires simultaneous processing of the elements within the cluster (i.e., element interactivity). Cluster 1 (*General Understanding of Emphasis*) contains element 1; the learner must understand the definition of emphasis. Cluster 2 (*Placement Method Understanding*) contains elements 1–4; the learner must understand the placement method used to create emphasis, recognize its application in the example image, and relate placement to the attention-guidance property of emphasis. Cluster 3 (*Size Method Understanding*) contains elements 1 and 5–7; the learner must understand the size method used to create emphasis, recognize its application in the example image, and relate size to the attention-guidance property of emphasis. Cluster 4 (*Color Method Understanding*) contains elements 1 and 8–14; the learner must understand the color method used to create emphasis, know that it can be implemented via two methods, recognize the application of the two methods in the example images, and relate color to the attention-guidance property of emphasis. Cluster 5 (*Integration of All Methods*) contains elements 1, 2, 5, 8 and 15–17; the learner must understand that the three previously discussed methods can be used together, recognize these methods in the example image, and understand that they draw attention. Elements 1–2, 3, 6, and 9 appear in multiple clusters to reflect their continued cognitive involvement across stages. The total number of interacting elements per cluster is 1, 4, 4, 9 and 7, respectively.

As shown in Table 2, the math lesson segment contains more interacting elements than the art lesson segment in three of the five clusters (Clusters 1–3), with large differences in each case (10 vs. 1, 12 vs. 4, and 8 vs. 4, respectively). The art segment has more interacting elements in the other two clusters (Clusters 4–5), with the difference being small (9 vs. 6 and 7 vs. 4, respectively). In addition, more elements are shared between clusters in the math segment than in the art segment (12 vs. 7). Together with the diagrams in Figures 9 and 11, these counts highlight the greater element interactivity in the math segment compared to the art segment.

Table 2 – Counts of interacting elements per cluster, elements shared between clusters, and total shared elements, organized by Lesson.

	Number of Interacting Elements					Number of Elements Shared between Clusters										
	Clusters					Cluster Pairs										
Lesson	1	2	3	4	5	1, 2	1, 3	1, 4	1, 5	2, 3	2, 4	2, 5	3, 4	3, 5	4, 5	Total
Math	10	12	8	6	4	3	0	0	1	3	1	0	3	0	1	12
Art	1	4	4	9	7	1	1	1	1	0	0	1	0	1	1	7

An additional examination of element interactivity in the two lessons is provided in Appendix C, along with diagrams (Figures C2 and C4) and tables of counts (Tables C1 and C2). Based on the examinations in the present section and the appendix, the math lesson contains a higher level of element interactivity than the art lesson. As previously noted, measures of element interactivity are only estimates, so small differences would not be expected to produce detectable performance differences. However, Chen et al. (2023) state that, "the effects of very large differences in element interactivity can be readily demonstrated" (p. 16). Given that most math clusters contain substantially more interacting elements than the art clusters, combined with participants' lower prior knowledge of the math content, the present study's lessons appear to successfully reflect the expected difference in element interactivity between the two domains.

3.5 Measures

3.5.1 Learning: Immediate & Retention

Learning was measured to assess participants' immediate understanding of the lesson content and their retention of that content one week later. Participants completed a pre-test, containing both subjective ratings of prior knowledge and objective test questions, before watching each of the two video lessons and completed a post-test immediately afterward. The post-test included recognition, recall, and transfer questions at three levels (near, medium, far). Immediate learning was operationalized as post-test scores after statistically adjusting for pre-test performance. Retention was assessed using a delayed test one week later. The retention test included the same types of questions as the immediate post-test. Appendix D defines each question type (Table D1) and provides examples and frequency counts from the math and art post-tests (Table D2 and Table D3). Appendix E and Appendix F contain example questions from the math and art pre-tests, respectively.

Because Lesson was a within-subjects factor, each participant had a pre-test score and post-test score for both the math and art lessons, which were used to assess immediate learning. In contrast, because the retention test was taken only once by each participant (during Session 2, for the Session 1 lesson), Lesson was treated as a between-subjects measure for this variable (see section *3.7 Procedure* for more details). For comparison purposes, all test scores were converted from raw scores to proportions (0–1 scale) prior to analysis.

3.5.2 Motivation

Motivation was measured to assess how participants' motivational responses to the learning materials differed across conditions. Motivation was measured using the reduced version of the 36-item Instructional Materials Motivation Survey (IMMS; Keller, 2010), the RIMMS (Loorbach et al., 2015; Appendix G). The original IMMS assesses motivation based on the ARCS model (Keller, 1987) and includes four subscales: attention, relevance, confidence, and satisfaction. A 2015 study by Loorbach et al. found that the IMMS had poor model fit ($\chi^2/df = 1.81$, TLI = 0.61, RMSEA = 0.09). The authors addressed this by removing items with correlated errors or poor factor loadings, resulting in a 12-item version.

The RIMMS demonstrated improved model fit ($\chi^2/df = 1.80$, TLI = 0.90, RMSEA = 0.09), likely due to the more parsimonious structure offered by the reduced number of items (Loorbach et al., 2015, p. 214). Thus, the RIMMS was used in the present study to measure motivation.

3.5.3 Subjective Workload

Subjective workload was measured to determine how participants' perceived cognitive load levels were affected by the study's conditions. Subjective workload was measured by the NASA-Task Load Index (NASA-TLX; Hart & Staveland, 1988; Appendix H). This survey uses six subscales to measure workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. The NASA-TLX was selected due to its long-standing use in the field of engineering psychology as well as its use in several studies involving PAs (Brachten et al., 2020; de Melo et al., 2020; Kim et al., 2020; Lin et al., 2013; Shah, 2020).

3.5.4 PA Persona

PA persona was measured to assess how the PA in the With-PA condition was perceived by participants compared to the arrows used in the No-PA condition. PA persona was measured using the revised version of the Agent Persona Instrument (API; Ryu & Baylor, 2005), the API-R (Schroeder et al., 2018; Appendix I). The original API is a widely used measure for PA persona that utilizes four subscales: humanlikeness, engagement, credibility, and ability to facilitate learning. A 2017 evaluation by Schroeder et al. found issues with the API, including a mediocre-to-poor model fit (CFI = 0.97, χ^2/df = 5.73, RMSEA = 0.11), items with cross-loadings, and a failure to meet the Rasch model's validity expectations due to multidimensionality. Therefore, in 2018, Schroeder et al. revised the API to the API-R by rewording several of the original API items and removing items that had cross-loadings.

Schroeder et al. (2018) found that the API-R showed improved model fit (CFI = 0.99, χ^2/df = 1.90, RMSEA = 0.066) and had strong internal consistency across its subscales: credibility (α = 0.86), humanlikeness (α = 0.87), ability to facilitate learning (α = 0.93), engagement (α = 0.91). It also exhibited less cross-loading than the original API. Thus, the API-R was used in the present study to measure PA persona.

3.5.5 Opinions Towards Lesson and PA/Arrows

An opinion-based survey designed by the author assessed participants' attitudes towards the video lesson and the PA or arrows (depending on their PA Assignment condition). The survey was comprised of 5-point Likert scale, multiple-choice, and text-entry questions and asked about:

1. Lesson content (e.g., difficulty, perceived level of learning)
2. Lesson design (e.g., background, fonts, length, speed)
3. PA/arrows characteristics (e.g., appearance, movements, naturalness)
4. Voice (e.g., naturalness, speed)
5. PA/arrows in general (e.g., level of distraction, desire to learn with aid in the future)

Participants in both the With-PA and No-PA conditions received questions on all topics. There were three questions that only participants in the With-PA condition received due to their inapplicability to the No-PA condition; these asked about the PA's gender, about the naturalness of its movements, and distinguished between gesturing and non-gesturing movements. When a question referenced the "pedagogical agent" for With-PA participants, the equivalent question referenced the "arrows" for No-PA participants. Given the length of this survey, it was distributed only once, at the end of Session 2. Example questions can be found in Appendix J.

No predictions were made for this survey. Instead, the survey's purpose was to guide the redesign of the PA prior to its use in Experiment 2 if Hypothesis 1.1 of Experiment 1 was not supported. This is further explained in Chapter 4: Experiment 2. Additionally, responses to this survey were examined in exploratory analyses, discussed in the section *3.8.4 Exploratory Analyses*.

For the exploratory analyses, survey responses were graded and converted to z-scores. The z-scores for questions on the content and design of the lesson were averaged to create a composite score reflecting overall opinions of the lesson. The z-scores for questions on the characteristics, voice, and general attitudes toward the PA/arrows were averaged to create a composite score reflecting overall opinions of the PA/arrows. The three questions received by only those in the With-PA condition were excluded from the composite scores to allow for better between-group comparisons.

3.6 Procedure

3.6.1 Session 1

Session 1 was in person, conducted in a university research laboratory, and had an average duration of approximately 64 minutes. The procedure for Session 1 was as follows:

1. Participants read and agreed to the informed consent statement.
2. Participants responded to the demographic survey (Appendix K).
3. Participants took the pre-test for their assigned Lesson condition, either Math (Bernoulli trials) or Art (principles of design).
4. Participants watched the video lesson corresponding to their Lesson condition, which either included a PA (With-PA) or arrows (No-PA), depending on their randomly assigned PA Assignment condition.
5. Participants responded to the NASA-TLX, RIMMS, and API-R.
6. Participants took the post-test matching their Lesson condition.

All materials were presented on the computer using Qualtrics, an online survey platform.

3.6.2 Session 2

Session 2 took place in person one week after Session 1 and had an average duration of approximately 56 minutes. It was conducted in the same university research laboratory as Session 1. The procedure for Session 2 was as follows:

1. Participants took the retention test corresponding to their Session 1 Lesson.
2. Participants took the pre-test for the Lesson condition they did not receive in Session 1.
    a. Those assigned to Math in Session 1 were assigned to Art in Session 2, and vice versa.
3. Participants watched the video lesson corresponding to their Session 2 Lesson condition. The PA Assignment condition remained the same as in Session 1.
4. Participants responded to the NASA-TLX, RIMMS, and API-R.
5. Participants took the post-test for their Session 2 Lesson.
6. Participants responded to the opinion-based survey.

All materials were presented on the computer using Qualtrics.

3.7 Results

3.7.1 Participants

Data from 33 participants (11 male, 21 female, 1 non-binary) were collected and used in the analyses for Experiment 1. For the distribution of participants across conditions, see Table 3.

All participants were undergraduate students, with ages ranging from 18 to 24 (M = 19.82; SD = 1.78). The majority of participants were either freshman (n = 13) or seniors (*n* = 11). Specific college majors were placed into broader categories; the majority of participants fell in the categories of either STEM (*n* = 14) or Psychology & Social Sciences (*n* =14), with the others being in Business & Management (*n* = 2) or Life Sciences (*n* = 2). Most participants had not taken and were not currently enrolled in a statistics course (*n* = 20) or an art or design course (*n* = 31), indicating an overall low level of prior knowledge on the experiment's lesson topics.

Table 3 – Experiment 1 participant counts by PA Assignment (between-subjects) and Lesson (within-subjects).

	Between-Subjects	
Within-Subjects	PA Assignment	
Lesson	PA	No-PA
Session 1 = Math Session 2 = Art	8	8
Session 1 = Art Session 2 = Math	9	8

3.7.2 Hypothesis 1.1

3.7.2.1 Analyses

A 2×2 mixed-design ANOVA was conducted on each of the following dependent variables: motivation, subjective workload, and PA persona. The outcome variables were RIMMS composite scores, NASA-TLX composite scores, and API-R composite scores, respectively. Each ANOVA included PA Assignment (between-subjects: With-PA vs. No-PA) and Lesson (within-subjects: Math vs. Art) as independent variables.

To account for baseline differences in prior knowledge, a 2×2 mixed-design ANCOVA was used for the immediate learning dependent variable. The same independent variables (PA Assignment and Lesson) were used, with post-test scores as the outcome variable and pre-test scores included as a covariate.

Retention was measured only for one lesson (during Session 2, to assess retention of the Session 1 lesson), so a 2×2 between-subjects ANCOVA was conducted. The same independent variables (PA Assignment and Lesson) were included with the covariate of pre-test scores.

Holm-Bonferroni correction was applied to the *p*-values for all analyses to adjust for multiple comparisons (reported as p_{adjusted}). Effect sizes (η_p^2) are also reported. For Hypotheses 1.1A–E, the primary effect of interest was the main effect of PA Assignment. However, the main effect of Lesson and the interaction effect of PA Assignment × Lesson are also reported. See Table 5 on page 55 for descriptive statistics and Table 7 on page 59 for a summary of results.

3.7.2.2 Assumption Checks

Shapiro-Wilk tests and Levene's tests were conducted for each ANOVA to assess the assumptions of normality and homogeneity of variances. In most cases, the Shapiro-Wilk tests did not indicate significant deviations from normality ($p > .05$). The exception was the within-subjects residuals of the NASA-TLX composite scores ($W = 0.93$, $p = .044$), but Levene's test indicated homogeneity of variances ($p = .757$). Given that ANOVAs are generally robust to deviations from normality when variances are equal and group sizes are approximately balanced (and that the Q-Q plot did not reveal major concerns), these results suggest no meaningful violations. Levene's tests for all other ANOVAs also indicated equal variances (i.e., $p > .05$).

Shapiro-Wilk and Levene's tests were also conducted for the immediate learning and retention ANCOVAs. The Shapiro-Wilk tests indicated a deviation from normality for both immediate learning ($W = 0.9$, $p < .001$) and retention ($W = 0.9$, $p = .02$), but Levene's tests showed homogeneity of variances ($p = .178$; $p = 0.111$). As with ANOVA, ANCOVA is generally robust to normality violations when variances are equal and group sizes are approximately balanced. Additionally, neither Q-Q plots were concerning. Scatterplots indicated weak but approximately linear relationships between pre-test scores and post-test scores and between pre-test scores and retention scores. The assumption of homogeneity of regression slopes was also met for both, as all interaction terms involving pre-test scores were nonsignificant ($p > .05$).

3.7.2.3 Hypothesis 1.1A

Hypothesis 1.1A predicted that participants in the With-PA condition would exhibit higher immediate learning than those in the No-PA condition. This hypothesis was not supported: a 2×2 mixed-design ANCOVA found no significant main effect of PA Assignment, $F(1, 30) = 0.93$, $p = .344$ ($p_{\text{adjusted}} = 1.0$), $\eta_p^2 = 0.03$. The interaction between PA Assignment and Lesson was also nonsignificant, $F(1, 30) = 0.44$, $p = .515$ ($p_{\text{adjusted}} = 1.0$), $\eta_p^2 = 0.01$.

The main effect of Lesson was significant, $F(1, 30) = 12.76$, $p = .001$ ($p_{\text{adjusted}} = .006$), $\eta_p^2 = 0.30$ (large effect; Table 4; Cohen, 1988), with participants demonstrating higher post-test scores in math ($M = 80.37$, $SD = 17.84$) than in art ($M = 71.54$, $SD = 10.83$).

Pre-test scores were included as a covariate but were not a significant predictor of post-test performance (p = .282).

Table 4 – Cohen's (1988) effect size guidelines for partial eta squared (η_p^2)

Effect Size	Partial Eta Squared (η_p^2)
Small	0.01
Medium	0.06
Large	0.14

3.7.2.4 Hypothesis 1.1B

Hypothesis 1.1B predicted that participants in the With-PA condition would exhibit higher motivation than those in the No-PA condition. This hypothesis was not supported: a 2×2 mixed-design ANOVA found no significant main effect of PA Assignment on composite RIMMS scores, $F(1, 31) = 2.26$, $p = .143$ ($p_{adjusted} = .715$), $\eta_p^2 = 0.07$. The effect of Lesson was also nonsignificant, $F(1, 31) = 0.02$, $p = .881$ ($p_{adjusted} = 1.0$), $\eta_p^2 = 0.0007$, as was the interaction between PA Assignment and Lesson, $F(1, 31) < 0.01$, $p = .944$ ($p_{adjusted} = 1.0$), $\eta_p^2 = 0.0002$.

3.7.2.5 Hypothesis 1.1C

Hypothesis 1.1C predicted that participants in the With-PA condition would exhibit lower subjective workload than those in the No-PA condition. This hypothesis was not supported: a 2×2 mixed-design ANOVA found no significant main effect of PA Assignment on composite NASA-TLX scores, $F(1, 31) = 0.91$, $p = .349$ ($p_{adjusted} = 1.0$), $\eta_p^2 = 0.03$. Lesson was also nonsignificant, $F(1, 31) = 0.06$, $p = .803$ ($p_{adjusted} = 1.0$), η_p^2 =

0.002, as was the interaction between PA Assignment and Lesson, $F(1, 31) = 0.49$, $p = .491$ ($p_{\text{adjusted}} = 1.0$), $\eta_p^2 = 0.02$.

3.7.2.6 Hypothesis 1.1D

Hypothesis 1.1D predicted that participants in the With-PA condition would provide higher PA persona evaluations than those in the No-PA condition. This hypothesis was not supported: a 2×2 mixed-design ANOVA found no significant main effect of PA Assignment on composite API-R scores, $F(1, 31) = 0.05$, $p = .832$ ($p_{\text{adjusted}} = 1.0$), $\eta_p^2 = 0.001$. Lesson was also nonsignificant, $F(1, 31) = 2.92$, $p = .098$ ($p_{\text{adjusted}} = .391$), $\eta_p^2 = 0.09$, as was the interaction between PA Assignment and Lesson, $F(1, 31) = 0.04$, $p = .851$ ($p_{\text{adjusted}} = 1.0$), $\eta_p^2 = 0.001$

3.7.2.7 Hypothesis 1.1E

Hypothesis 1.1E predicted that participants in the With-PA condition would exhibit higher retention than those in the No-PA condition. This hypothesis was not supported: a 2×2 between-subjects ANCOVA found no significant main effect of PA Assignment on retention scores, $F(1, 28) = 1.12$, $p = .298$ ($p_{\text{adjusted}} = 1.0$), $\eta_p^2 = 0.04$. The main effect of Lesson was also insignificant, $F(1, 28) = 2.41$, $p = .132$ ($p_{\text{adjusted}} = .396$), $\eta_p^2 = 0.08$, as was the interaction between PA Assignment and Lesson was also nonsignificant, $F(1, 28) = 1.96$, $p = .172$ ($p_{\text{adjusted}} = 0.86$), $\eta_p^2 = 0.07$. Pre-test scores were included as a covariate but were not a significant predictor of retention performance ($p = .053$).

Table 5 – Descriptive statistics by PA Assignment, Lesson, and their interaction for Hypotheses 1.1A–E analyses.

	PA Assignment				Lesson				Interaction							
	With-PA		No-PA		Math		Art		With-PA/Math		With-PA/Art		No-PA/Math		No-PA/Art	
Dependent Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Post-Test	73.76	17.95	78.29	11.71	80.37	17.84	71.54	10.83	77.22	22.01	70.29	12.44	83.72	11.80	72.86	9.02
Motivation	2.88	0.55	3.14	0.69	3.02	0.66	2.99	0.61	2.89	0.53	2.86	0.58	3.15	0.77	3.14	0.63
Subjective Workload	5.05	2.79	5.84	2.58	5.49	2.86	5.38	2.58	5.27	3.02	4.84	2.62	5.73	2.75	2.97	0.61
PA Persona	3.07	0.51	3.03	0.68	3.13	0.64	2.98	0.55	3.16	0.53	2.99	0.49	3.10	0.75	5.95	2.48
Retention	0.75	0.19	0.82	0.16	85.34	16.54	71.95	16.59	0.78	0.21	71.79	17.73	92.24	6.32	72.12	16.41

Note. Possible ranges of the variables: post-test (0–100), motivation (1–5), subjective workload (1–21), PA persona (1–5), retention (0–100).

3.7.3 Hypothesis 1.2

3.7.3.1 Analyses

For Hypotheses 1.2A–E, only data from participants assigned to the With-PA condition were analyzed (n = 17). To account for the use of repeated measures, a paired samples t-test with the independent variable Lesson (Math vs. Art) was conducted on each of the following dependent variables: motivation, subjective workload, and PA persona.

To account for baseline differences in prior knowledge, a one-way ANCOVA with repeated measures was used for the immediate learning dependent variable. Lesson was the independent variable, post-test scores were the dependent variable, and pre-test scores were included as a covariate. Similarly, a one-way ANCOVA was conducted for retention with the same independent variables and covariate and with retention scores as the dependent variable.

Holm-Bonferroni correction was applied to p-values to adjust for multiple comparisons (reported as p_{adjusted}). Effect sizes (η_p^2, d_z, d) are also reported. See Table 6 on page 58 for descriptive statistics and Table 7 on page 59 for a summary of results.

3.7.3.2 Assumption Checks

For the paired samples *t*-tests, box plots were generated to inspect for outliers and skew. The distribution of the composite API-R difference scores appeared approximately symmetric and showed no outliers, the composite RIMMS difference scores showed moderate negative skew with two mild outliers, and the composite NASA-TLX difference scores showed positive skew with two lower outliers. However, Shapiro-Wilk tests for all three variables indicated no significant deviations from normality ($p > .05$), suggesting that the observed skew and outliers were not cause for concern.

Shapiro-Wilk and Levene's tests were conducted for the immediate learning and retention ANCOVAs. The Shapiro-Wilk test for immediate learning indicated a deviation from normality ($W = 0.88$, $p = .001$). However, the Q-Q plot was not concerning, and Levene's test indicated homogeneity of variances ($p = .277$). For retention, neither the Shapiro-Wilk test ($W = .91$, $p = .1$) nor Levene's test ($p = .524$) indicated assumption violations. The scatterplots indicated a weak but roughly linear relationship between pre-test scores and post-test scores and between pre-test scores and retention scores. The assumption of homogeneity of regression slopes was also met for both, as the interaction terms involving pre-test scores were nonsignificant ($p > .05$).

3.7.3.3 Hypothesis 1.2A

Hypothesis 1.2A predicted that, for participants who learned with a PA, immediate learning would be higher in math than in art. This hypothesis was not supported: a one-way ANCOVA found no significant effect of Lesson on post-test scores when pre-test scores were included as a covariate, $F(1, 15) = 3.92$, $p = 0.07$ ($p_{\text{adjusted}} = .333$), $\eta_p^2 = 0.21$.

Additionally, pre-test scores were not a significant predictor of post-test performance (p = .697).

3.7.3.4 Hypothesis 1.2B

Hypothesis 1.2B predicted that, for participants who learned with a PA, motivation would be higher in math than art. This hypothesis was not supported: a paired samples t-test found no significant difference in composite RIMMS scores between the math and art groups, $t(16) = 0.16$, $p = .872$ ($p_{adjusted} = .872$), $d_z = 0.04$.

3.7.3.5 Hypothesis 1.2C

Hypothesis 1.2C predicted that, for participants who learned with a PA, subjective workload would be lower in math than in art. This hypothesis was not supported: a paired samples t-test found no significant difference in composite NASA-TLX scores between the math and art groups, $t(16) = 1.45$, $p = .166$ ($p_{adjusted} = .387$), $d_z = 0.36$.

3.7.3.6 Hypothesis 1.2D

Hypothesis 1.2D predicted that, for participants who learned with a PA, PA persona ratings would be higher in math than art. This hypothesis was not supported: a paired samples t-test found no significant difference in composite API-R scores between the math and art groups, $t(16) = 1.74$, $p = .101$ ($p_{adjusted} = .387$), $d_z = 0.43$.

3.7.3.7 Hypothesis 1.2E

Hypothesis 1.2E predicted that, for participants who learned with a PA, retention would be higher in math than art. This hypothesis was not supported: a one-way ANCOVA found no significant effect of Lesson on retention scores when pre-test scores were included as a covariate, $F(1, 14) = 3.17$, $p = 0.1$ ($p_{adjusted} = .387$), $\eta_p^2 = 0.18$. Additionally, pre-test scores were not a significant predictor of retention performance ($p = .081$).

Table 6 – Descriptive statistics by Lesson for Hypotheses 1.2A–E analyses.

	Lesson			
	Math		Art	
Dependent Variable	*M*	*SD*	*M*	*SD*
Post-Test	77.22	22.01	70.29	12.44
Motivation	2.89	0.53	2.86	0.58
Subjective Workload	5.27	3.02	4.84	2.62
PA Persona	3.16	0.53	2.99	0.49
Retention	78.45	20.91	71.79	17.73

Note. Analyses performed only on participants in With-PA condition (n = 17).
Possible ranges of the variables: post-test (0–100), motivation (1–5), subjective workload (1–21), PA persona (1–5), retention (0–100).

Table 7 – Summary of results for Experiment 1 Hypotheses 1.1A–E and 1.2A–E, with mean directions shown for significant effects.

Data	Hypothesis	Analysis	Dependent Variable	Main Effect: PA Assignment	Main Effect: Lesson	Interaction: PA Assignment × Lesson
				p_{adjusted}		
All Participants	Hyp1.1A	2×2 mixed ANCOVA	Immediate learning	1.0	.006** Math > Art	1.0
	Hyp1.1B	2×2 mixed ANOVA	Motivation	.715	1.0	1.0
	Hyp1.1C	2×2 mixed ANOVA	Subjective workload	1.0	1.0	1.0
	Hyp1.1D	2×2 mixed ANOVA	PA persona	1.0	.391	1.0
	Hyp1.1E	2×2 between-subjects ANCOVA	Retention	1.0	.396	.86
Only With-PA Participants	Hyp1.2A	One-way within-subjects ANCOVA	Immediate learning		.333	
	Hyp1.2B	Paired samples t-test	Motivation		.872	
	Hyp1.2C	Paired samples t-test	Subjective workload		.387	
	Hyp1.2D	Paired samples t-test	PA persona		.387	
	Hyp1.2E	One-way between-subjects ANCOVA	Retention		.387	

p adjusted = p-value after Holm-Bonferroni correction

**Statistically significant at $p < .05$

3.7.4 Exploratory Analyses

3.7.4.1 <u>Analyses</u>

As described in the earlier section *3.6.5 Opinions Towards Lesson and PA/Arrows*, responses to the opinion-based survey were averaged to create two composite scores, one for lesson-based opinions and another for opinions on the PA or arrows. Separate 2×2 between-subjects ANOVAs were conducted on the two composite scores; PA Assignment and Lesson were included as the independent variables. Because the opinion-based survey was taken only once by each participant at the end of Session 2, there was no need to account for repeated measures. Effect sizes (η_p^2) are reported. See Table 8 for descriptive statistics.

3.7.4.2 Assumption Checks

To assess the assumptions of normality and homogeneity of variances, Shapiro-Wilk tests and Levene's tests were conducted for each ANOVA. All tests were nonsignificant ($p > .05$), indicating no assumption violations.

3.7.4.3 Findings

For the lesson opinions composite *z*-scores, a two-way ANOVA found no significant main effect of PA Assignment, $F(1, 29) = 0.07$, $p = .794$, $\eta_p^2 = .002$, or of Lesson, $F(1, 29) = 1.20$, $p = .283$, $\eta_p^2 = .04$. The interaction between PA Assignment and Lesson was also nonsignificant, $F(1, 29) = 0.03$, $p = .867$, $\eta_p^2 = .0009$.

For the PA/arrows opinions composite *z*-scores, a two-way ANOVA found no significant main effect of PA Assignment, $F(1, 29) = 0.83$, $p = .369$, $\eta_p^2 = .03$, or of Lesson, $F(1, 29) = 1.98$, $p = .170$, $\eta_p^2 = .06$. The interaction between PA Assignment and Lesson was also nonsignificant, $F(1, 29) = 1.45$, $p = .239$, $\eta_p^2 = .05$.

Table 8 – Descriptive statistics for *z*-scores by PA Assignment, Lesson, and their interaction for Experiment 1's exploratory analyses.

	PA Assignment				Lesson				Interaction							
	With-PA		No-PA		Math		Art		With/Math		With/Art		No/Math		No/Art	
Dependent Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Lesson Opinions	−0.02	0.36	0.02	0.36	0.07	0.36	−0.07	0.34	0.04	0.46	−0.08	0.22	0.10	0.25	−0.06	0.44
PA/Arrows Opinions	−0.08	0.51	0.08	0.50	−0.12	0.57	0.13	0.42	−0.29	0.49	0.16	0.45	0.07	0.62	0.10	0.40

Note. Descriptive statistics are for standardized *z*-scores. Observed *z*-score ranges: lesson opinions (−0.90 to 0.68) and PA/arrows opinions (−1.53 to 1.21).

3.7.4.4 Specific Responses

Particular attention was given to questions about the PA's adjustable characteristics, as responses to these items were used to determine what changes should be made to the PA before its use in Experiment 2. The responses that were most useful were to questions about the naturalness of the audio (i.e., voice) and the rate of speech. Almost half of participants (48.5%) indicated that they found the voice to be either "very unnatural" or "sort of unnatural." The same proportion of participants (48.5%) reported feeling like the voice spoke "way too slow" or "a bit too slow." Changes to the PA and voice prior to use in Experiment 2 are discussed later in section *4.4 Pedagogical Agent*.

3.8 Discussion

Hypothesis 1.1 predicted that participants in the With-PA condition would demonstrate better outcomes than those in the No-PA condition, meaning that participants who learned with a PA would perform better than those who learned with arrows instead. This hypothesis was not supported, as PA Assignment had no significant effect on any dependent variables. The lack of difference in motivation between the two conditions is contrary to previous studies that found a positive PA effect (Dincer & Doganay, 2017), while for cognitive load, the null result aligns with prior research suggesting that PAs have little or no impact on this variable (Davis, 2018; Frechette & Moreno, 2010; Moreno et al., 2001; Schroeder, 2017; Yusoff et al., 2023). Further, the absence of a PA Assignment effect on immediate learning or retention is contrary to the claim of PA proponents that the presence of a PA enhances learning (Kizilkaya & Asher, 2008; Moreno et al., 2001; Schroeder et al., 2013). It also challenges the embodiment principle (Mayer & DaPra, 2012), which emphasizes the benefits of humanlike social cues such as "gesturing, facial expression, eye gaze, and human-like movement" (p. 239). In the present study, the arrows in the No-PA condition provided only one of these cues (gesture), while the PA provided all four, yet no learning differences were observed.

One explanation for the null effect on learning lies in Social Agency Theory (see section *1.1.1.1 Social Agency Theory*; Mayer et al., 2003b), which posits that social cues presented in an educational context elicit social conversation schemas, fostering deeper cognitive processing and more meaningful learning. Although the arrows in the No-PA condition lacked a humanlike appearance, they still conveyed social cues, including deictic gestures and voice narration, which might have been sufficient to activate the social schema. The With-PA condition included additional cues, such as eye gaze, facial expressions, and non-deictic movements. However, if the cues in the No-PA condition were adequate to elicit the benefits proposed in Social Agency Theory, the PA's additional cues might not have provided further benefit. This ceiling effect interpretation suggests that the learning benefits often attributed to PAs could stem from their basic communicative features rather than from their full embodiment or complex social presence.

The statistically equivalent results between the PA and the arrows suggest two possible interpretations. First, this could mean that PAs and arrows are equally effective at enhancing learning outcomes. This would have practical implications for lesson designers and instructors, as it would mean the additional time and effort required to create and implement a PA is not necessary. This interpretation is supported by the similar PA persona ratings between the conditions, suggesting that participants found the PA and arrows to be equally credible, humanlike, engaging, and able to facilitate learning. Again, this is consistent with Social Agency Theory, due to the No-PA condition presenting some social cues. Alternatively, these results could mean that neither PAs nor arrows improve learning relative to having no learning aid. The inclusion of a traditional control condition without any such aid would be required to test these explanations.

Only the main effect of Lesson was found to be significant in the analyses for Hypothesis 1.1. Participants showed significantly higher immediate learning for the math lesson compared to the art lesson. Although the higher performance in math is counterintuitive given the estimated higher element interactivity of the lesson (see section *3.4.2.1 Measuring Element Interactivity*), it could reflect characteristics of the participant sample. Experiment 1's participants were recruited from a technical university where the most common majors are computer science and engineering, as evidenced by 46.47% of participants reporting a STEM major (45% in computer science or engineering specifically), and none reporting a major in the humanities. As such, it is possible that participants were generally more familiar with or interested in math than art.

Hypothesis 1.2 predicted that within the With-PA condition, participants would have better outcomes in math than in art. This hypothesis was informed by Schroeder et al.'s (2013) meta-analysis, which found that PAs had a larger effect in studies involving mathematics ($g = .27$, $p < .05$) compared to those involving the humanities ($g = .06$, $p > .05$). This hypothesis was not supported, with no significant effect of Lesson on any dependent variables. One explanation for this null result is that the PA used in the With-PA condition was simply not effective enough to induce similar findings to Schroeder et al. (2013), as supported by the absence of a main effect of PA Assignment. An alternative explanation is that the math lesson was too high in element interactivity (see Appendix C), with its complexity bringing performance levels closer to that of the art lesson.

Ultimately, the results of Experiment 1 challenge prior conclusions that the presence of a PA inherently improves learning outcomes. Although many previous studies have found positive effects of PAs, the current findings suggest that when functionality is controlled (by using arrows to provide the same attentional cues as the PA), the social presence added by a PA might not be enough to enhance learning or motivation, or to reduce cognitive load. Although not the original intention of the present experiment, it contributes to the literature by assessing the impact of the PA's social presence separate from its instructional role. However, without a true control condition (i.e., no PA or arrows), it is unclear whether the two learning aids were equally *effective* or equally *ineffective*.

The next chapter discusses Experiment 2, which expanded on the findings of Experiment 1 by examining whether individual differences influence PA-enhanced learning. Specifically, it investigated whether outcomes differed when participants were assigned to a PA condition that aligned or conflicted with their indicated preference.

CHAPTER 4. EXPERIMENT 2

As discussed in the previous section *1.2.3 Preference*, the effect of preference matching on learning outcomes in PA-enhanced environments has not yet been investigated, although the existence of such an effect has been observed in non-PA learning (Cinkara and Bagceci, 2013; Freitag and Sullivan, 1995). This lack of research is problematic given that some learners dislike PAs (Hook et al., 2000) and that many environments utilizing PAs do not offer learners the option to disable them (Graesser et al., 2005; Hone et al., 2005; Krämer et al., 2007). Given this context, it is important to examine whether individual preference for a PA impacts learning outcomes. Therefore, Experiment 2 examined how congruence and incongruence between a learner's preference for a PA and their assigned PA condition affected learning outcomes. Based on the results of Experiment 1, Experiment 2 used the same math lesson and a modified version of the original PA (see section *4.4 Materials*).

4.1 Design

A 2×2 between-subjects design was used with the dependent variables learning, subjective workload, motivation, PA persona, opinions on the lesson, and opinions on the assigned learning aid. The first independent variable was PA Preference with two levels: prefers a PA versus does not prefer a PA. PA Preference was a subject variable, rather than a random variable, because it was determined for each participant by their responses to a survey (Educational Technology Preferences Survey, discussed in section *4.6 Measures* of Experiment 2). The second independent variable was PA Assignment with two levels: assigned to learn with a PA (With-PA) versus without a PA (No-PA).

The combination of PA Preference and PA Assignment resulted in two congruent conditions and two incongruent conditions. In the congruent conditions, the participant's PA Preference and PA Assignment matched. In the incongruent conditions, they conflicted. These condition combinations are shown in Table 9.

Table 9 – Combinations of PA Preference and PA Assignment conditions to form two congruent and two incongruent conditions.

	PA Assignment	
PA Preference	Assigned a PA	Not assigned a PA
Prefers a PA	Congruent	Incongruent
Does not prefer a PA	Incongruent	Congruent

4.2 Hypotheses

1. Hypothesis 2.1: Participants whose PA preferences are honored will have better outcomes than those whose PA preferences are not honored (congruent conditions will have better outcomes than incongruent conditions), such that:
    A. They will exhibit higher learning (recall, recognition, and transfer).
    B. They will report higher motivation.
    C. They will report lower workload.
    D. They will report higher PA persona.
2. Although no significant differences were observed between the With-PA and No-PA conditions in Experiment 1, the PA was revised prior to use in Experiment 2. Therefore, it is hypothesized that participants in the With-PA condition will have better outcomes than those in the No-PA condition, such that:

A. They will exhibit higher learning (recall, recognition, and transfer).
B. They will report higher motivation.
C. They will report lower workload.
D. They will report higher PA persona.

4.3 Participant Recruitment

A priori analysis for an ANOVA was conducted in G*Power (Faul et al., 2007) with a medium effect size of $f = .25$. This test indicated that a sample size of 128 total participants would be required to achieve a power of .8 using the described design.

Initially, participants were recruited from the same large technical university in the southeastern United States as in Experiment 1. These participants were given the choice between SONA credit and $10 for compensation. Inclusion criteria comprised normal or corrected-to-normal hearing, vision, and color vision. Students who (1) participated in a previous PA study by the author or (2) participated in Experiment 1 were ineligible to participate in Experiment 2.

After several months of low university-based recruitment, a switch was made to the online recruitment platform Prolific (2024). For Experiment 2's statistical analyses, only the data from Prolific participants were used, as discussed in the section *4.8.2 Prolific versus University Participants*. These participants were awarded $12 for successful completion of the study, consistent with Prolific's $8 per hour policy.

For inclusion criteria, Prolific's pre-screening questions were utilized. These are questions taken by Prolific participants before they become eligible to participate in studies. For Experiment 2, participants were required to reside in the United States, have English as their first language, maintain a prior study approval rating of 99–100%, and not have participated in the researcher's previous Prolific PA studies. To better match the university-based sample, participants also had to be (1) either enrolled in a technical/community college or in an undergraduate program and (2) be between 18 and 30 years old. Prolific participants could be excluded early for two reasons: (1) their condition's quota had already been met (explained in section *4.8.1 Participants*), or (2) they failed more than one attention check (Appendix L). Participants excluded for the first reason received partial compensation of $4 for 15–30 minutes of participation, whereas those excluded for the second reason did not receive any compensation. This compensation plan was outlined in the informed consent document (Appendix M).

4.4 Materials

4.4.1 Pedagogical Agent

Because Hypothesis 1.1 from Experiment 1 was not supported, meaning the PA did not significantly improve learning outcomes compared to the No-PA condition, the PA was modified before use in Experiment 2. Changes were informed by the results of the opinion-based survey from Experiment 1 (discussed in the previous section *3.6.5 Exploratory Analyses: Opinion-Based Survey*). The most common point of criticism across both PA Assignment conditions was the voice, with participants deeming it unnatural and too slow.

In Experiment 1, the AI Voice, "Ava", from the text-to-speech (TTS) engine NaturalReader was used at a rate of 180 words per minute (WPM). For Experiment 2, this was changed to "Onyx" with a faster speaking rate of 185 WPM. Although 185 WPM falls slightly outside the typical speaking range of 150–180 WPM (Pastore & Ritzhaupt, 2015; Ritzhaupt & Barron, 2008; Ritzhaupt et al., 2015), it remains well below the undesirable range of 219–275 WPM (Barabasz, 1968; Reid, 1968; Ritzhaupt & Barron, 2008; Ritzhaupt et al., 2015).

The change from Ava to Onyx was made to enhance the naturalness of the audio. Ava is a “Plus” voice from NautralReader, developed using a neural TTS model and designed to be more natural than traditional robotic-sounding TTS audio. However, these voices can still sound somewhat monotone or artificial. Alternatively, Onyx is a cloned voice generated from audio samples submitted by a user through NaturalReader’s “Community” voice feature. Although trained on a smaller dataset, these cloned voices often sound more lifelike due to their ability to replicate the speaker’s unique vocal characteristics. To address minor inconsistencies introduced by the limited training data, the researcher manually edited the audio to produce more natural narration.

The other features of the PA remained the same as in Experiment 1. The No-PA condition also stayed the same, consisting of a pointing arrow in place of the PA. Again, the arrow performed the same deictic movements as the PA, with identical audio used in both conditions. To reiterate the justification for using arrows in the No-PA condition, the intention was to ensure the same level of visual attention guidance in both levels of the PA Assignment condition.

4.4.2 Lesson

The lesson for Experiment 2 was selected based on the results of Experiment 1. Because Hypotheses 1.2A–D were not supported, indicating no differences in the PA's effectiveness between the math and art lessons, the selection of the lesson for Experiment 2 was based on the main effect of Lesson from Experiment 1's analyses. As described in the previous section *3.8.2.3 Hypotheses 1.1A*, this effect was significant (p_{adjusted} = .006, η_p^2 = .30) on post-test scores after controlling for pre-test performance. On average, participants scored 8 percentage points higher on the math post-test than on the art post-test; scores were out of 100. Therefore, the math lesson on Bernoulli Trials was selected for use in Experiment 2.

4.5 Measures

4.5.1 PA Preference

4.5.1.1 Pre-Task

Preference for a PA was assessed using the Educational Technology Preferences Survey developed by the author. This survey was the first measure completed after the demographic questionnaire. It consisted of nine filler items and two target items. Each item contained a Yes/No question on preference (e.g., "…would you want to learn with x?") and a 1–5 Likert scale question on the strength of that preference (example questions in Appendix N). For each item, participants were prompted to watch a short (15-30 seconds) video prior to responding, which allowed them to see and/or hear each educational technology before providing their opinion on it.

Filler items (Appendix N, Figure N1) gauged opinions on the following educational technologies: digital flashcards, online educational games, educational phone applications, eBooks, websites with educational videos, and massive open online courses.

The two target questions were about PAs. One of the target questions included a video on computer networks featuring the PA used in the study (Appendix N, Figure N2); the other included a video on bird migration featuring an alternative PA (i.e., PA not used in the present study; Figure 11). Each target question stated, “The video above depicts an example of a **virtual instructor**. Please watch the video before answering the following questions.”

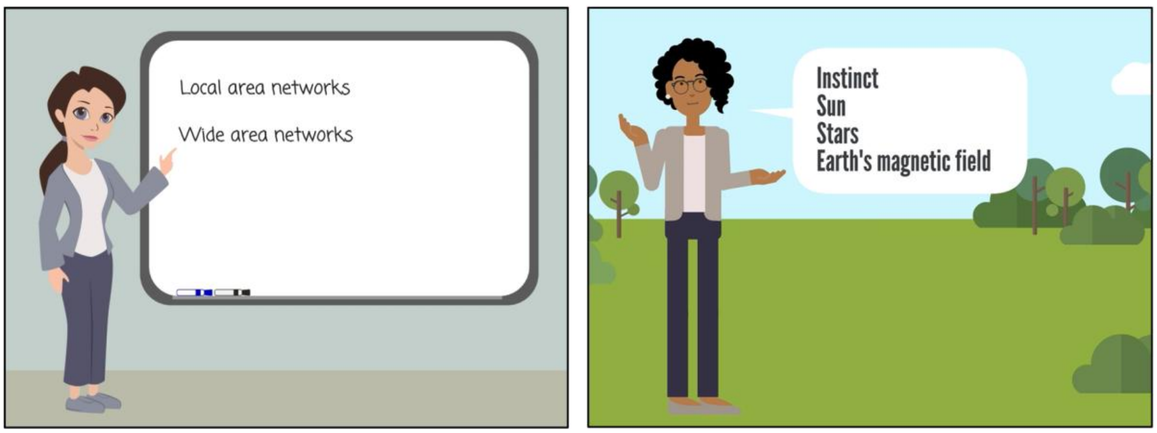


Figure 11 – Screen captures of target-question videos for the study’s PA (left) and alternative PA (right).

Two target questions were used instead of one to assess the consistency of a participant's responses. To prevent the PA target questions from appearing suspicious due to there being two of them, the survey also included two questions for the following filler educational technologies: digital flashcards, eBooks, and online educational games.

The target question about the study’s PA (Appendix N, Figure N2) was of primary importance. A participant’s response to this question identified their PA Preference condition; “Yes” indicated preference for a PA, and “No” indicated non-preference. If participants responded “Yes” for the study’s PA but “No” for the alternative PA, they were marked as “inconsistent.” The decision to not exclude inconsistent participants outright was based on two considerations: (1) the participant pool was limited when recruiting at the university, and (2) a person's opinions of a given PA might vary based on its unique attributes. Therefore, it was deemed most important to assess preference toward the specific PA that participants would see if assigned to the With-PA condition.

Pilot data were collected using the Educational Technology Preferences Survey to ensure that a sufficient number of participants who preferred and did not prefer a PA could be expected in the main study. Out of 21 university-recruited participants, 11 indicated a preference for a PA, and 10 indicated a preference against a PA.

4.5.1.2 Post-Task

The final question of the study reassessed each participant's PA preference, which was used to determine whether preference changed from pre- to post-task. This question can be seen in Appendix O.

4.5.2 Learning

The same math pre-test and immediate post-test from Experiment 1 were used in Experiment 2. Again, immediate learning was operationalized as post-test scores after controlling for pre-test performance, and all test scores were converted from raw scores to proportions (0–1 scale) prior to analysis to comparison purposes.

4.5.3 Motivation

As in Experiment 1, the Reduced Instructional Materials Motivation Survey (RIMMS; Loorbach et al., 2015; Appendix G) was used to assess motivation in Experiment 2.

4.5.4 *Subjective Workload*

As in Experiment 1, the NASA-Task Load Index (NASA-TLX; Hart & Staveland, 1988; Appendix H) was used to measure subjective workload in Experiment 2.

4.5.5 *PA Persona*

As in Experiment 1, the revised Agent Persona Instrument (API-R; Schroeder et al., 2018; Appendix I) was used to measure PA persona in Experiment 2.

4.5.6 *Opinions Towards PA and Lesson*

The same opinion-based survey used in Experiment 1 was used in Experiment 2. Like in Experiment 1, participants in both PA Assignment conditions received questions on all topics. The three questions not applicable to the No-PA condition (those on the PA's gender, naturalness of movements, and different gesture types) were presented only to participants in the With-PA condition. In questions where the "pedagogical agent" was referred to for With-PA participants, "arrows" were referred to for No-PA participants. See Appendix J for example questions from this survey.

Also analogous to Experiment 1, no results were hypothesized for this survey. Its purpose was to collect participant feedback on the lesson and learning aids. Survey responses were analyzed in exploratory analyses, with *z*-scores averaged into composite scores reflecting opinions of the lesson and of the PA/arrows. Again, the three questions given only to the With-PA condition were not used in the *z*-score calculations to allow for better between-group comparisons.

4.6 Procedure

For the university-recruited participants, Experiment 2 took place in person in the same university research laboratory as Experiment 1 and had an average duration of approximately 73 minutes. The procedure was as follows:

1. Participants read and agreed to the informed consent statement.
2. Participants responded to the demographic survey (Appendix K).
3. Participants read instructions for and responded to the Educational Technology Preference Survey (question order randomized for each participant).
4. Participants took the math pre-test.
5. Participants watched the math video lesson (either With-PA or No-PA).
6. Participants responded to the NASA-TLX, RIMMS, and API-R.
7. Participants took the math post-test.
8. Participants responded to the opinion-based survey.
9. Participants answered the post-task preference question.

All materials were presented on the computer using Qualtrics.

For the Prolific-recruited participants, a similar procedure was followed. Each participant was provided with a unique link to the Qualtrics survey, where the study was hosted. The average duration was approximately 92 minutes. Changes to the procedure for Prolific participants were as follows:

1. The first page contained a list of study requirements intended to enhance data quality (Figure 12).
2. Additional questions regarding current academic status were added to the demographic survey (Appendix P).
3. Attention checks were added at three points in the study (Appendix L).

Before you begin, please ensure you can meet the following requirements:

1. Complete this study on a **computer** (rather than a smartphone or tablet).
2. Use **headphones or speakers**, as the study includes **audio**. Make sure you are in a quiet environment where you can listen without distractions.
3. Complete this study in **one sitting**. Brief interruptions (e.g., for water or a restroom break) are completely understandable, but **extended breaks should be avoided**.

Figure 12 – List of requirements added to the first page of the study for Prolific participants.

4.7 Results

4.7.1 Participants

From the university recruitment site, data from 47 participants (17 male, 29 female, 1 non-binary) were collected. All participants were undergraduate or graduate (n = 3) students, with ages ranging from 18 to 30 (M = 20.47; SD = 2.14). The distribution of classes was roughly equal, with 11 freshmen, 12 sophomores, 8 juniors, and 12 seniors; one participant indicated they were a transfer student. Specific college majors were placed into broader categories. The majority of participants fell in the category of STEM (*n* = 28), the second largest category was Psychology & Social Sciences (*n* =13), and the others were in Arts, Design & Humanities (*n* = 2) or Life Sciences (*n* = 4). Around 60% of participants had not taken and were not currently enrolled in a statistics course (*n* = 28), indicating an overall mixed level of prior knowledge on the experiment's lesson topic. For the distribution of participants across conditions, see Table 10; given that the majority of participants did not prefer the study's PA (*n* = 31), and data collection was paused prior to completion, the group sizes are unbalanced.

Table 10 – Experiment 2 university-recruited participant assignment by PA Preference and PA Assignment (both between-subjects).

	PA Preference	
PA Assignment	Prefers	Does Not Prefer
With	8	15
Without	8	16

Data from 126 Prolific participants (49 male, 74 female, 3 non-binary) were used for analysis. All participants were undergraduate students, with ages ranging from 18 to 30 (*M* = 23.03; *SD* = 3.41). The majority of participants were sophomores (*n* = 30), juniors (*n* = 37), or seniors (*n* = 44), with only 14 freshmen. One student indicated they were non-traditional, pursuing their second Bachelor's degree. The majority of participants fell in the category of STEM (*n* = 46), the second largest category was Business & Management (*n* = 26), and the others were in Psychology & Social Sciences (*n* = 20), Arts, Design & Humanities (*n* = 12), Health & Medical Fields (*n* = 12), Life Sciences (*n* = 5), or Education (*n* = 3); two participants indicated an undeclared or general studies major. Around 69% of participants had not taken and were not currently enrolled in a statistics course (n = 87), indicating an overall mixed level of prior knowledge on the experiment's lesson topic. For the distribution of participants across conditions, see Table 11.

Table 11 – Experiment 2 Prolific participant assignment by PA Preference and PA Assignment (both between-subjects).

	PA Preference	
PA Assignment	Prefers	Does Not Prefer
With	30	31
Without	33	32

In total, 122 Prolific participants were excluded from analysis. The target sample size was 128, with equal group sizes requiring 64 participants who preferred a PA and 64 who did not. However, among the 248 total Prolific participants, approximately 71% indicated a preference for the study's PA, while only 29% did not. As a result, most exclusions (n = 104) occurred because the quota for the participant's PA Preference group had already been met. Additional exclusion factors were poor-quality or AI-suspected responses (n = 12) and not being an undergraduate student at the time of participation (n = 6). No participants were excluded due to attention checks, as no participants failed more than one.

4.7.2 Prolific versus University Participants

Several independent samples t-tests and a one-way between-subjects ANCOVA were run on the combined Prolific and university data to look for differences between the two data collection sites. The ANCOVA was used to assess immediate learning, with post-test scores as the outcome variable, pre-test scores as the covariate, and collection site as the independent variable.

Significant differences were found for all dependent variables besides RIMMS composite scores; see Table 12 for a summary of the results. Therefore, the data for the Prolific participants and for the university participants were analyzed separately. Given that the sample size for the university participants ($N = 47$) was smaller than the required 128 for power of .8, only descriptive statistics are reported for this dataset (Table 19 in section *4.7.5 University Participants: Descriptive Statistics*).

Table 12 – Summary of results for analyses of data collection site, with mean directions shown for significant effects between Prolific (Pro) and University (Uni) groups.

				Prolific		University	
Dependent Variable	Analysis	p_{adjusted}	Effect Size	*M*	*SD*	*M*	*SD*
Pre-test	Independent samples *t*-test	.001* Pro > Uni	$d = -0.64$	32.54	20.10	17.38	32.62
Post-test	Independent samples *t*-test	< .001* Pro < Uni	$d = 0.87$	66.27	21.20	79.66	12.26
Immediate Learning	One-way between-subjects ANCOVA	< .001* Pro < Uni	$\eta_p^2 = 0.10$	*EMM* 65.9	*SE* 1.71	*EMM* 80.7	*SE* 2.83
Motivation	Independent samples *t*-test	.362 Pro > Uni	$d = -0.18$	3.31	1.06	3.17	0.80
Subjective Workload	Independent samples *t*-test	< .001* Pro > Uni	$d = -1.86$	9.48	3.19	5.10	2.43
PA Persona	Independent samples *t*-test	.044* Pro > Uni	$d = -0.43$	3.36	0.86	3.09	0.60

* Statistically significant

Note. Possible ranges of the variables: pre-test (0-100), post-test (0-100), immediate learning (0–100), motivation (1–5), subjective workload (1–21), PA persona (1–5).

4.7.3 Prolific Participants: Hypotheses 2.1 & 2.2

4.7.3.1 Inconsistent Participants

As described in the section *4.6.1 PA Preference*, a participant's PA preference was determined by their responses to two target questions: one that asked about their preference for the study's PA and another that asked about their preference for an alternative PA. They were marked as "prefers" or "does not prefer" based on their response to the question about the study's PA. If they answered differently for the alternative PA ("Yes" to preferring the study's PA but "No" to preferring the alternative PA, or vice versa), they were marked as "inconsistent" using an embedded Qualtrics variable.

Preliminary analyses were conducted to determine whether the inconsistent participants' data could be retained in the primary dataset. To assess this, the same tests described in the next section were run using only the consistent participants (n = 115). The results were similar to those obtained when the consistent and inconsistent (n = 11) participants were analyzed together. Therefore, all 126 participants were included in the main analyses.

4.7.3.2 Analyses

A 2×2 between-subjects ANOVA was conducted on each of the following dependent variables: motivation, subjective workload, and PA persona. Each ANOVA included PA Preference (prefers vs. does not prefer a PA) and PA Assignment (With-PA vs. No-PA) as independent variables. The outcome variables were RIMMS composite scores, NASA-TLX composite scores, and API-R composite scores, respectively.

To account for baseline differences in prior knowledge, a 2×2 between-subjects ANCOVA was used to analyze the learning dependent variable, with PA Preference and PA Assignment as independent variables, post-test scores as the outcome, and pre-test scores as a covariate. As described in the next section, two ANCOVA assumptions were marginally violated. Therefore, a multiple linear regression was conducted as a follow-up, as it allows for non-parallel slopes. However, this model exhibited a significant violation of the homoscedasticity assumption. Given these trade-offs and the lack of significant effects in the ANCOVA, the ANCOVA was retained as the primary analysis. Holm-Bonferroni correction was applied to control across multiple comparisons, and the adjusted *p*-values are provided (p_{adjusted}). Effect sizes (η_p^2) are also reported. See Table 13 on page 83 for descriptive statistics and Table 14 on page 84 for a summary of results.

4.7.3.3 Assumption Checks

Shapiro-Wilk tests and Levene's tests were conducted for each of the three ANOVAs to assess the assumptions of normality and homogeneity of variances. No concerning assumption violations were found (all $p \geq .05$).

For the learning ANCOVA, Shapiro-Wilk and Levene's tests were conducted to assess normality and homogeneity of variance, respectively. The Shapiro-Wilk test indicated a deviation from normality ($W = 0.96$, $p < .001$), but Levene's test supported the assumption of equal variances ($p = .658$). Given that ANCOVAs are generally robust to deviations from normality when variances are equal and group sizes are approximately balanced (and that the Q-Q plot did not reveal major concerns), these results suggest no meaningful violations. A scatterplot also indicated a weak but approximately linear relationship between pre-test and post-test scores. A marginal violation of the homogeneity of regression slopes was detected, with a significant interaction between pre-test scores and

PA Assignment (p = .045). Given the modest size of this effect and the absence of significant group differences, the ANCOVA was retained as the primary analysis.

For Hypotheses 2.1A–E, the primary effect of interest was the interaction between PA Preference and PA Assignment. For Hypotheses 2.2A–E, the primary effect of interest was the main effect of PA Assignment. The main effect of PA Preference is discussed in the section *4.8.4 Prolific Participants: Exploratory Analyses*.

4.7.3.4 Hypotheses 2.1A & 2.2A

Hypothesis 2.1A predicted that participants whose PA preferences were honored would exhibit higher learning than those whose preferences were not honored. This hypothesis was not supported: a 2×2 between-subjects ANCOVA found the interaction between PA Preference and PA Assignment to be nonsignificant, $F(1, 121) = 0.004$, $p = .945$ ($p_{\text{adjusted}} = 1.0$), $\eta_p^2 = 0.00004$.

Hypothesis 2.2A predicted that participants who learned with the PA would exhibit higher learning than those who learned without a PA. This hypothesis was not supported: a 2×2 between-subjects ANCOVA found the main effect of PA Assignment to be nonsignificant, $F(1, 121) = 0.58$, $p = .445$ ($p_{\text{adjusted}} = 1.0$), $\eta_p^2 = 0.005$. Pre-test scores were included as a covariate and were a marginally significant predictor of post-test performance ($p = .05$).

4.7.3.5 Hypotheses 2.1B & 2.2B

Hypothesis 2.1B predicted that participants whose PA preferences were honored would report higher motivation than those whose preferences were not honored. This hypothesis was not supported: a 2×2 between-subjects ANOVA found the interaction between PA Preference and PA Assignment to be nonsignificant on composite RIMMS scores, $F(1, 122) = 0.24$, $p = .626$ ($p_{\text{adjusted}} = 1.0$), $\eta_p^2 = 0.002$.

Hypothesis 2.2B predicted that participants who learned with the PA would report higher motivation than those who learned without a PA. This hypothesis was not supported: a 2×2 between-subjects ANOVA found the main effect of PA Assignment on composite RIMMS scores to be nonsignificant, $F(1, 121) = 0.31$, $p = .578$ ($p_{\text{adjusted}} = 1.0$), $\eta_p^2 = 0.003$.

4.7.3.6 Hypotheses 2.1C & 2.2C

Hypothesis 2.1C predicted that participants whose PA preferences were honored would report lower subjective workload than those whose preferences were not honored. This hypothesis was not supported: a 2×2 between-subjects ANOVA found the interaction between PA Preference and PA Assignment to be nonsignificant on composite NASA-TLX scores, $F(1, 122) = 0.14$, $p = .711$ ($p_{\text{adjusted}} = 1.0$), $\eta_p^2 = 0.001$.

Hypothesis 2.2C predicted that participants who learned with the PA would report lower subjective workload than those who learned without a PA. This hypothesis was not supported: a 2×2 between-subjects ANOVA found the main effect of PA Assignment on composite NASA-TLX scores to be nonsignificant, $F(1, 121) = 0.37$, $p = .544$ ($p_{\text{adjusted}} = 1.0$), $\eta_p^2 = 0.003$.

4.7.3.7 Hypotheses 2.1D & 2.2D

Hypothesis 2.1D predicted that participants whose PA preferences were honored would report higher PA persona ratings than those whose preferences were not honored.

This hypothesis was not supported: a 2×2 between-subjects ANOVA found the interaction between PA Preference and PA Assignment to be nonsignificant on composite API-R scores, $F(1, 122) = 0.49$, $p = .483$ ($p_{\text{adjusted}} = 1.0$), $\eta_p^2 = 0.004$.

Hypothesis 2.2D predicted that participants who learned with the PA would report higher PA persona ratings than those who learned without a PA. This hypothesis was not supported: a 2×2 between-subjects ANOVA found the main effect of PA Assignment on composite API-R scores to be nonsignificant, $F(1, 121) = 1.38$, $p = .242$ ($p_{\text{adjusted}} = .968$), $\eta_p^2 = 0.01$.

Table 13 – Descriptive statistics by PA Assignment and the interaction between PA Assignment and PA Preference for Hypotheses 2.1A–D and 2.2A–D analyses.

	PA Assignment				Interaction							
	With-PA		No-PA		With-PA/ Prefers PA		No-PA/ Prefers PA		With-PA/ Does Not Prefer PA		No-PA/ Does Not Prefer PA	
Dependent Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Post-Test	64.75	22.80	67.68	19.66	63.12	23.47	66.58	19.32	66.34	22.40	68.82	20.25
Motivation	3.26	1.04	3.36	1.08	3.80	0.90	3.96	0.91	2.73	0.90	2.74	0.88
Subjective Workload	9.30	3.13	9.66	3.25	9.92	2.79	10.47	3.13	8.69	3.37	8.82	3.21
PA Persona	3.27	0.78	3.44	0.92	3.70	0.66	3.94	0.76	2.87	0.67	2.93	0.79

Note. Possible ranges of the variables: post-test (0–100), motivation (1–5), subjective workload (1–21), PA persona (1–5).

Table 14 – Summary of results for Experiment 2 Hypotheses 2.1A–D and 2.2A–D.

			Main Effect	Interaction
			Lesson	PA Assignment × Lesson
Hypothesis	Analysis	Dependent Variable	p adjusted	
Hyp2.1A	2×2 between-subjects ANCOVA	Immediate learning		1.0
Hyp2.1B	2×2 mixed ANOVA	Motivation		1.0
Hyp2.1C	2×2 mixed ANOVA	Subjective workload		1.0
Hyp2.1D	2×2 mixed ANOVA	PA persona		1.0
Hyp2.2A	2×2 between-subjects ANCOVA	Immediate learning	1.0	
Hyp2.2B	One-way ANOVA	Motivation	1.0	
Hyp2.2C	One-way ANOVA	Subjective workload	1.0	
Hyp2.2D	One-way ANOVA	PA persona	.968	

p adjusted = *p*-value after Holm-Bonferroni correction

4.7.4 Prolific Participants: Exploratory Analyses

4.7.4.1 <u>Analyses</u>

<u>PA Preference.</u> ***Pre-Task Preference.*** The main effect of PA Preference was examined in the previously described 2×2 between-subjects ANCOVA and ANOVAs for learning, motivation, subjective workload, and PA persona. See Table 15 on page 88 for descriptive statistics and Table 16 on page 89 for a summary of results.

Post-Task Preference. A binary logistic regression was conducted to examine whether post-task PA preference was affected by any of the following predictor variables: pre-task preference (yes/no), strength of the pre-task preference (1-5), PA Assignment condition (With-PA vs. No-PA). The outcome variable was whether a participant's PA preference changed from pre- to post-task.

Opinion-Based Survey. Responses to the opinion-based survey were graded and converted to *z*-scores. The *z*-scores for questions on the content and design of the lesson were averaged to create a composite score reflecting overall opinions of the lesson. The *z*-scores for questions on the characteristics, voice, and general attitudes toward the PA/arrows were averaged to create a composite score reflecting overall opinions of the PA/arrows. The three questions given only to participants in the With-PA condition were excluded from the composite scores to allow for better between-group comparisons. Separate 2×2 between-subjects ANOVAs were conducted on the two composite scores, with PA Assignment and PA Preference included as the independent variables. See Table 18 on page 91 for descriptive statistics.

4.7.4.2 Assumption Checks

PA Preference. ***Pre-Task Preference.*** The assumptions of the ANCOVA and ANOVAs were discussed in the assumptions section of *4.8.3 Prolific Participants: Hypotheses 2.1 & 2.2*.

Post-Task Preference. The assumptions of binary logistic regression were assessed. Multicollinearity was not a concern, with all variance inflation factors (VIFs) near 1. The linearity of the logit assumption for the numeric predictor, preference strength, was assessed using an interaction term between preference strength and its log transformation; interaction was not significant (p = .452), indicating no violation. Outliers and leverage points were examined, and no problematic cases were found. Model fit was supported by a nonsignificant Hosmer-Lemeshow test, $\chi^2(5) = 3.93$, $p = .559$, and McFadden's pseudo-R^2 indicated modest explanatory power ($R^2 = .049$).

Opinion-Based Survey. Shapiro-Wilk and Levene's tests were conducted for the two ANOVAs. Shapiro-Wilk tests found a violation of the normality assumption for both ($p < .05$). Levene's test indicated no violation of the homogeneity of variances assumptions for the lesson opinions ANOVA ($p = .178$) but did for the PA/arrows opinions ANOVA ($p = .033$). Given that ANOVA is robust to these violations when group sizes are similar, the analysis was still used but interpreted with caution.

4.7.4.3 Findings.

PA Preference. ***Pre-Task Preference.*** After controlling for pre-test performance, a 2×2 between-subjects ANCOVA found a marginal main effect of PA Preference on post-test scores, $F(1,121) = 3.60$, $p = .06$ ($p_{\text{adjusted}} = .06$), $\eta_p^2 = 0.03$, indicating a small to moderate effect (Cohen, 1988; Table 4, p. 53). Although the raw post-test scores were similar between participants who preferred a PA ($M = 64.93$, $SD = 21.29$) and those who did not ($M = 67.60$, $SD = 21.20$), pre-test scores differed substantially: PA-preferring participants scored significantly higher ($M = 48.15$, $SD = 33.78$) than non-preferring participants ($M = 16.93$, $SD = 22.55$), $F(1, 122) = 36.65$, $p < .001$, $\eta_p^2 = 0.23$. The ANCOVA adjusted for this initial disparity, suggesting that if both groups had started at similar pre-test levels, those who did not prefer a PA might have performed better on the post-test than those who did prefer a PA, as indicated by the estimated marginal means ($EMM = 70.3$, $SE = 2.82$ vs. $EMM = 62.2$, $SE = 2.82$).

A series of 2×2 between-subjects ANOVAs revealed significant main effects of PA Preference on motivation, PA persona ratings, and subjective workload. Participants who preferred a PA reported significantly higher motivation, as reflected in higher composite RIMMS scores ($M = 3.88$, $SD = 0.90$) than those who did not prefer a PA ($M = 2.74$, $SD = 0.88$), $F(1, 122) = 51.59$, $p < .001$ ($p_{adjusted} < .001$), $\eta_p^2 = 0.30$. They also rated their learning aid more favorably in terms of PA persona, as indicated by significantly higher composite API-R scores ($M = 3.83$, $SD = 0.72$) compared to those who did not prefer a PA ($M = 2.74$, $SD = 0.73$), $F(1, 122) = 51.67$, $p < .001$ ($p_{adjusted} < .001$), $\eta_p^2 = 0.30$. Both effects were large in magnitude (Cohen, 1988). To summarize, these results suggest that PA-preferring participants responded more favorably on measures of motivation and PA persona.

This advantage of PA-preferring participants over non-preferring participants did not extend to workload. Although PA-preferring participants did report higher NASA-TLX scores ($M = 10.21$, $SD = 2.96$) than non-preferring participants ($M = 8.76$, $SD = 3.26$), this reflects a negative outcome, as higher scores indicate greater subjective workload. This difference was statistically significant, $F(1, 122) = 6.78$, $p = .01$ ($p_{adjusted} = .021$), $\eta_p^2 = 0.05$, representing a small to moderate effect.

Follow-up multiple regression models were constructed to determine whether the observed effects were driven by PA preference specifically or by broader preferences for educational technology. Each model included both PA preference and general educational technology preference (measured by responses to the filler items on the Educational Technology Preferences Survey) as predictors. The dependent variables were learning, motivation, subjective workload, and PA persona ratings. Across all models, general educational technology preference was not a significant predictor ($p = .688$ for learning,

.102 for motivation, .978 for workload, and .200 for PA persona). In contrast, PA preference significantly predicted motivation ($p < .001$) and PA persona ratings ($p < .001$) and showed marginal significance for learning ($p = .074$) and subjective workload ($p = .065$), consistent with the findings from the previous ANCOVA and ANOVAs.

Table 15 – Descriptive statistics by PA Preference for Experiment 2's exploratory analyses on pre-task PA preference.

	PA Preference			
	Prefers PA		Does Not Prefer PA	
Dependent Variable	*M*	*SD*	*M*	*SD*
Pre-Test	48.15	33.78	16.93	22.55
Post-Test	64.93	21.29	67.60	21.20
Motivation	3.88	0.90	2.74	0.88
Subjective Workload	10.21	2.96	8.76	3.26
PA Persona	3.82	0.72	2.90	0.73

Note. Immediate learning ANCOVA included pre-test scores as a covariate and post-test scores as the outcome variable.
Possible ranges of the variables: pre-test (0-100), post-test (0–100), motivation (1–5), subjective workload (1–21), PA persona (1–5).

Table 16 – Summary of results for exploratory analyses of PA Preference, with mean directions shown for significant effects between Prefers (P) and Does Not Prefer (DNP) conditions.

		Main Effect
		PA Preference
Analysis	Dependent Variable	p adjusted
2×2 between-subjects ANCOVA	Immediate learning	.06* *EMM*s: P < DNP Pre-test: P > DNP Post-test: P ≈ DNP
2×2 between-subjects ANOVA	Motivation	< .001*** P > DNP
2×2 between-subjects ANOVA	PA persona	< .001*** P > DNP
2×2 mixed ANOVA	Subjective workload	.021** P > DNP

p adjusted = *p* -value after Holm-Bonferroni correction
*Marginally significant at $p < .10$
**Statistically significant at $p < .05$
***Statistically significant at $p < .001$

Post-Task Preference. A binary logistic regression was conducted to examine if PA Assignment and initial PA preference strength predicted whether participants changed their preference after completing the learning task. The inclusion of these predictors significantly improved model fit, $\chi^2(2) = 5.99$, $p = .0499$, indicating that they reliably distinguished between participants who did and did not change their preference.

The effect of PA Assignment was not significant, $\beta = –0.77$, $SE = 0.48$, $z = –1.62$, $p = .105$, $OR = 0.46$, meaning that being assigned to the With-PA versus No-PA condition did not significantly influence preference change. Preference strength showed a trend toward significance, $\beta = –0.47$, $SE = 0.25$, $z = –1.88$, $p = .061$, $OR = 0.63$. This indicates that for each one-point increase in strength of pre-task preference, the odds of changing one's preference decreased by approximately 37%. In other words, having a stronger initial PA preference was associated with a lower likelihood of changing PA preference from pre-task to post-task. Pre-task and post-task preference counts are shown in Table 17.

Table 17 – Pre-task and post-task PA preference frequency counts for Prolific participants.

	Post-Task PA Preference		
Pre-Task PA Preference	Prefer PA	Do Not Prefer PA	*Total*
Prefer PA	59	4*	*63*
Do Not Prefer PA	20*	43	*63*
Total	*79*	*47*	

*Participants who changed their PA preference from pre- to post-task.

Opinion-Based Survey. A 2×2 between-subjects ANOVA on participants' opinions of the lesson (composite *z*-scores) did not find a significant main effect of PA Preference, $F(1, 122) = 1.48$, $p = .227$, $\eta_p^2 = 0.01$, or of PA Assignment, $F(1, 122) = 0.09$, $p = .768$, $\eta_p^2 = 0.0007$. The interaction effect was also insignificant, $F(1, 122) = .261$, $p = .61$, $\eta_p^2 = 0.002$.

The 2×2 between-subjects ANOVA on participants' opinions of the PA/arrows (composite *z*-scores) found the interaction effect to be nonsignificant, $F(1, 122) = 2.07$, $p = .153$, $\eta_p^2 = 0.02$. However, both main effects were significant: PA Preference, $F(1, 122) = 63.51$, $p < .001$, $\eta_p^2 = 0.34$, and PA Assignment, $F(1, 122) = 7.04$, $p = .009$, $\eta_p^2 = 0.05$. The effect size was large for PA Preference and small to moderate for PA Assignment

(Cohen, 1988). Participants who preferred a PA provided higher opinions of the PA or arrows (descriptive statistics are for standardized scores; $M_z = 0.34$, $SD_z = 0.44$) than those who did not prefer a PA ($M_z = -0.34$, $SD_z = 0.55$). Participants in the With-PA condition rated the PA less favorably ($M_z = -0.12$, $SD_z = 0.69$) than participants in the No-PA condition rated the arrows ($M_z = 0.12$, $SD_z = 0.48$).

Table 18 – Descriptive statistics for *z*-scores by PA Assignment, PA Preference, and their interaction for Experiment 2's exploratory analyses on the opinion-based survey.

	PA Assignment				PA Preference				Interaction							
	With-PA		No-PA		Prefers PA (P)		Does Not Prefer PA (DNP)		With-PA/P		No-PA/P		With-PA/ DNP		No-PA/DNP	
Dependent Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Lesson Opinions	−0.01	0.48	0.01	0.42	0.05	0.44	−0.05	0.46	0.01	0.51	0.08	0.38	−0.04	0.47	−0.06	0.46
PA/Arrows Opinions	−0.12	0.69	0.12	0.48	0.34	0.44	−0.34	0.55	0.29	0.50	0.39	0.38	−0.52	0.62	−0.17	0.41

Note. Descriptive statistics are for standardized *z*-scores. Observed *z*-score ranges: lesson opinions (−1.08 to 0.85) and PA/arrows opinions (−2.18 to 1.13).

4.7.5 University Participants: Descriptive Statistics

Table 19 shows the descriptive statistics (means, standard deviations, and group sample sizes) for all dependent variables among university participants, organized by PA Assignment, PA Preference, and their interaction.

Table 19 – Descriptive statistics by PA Assignment, PA Preference, and their interaction for university participants in Experiment 2.

	PA Assignment				PA Preference				Interaction							
	With-PA *n* = 23		No-PA *n* = 24		Prefers PA (P) *n* = 16		Does Not Prefer PA (DNP) *n* = 31		With-PA/P *n* = 8		No-PA/P *n* = 8		With-PA/ *n* = 15		No-PA/DNP *n* = 16	
Dependent Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Pre-Test	31.83	33.11	33.21	32.39	23.44	26.91	14.25	15.09	46.94	34.36	49.24	33.75	17.20	24.53	16.67	20.85
Post-Test	82.20	11.17	77.23	13.00	82.12	11.81	78.39	12.49	88.37	6.87	75.87	12.75	78.91	11.80	77.91	13.48
Motivation	3.16	0.78	3.19	0.83	3.65	0.61	2.93	0.78	3.59	0.45	3.70	0.76	2.92	0.83	2.94	0.76
Subjective Workload	4.96	2.16	5.23	2.70	5.64	2.08	4.82	2.58	4.94	1.85	6.33	2.18	4.97	2.37	4.68	2.83
PA Persona	3.08	0.61	3.10	0.60	3.53	0.26	2.86	0.60	3.56	0.27	3.51	0.27	2.83	0.60	2.89	0.62
Lesson Opinions	0.08	0.43	−0.07	0.36	0.20	0.40	−0.10	0.36	0.33	0.29	0.08	0.46	−0.06	0.44	−0.15	0.28
PA/Arrows Opinions	−0.19	0.61	0.18	0.35	0.26	0.17	−0.14	0.59	0.24	0.19	0.28	0.17	−0.41	0.64	0.13	0.40

Note. Possible ranges of the variables: pre-test (0-100), post-test (0–100), motivation (1–5), subjective workload (1–21), PA persona (1–5).
Descriptive statistics for Lesson Opinions and PA/Arrows Opnions are for standardized z-scores. Observed *z*-score ranges: lesson opinions (−0.95 to 0.79) and PA/arrows opinions (−1.62 to 0.96).

Although statistical analyses could not be conducted on these data due to the small sample size and unbalanced groups, speculations can be made based on the descriptive statistics. First, there was a substantial difference in pre-test scores between the PA-preferring ($M = 23.44$, $SD = 26.91$) and non-preferring ($M = 14.31$, $SD = 15.09$) participants. However, this difference did not remain in the post-test ($M = 82.12$, $SD = 11.81$ vs. $M = 78.42$, $SD = 12.49$), indicating that the initial discrepancy in prior knowledge did not impact post-test performance. The difference in pre-test performance is the same as what was found in Prolific participants, where PA-preferring participants scored significantly higher on the pre-test than non-preferring participants. No large differences were observed between any groups for motivation, subjective workload, or PA persona ratings.

For opinions on the lesson, participants who preferred a PA ($M_z = 0.21$, $SD_z = 0.40$) rated the lesson more favorably than those who did not prefer the PA ($M_z = -0.11$, $SD_z = 0.36$). The same can be seen for opinions on the PA/arrows ($M_z = 0.26$, $SD_z = 0.17$ vs. $M_z = -0.14$, $SD_z = 0.59$). This is similar to the Prolific participants, where participants who preferred a PA provided higher opinions of the PA/arrows.

Participants in the With-PA condition ($M_z = 0.14$, $SD_z = 0.43$) provided more favorable ratings of the lesson than those in the No-PA condition ($M_z = -0.04$, $SD_z = 0.36$), although this difference was not very large. Additionally, participants in the With-PA condition rated the PA less favorably ($M_z = -0.19$, $SD_z = 0.61$) than participants in the No-PA condition rated the arrows ($M_z = 0.18$, $SD_z = 0.35$). Again, this is analogous to the findings from the Prolific participants, where participants who were assigned a PA provided lower opinions of their learning aid than those assigned arrows.

A difference between the university and Prolific participants was found in the proportion that preferred versus did not prefer the PA. In the Prolific participants (N = 248 before exclusions), around 71% (n = 176) indicated preference for the PA. In contrast, only ~34% (n = 16) of university participants (N = 47) preferred the PA. Additionally, while roughly 39% (n = 24) of Prolific participants changed their PA preference from pre- to post-task, only 19% (n = 15) of university participants did. However, it is impossible to know whether these proportions would remain the same with a larger sample of university participants. The counts for pre- and post-task PA preference in university participants can be seen in Table 20.

Table 20 – Pre-task and post-task PA preference frequency counts for university participants.

	Post-Task PA Preference		
Pre-Task PA Preference	Prefer PA	Do Not Prefer PA	*Total*
Prefer PA	10	4*	*14*
Do Not Prefer PA	11*	13	*24*
Total	*21*	*17*	

*Participants who changed their PA preference from pre- to post-task.
Note. Post-task preference question was not implemented until partway through data collection, so table only shows data for 38 participants.

4.8 Discussion

Hypotheses 2.1A–D predicted that honoring participants' PA preferences would lead to better outcomes in learning, motivation, subjective workload, and PA persona ratings, compared to when their preferences were not honored. However, these hypotheses were not supported, as no significant interactions were found between PA Preference and PA Assignment. To confirm this, follow-up independent samples *t*-tests comparing participants in congruent versus incongruent preference-assignment conditions were conducted, and none were significant.

This null interaction effect contradicts previous studies that reported benefits of preference matching in non-PA learning environments (Cinkara & Bagceci, 2013; Freitag & Sullivan, 1995). It also contradicts the opinion of numerous PA researchers that catering to individual differences should have a positive effect on PA-enhanced learning outcomes (Gulz, 2004; Dincer & Doganay, 2017; Ozogul et al., 2013). Instead, it is more consistent with Andre and Wickens' (1995) conclusion that user preference and performance are often disassociated. The authors described several studies in which preference for an interface did not correspond with performance (Blanchard et al., 1993; Jeffrey & Beck, 1972). One explanation for why this occurred in the present study is that preference might have reflected participants' general attitudes or affective responses to the PA, rather than being based on how they felt about its instructional efficacy. Although the preference question asked participants whether they would "want to learn" with the PA, this does not necessarily mean they evaluated their preference from a learning-oriented perspective. Thus, a validated instrument for PA preference and an understanding of the constructs it assesses would be advantageous for this line of research.

Another explanation for the null interaction effect, as well as for the lack of a main effect of PA Assignment, is the high degree of similarity between the With-PA and No-PA conditions. Both presented the same educational content, used the same voice, and provided equivalent deictic gestures. The only meaningful difference between the two was the PA's additional humanlike features, such as its human appearance, eye gaze, facial expressions, and non-deictic gestures. Consider Social Agency Theory (see section *1.1.1.1 Social Agency Theory*; Mayer et al., 2003b), which posits that social cues trigger social conversation schemas, fostering deeper cognitive processing and more meaningful learning. Perhaps the social cues presented by the arrows (voice and deictic gesture) were sufficient to elicit the social schema, meaning the additional cues in the With-PA condition did not provide further benefit. This interpretation contradicts the embodiment principle (Mayer & DaPra, 2012), which would predict enhanced learning from the additional humanlike social cues in the With-PA condition. However, it is consistent with the Media Equation (Reeves & Nass, 1996), which argues that people often respond socially to media, even when it lacks traditional human characteristics. These findings suggest that the basic social cues provided by the arrows might have sufficiently satisfied participants' social expectations, limiting the potential impact of the additional cues in the PA condition. Further, the results align with prior work by Choi and Clark (2006), who similarly found no significant differences in learning or mental effort between an arrows condition and a gesturing PA condition.

An alternative explanation for the null interaction effect and PA Assignment main effect is that participants interpreted the arrows not as inherently social, but as evidence of human involvement in the lesson design. This recognition of intentional design might have prompted a subtle social response due to the awareness of a human presence behind the instruction. Both of these explanations might also account for PA-preferring participants reporting higher motivation and PA persona ratings regardless of PA Assignment; even those in the No-PA condition might have perceived the arrows as sufficiently satisfying their desire for social qualities in a learning aid, whether from the arrows' social cues or from their perception of a human designer.

The main effect of PA Preference can be interpreted independently as well. In addition to reporting higher motivation and PA persona ratings, PA-preferring participants exhibited significantly higher pre-test scores and greater subjective workload than non-preferring participants. Although random sampling was used for participant selection, one possibility is that PA-preferring participants entered the study with greater prior knowledge or stronger baseline academic skills, contributing to their higher pre-test scores. However, another possibility is that these participants invested more effort throughout the study, as suggested by their higher perceived workload and motivation. If so, PA preference, as measured in the present study, might reflect an underlying personality trait that fosters interest or engagement. For example, individuals high in conscientiousness are described as responsible, dependable, and precise, with this trait facilitating "task- and goal-directed behavior" (John et al., 2008, p. 120). Those high in openness are described as intelligent, insightful, and curious, often "tak[ing] the time to learn something simply for the joy of learning" (p. 120). Either of these traits, traits from other personality paradigms, or other cognitive factors could contribute to a greater willingness to participate in the study or engage with the instructional material. Future research incorporating a battery of personality assessments is needed to evaluate this interpretation.

The results of the opinion-based survey could support either argument: (1) that the arrows were sufficient to produce a social effect, or (2) that PA preference reflects an underlying trait. On the survey, PA-preferring participants provided more favorable opinions of their assigned learning aid than non-preferring participants, regardless of PA Assignment condition. This could indicate either (1) that both learning aids were perceived similarly in terms of their social qualities, leading PA-preferring participants to rate them both highly, or (2) that PA-preferring participants were more engaged in the study, leading to more favorable responses. However, the absence of similar elevated ratings for the lesson content and design among PA-preferring participants lends credence to the first explanation. If the second were true, higher ratings would likely extend across all aspects of the study, not just the learning aid.

Lastly, there were notable differences between the Prolific and university samples. Approximately 71% of Prolific participants preferred learning with a PA, compared to only 34% of university participants. This could be a reflection of differences in the populations from which the samples were drawn. The university participants might have had more prior experience with online education, potentially shaping their opinions on educational technology in different ways than Prolific participants. However, university recruitment was incomplete, and it is possible that the observed differences would shrink with more balanced sample sizes.

Another explanation for this discrepancy is response bias among Prolific participants. Because their study submissions can be rejected, potentially leading to their Prolific account being banned and them losing income, these participants might have aimed to provide responses they believed the researcher wanted. As such, it is possible that Prolific participants erroneously assumed that the researcher wanted them to prefer the PA, leading to an influx of supposed PA-preferring participants. This could also explain Prolific participants' higher scores on the pre-test, PA persona ratings, and subjective workload compared to university participants. In contrast, university participants outperformed Prolific participants on the post-test, possibly due to greater familiarity with the math content, as the university sample included a higher proportion of STEM majors (60% vs. 37%).

In conclusion, although Experiment 2 did not support the hypothesis that honoring PA preferences would improve learning outcomes, the results offer several meaningful contributions. They challenge beliefs that preference matching enhances instructional effectiveness, particularly in situations where the functional differences between conditions are limited. The results also raise questions about how learners interpret and respond to social cues in educational environments, suggesting that even simple learning aids like arrows might be sufficient to activate social schema (Mayer et al., 2003b). Further, the main effect of PA Preference highlights the potential influence of individual learner characteristics, which might impact both subjective and objective outcomes regardless of the instructional manipulation. These findings underscore the complexity of studying and designing digital learning environments and emphasize the need for future work on validating preference measures, identifying which cues are optimal for eliciting a sense of social presence, and exploring how personality traits shape performance in online learning. Further, research could examine differences in participant responses to simple versus complex learning aids to determine whether simple aids are perceived as having social qualities or merely signal the involvement of a human designer in the lesson's creation.

CHAPTER 5. CONCLUSION

5.1 Summary & General Discussion

Pedagogical agents (PAs) are virtual characters implemented in online learning environments, utilized for their ability to enhance users' learning, motivation, and cognitive load. Despite their recognized benefits, some learners exhibit a dislike towards PAs (Hook et al., 2000), and many online platforms lack the option to disable them (Graesser et al., 2005; Hone et al., 2005; Krämer et al., 2007). This raises concerns about the potential adverse impact on the educational outcomes of these learners. Furthermore, it has been suggested that the effectiveness of PAs varies across different content domains, such as mathematics and humanities (Schroeder et al., 2013), a phenomenon that has not been thoroughly explored in experimental studies.

The present research examined (1) the effectiveness of a PA designed by the author, as compared to a No-PA condition that instead included pointing arrows, (2) the impacts of a PA in a math lesson compared to an art lesson, and (3) the effects of congruence versus incongruence between a participant's preference for a PA and their assignment to a PA. None of the study's hypotheses were supported, calling into question both the added value of PAs over simpler instructional aids and the extent to which learner preference meaningfully influences educational outcomes.

Consistent across both experiments, participants who learned with the PA and participants who learned with the arrows performed similarly on learning, motivation, subjective workload, and PA persona ratings. These null results challenge the common assumption that the presence of a PA in the learning environment enhances outcomes (Atkinson, 2002; Dincer & Doganay, 2017; Kizilkaya & Askar, 2008; Moreno et al., 2001; Moreno et al., 2010; Paas & van Merrienboer, 1993; Schroeder et al., 2013; van Mulken et al., 1998). They also contradict the embodiment principle (Mayer & DaPra, 2012), which would predict that the additional humanlike features provided by the PA should result in increased learning. Instead, the findings align more closely with Social Agency Theory (Mayer et al., 2003b), as it is possible that the minimal cues provided by the arrows (voice and deictic gestures) were sufficient to elicit the social conversation schema described by the theory, making the PA's additional cues ineffective at producing further benefits. Another possibility is that the design of the No-PA condition signaled human involvement in the lesson's creation, producing a subtle social effect that the With-PA condition did not enhance further. Both interpretations point to the need to identify which cues most effectively foster social engagement in learning contexts.

While many of the present study's hypotheses were not supported, the results do not support the "agents-as-distractors" viewpoint held by some opponents of PAs (Frechette & Moreno, 2010, p. 64; Clark & Choi, 2007), given that participants in the No-PA condition did not perform better than those in the With-PA condition. Instead, the results of this study support a neutral position, consistent with previous research reporting no significant differences between With-PA and No-PA conditions across various learning outcomes (Davis, 2018; Frechette & Moreno, 2010; Heidig & Clarebout, 2011; Moreno et al., 2001; Moundridou & Virvou, 2002; Schroeder et al., 2017; van Mulken et al., 1998; Yusoff et al., 2023). Further, the present study lends support to the idea that pointing arrows are equally as beneficial as a gesturing PA for facilitating learning (Choi & Clark, 2006). In fact, responses to the opinion-based survey in Experiment 2 suggest that arrows could result in more favorable subjective outcomes; participants assigned to the arrow condition reported significantly more positive opinions of their learning aid than those assigned to the PA condition. The results of these experiments could also be interpreted as indicating that the PA and arrows were equally *ineffective*, meaning neither offer advantages over no learning aid.

The other two hypotheses of this study were that (1) in the With-PA condition, better outcomes would be found for the math lesson than for the art lesson (Experiment 1) and (2) participants whose PA preferences were honored would have better outcomes than those whose preferences were not honored (Experiment 2). The fact that neither of these hypotheses were supported could be rooted in the absence of a PA Assignment effect. In other words, the lack of differences between the With-PA and No-PA conditions could have eliminated the ability for these other effects to appear. The present study's null effects stand in contrast to prior findings supporting domain-specific PA effects (Schroeder et al., 2013) and the benefits of preference matching (Cinkara & Bagceci, 2013; Freitag & Sullivan, 1995).

The only meaningful significant main effect in this study's two experiments was PA Preference in Experiment 2. The results showed that participants who preferred a PA had higher pre-test scores, motivation, PA persona ratings, subjective workload, and ratings of the learning aid, regardless of whether they learned with the PA or the arrows. One explanation for these results is that preference for a PA is indicative of some personality trait or cognitive factor which led PA-preferring participants to learn or participate more intensely.

Overall, the findings from this study contribute to the ongoing discussion regarding the value of PAs by challenging key assumptions in the literature. While past research has emphasized the instructional advantages of PAs, the current results suggest that these benefits might not be robust or generalizable. The present study directly compared a gesturing, facially expressive PA to a non-anthropomorphic pointing arrow across two content domains while also examining the role of preference matching. By doing so, it provided a systematic test of several conditions in which PAs might be effective. The null effects observed across most outcomes emphasize the need for a more thorough investigation of PA design and the creation and validation of measures assessing individual preference in digital learning environments. Ultimately, these findings highlight the importance of using controlled experimental designs to disentangle when, how, and for whom PAs offer instructional value, and when simpler alternatives might be just as effective.

5.2 Limitations & Future Research

A few limitations should be considered when interpreting the findings of this study, many of which offer clear directions for future research. One limitation was the high level of similarity between the With-PA and No-PA conditions. Future research should include a true control condition with no learning aid, allowing for comparisons across all three groups. This would clarify whether both the PA and arrows improve learning relative to no aid or whether neither offers a significant advantage over a control condition.

As discussed throughout this paper, the effectiveness of a PA depends on many factors, one of which is the content of the lesson in which it is used (Schroeder et al., 2013).

Thus, another limitation is that the math lesson used in the present study might not have effectively promoted differences between the With- and No-PA conditions. Further, visual and auditory aspects of the PA, such as its appearance and voice, can influence its pedagogical benefits (Heidig & Clarebout, 2011; Mayer, 2014), meaning it is possible that the design of the PA in the present study was not conducive to enhancing outcomes relative to the arrows. Future research should establish the instructional effectiveness of a PA for a specific lesson through pilot testing before evaluating the impact of PA preference.

As highlighted by the issues described in the previous paragraph, a broader challenge in PA research is the difficulty of producing generalizable findings, given the vast range of possible designs and instructional contexts. Minor changes in appearance, voice, gesture style, or other behaviors can shift learner perceptions and outcomes, yet these factors are often varied without strong theoretical or empirical justification. A literature-informed, rationally guided review could help identify which design features most consistently effect learning and related outcomes, allowing researchers to target the most promising variables rather than exploring design elements in an unguided manner. Such a review would also provide support for systematic comparisons of alternative PA and lesson designs, strengthening the theoretical foundation of PA research and increasing the likelihood that findings are replicable and applicable across contexts.

Another limitation of the present study is that the preference measure used in Experiment 2 might have lacked validity. Preference was assessed with a single yes or no question in the Educational Technology Preferences Survey (discussed in section *4.5.1 PA Preference*). In contrast, Freitag and Sullivan (1995), who found a preference matching effect in a non-PA instructional context, used a 10-item Likert-style questionnaire to measure preference. The need for a more comprehensive preference measure is also supported by Andre and Wickens (1995), who noted that to accurately compare preference and performance, researchers should "match the dimension(s) on which user preference ratings are based with the dimension(s) on which performance measures are based" (p. 12). While preference was measured simply in the current study, performance was measured across multiple constructs and subconstructs. Andre and Wickens further note that preference itself has many dimensions, including "performance, usability, familiarity, excitement, comfort, and effort" (p. 12). Thus, another avenue for further research would be the development and validation of a multi-dimensional PA Preference instrument aligned with the relevant performance measures. This would be beneficial for future investigations of PA preference and could also be adapted to assess preferences for other types of educational technology.

As previously discussed, PA preference as measured in Experiment 2 might have been indicative of some underlying personality trait. Future research should utilize additional surveys to investigate this. If underlying traits are identified, investigations examining whether they also predict enhanced outcomes with other types of educational technologies could follow. Such findings would help inform the design of learning or training experiences tailored to individuals with specific traits, tendencies, or dispositions.

Also in Experiment 2, there was a higher potential for response bias in Prolific participants because they were likely acting to avoid a submission rejection. Data collection is ongoing with university participants, who do not face the same risk. If the effect of PA Preference is not also observed in this sample, the bias explanation will be supported.

However, if the effect is still observed, it is more likely caused by an underlying personality trait associated with PA preference.

Finally, although no domain or preference matching effects were found in the present study, the author emphasizes the importance of continuing this line of research given the support for these hypotheses found in the literature (Cinkara & Bagceci, 2013; Freitag & Sullivan, 1995; Schroeder et al., 2013). The modifications outlined in the current section should allow for a more controlled and informative investigation of these phenomena.

5.3 Final Remarks

This study highlights several important considerations regarding the use of PAs in digital learning environments. The present PA did not facilitate enhanced performance compared to the simpler instructional aid of pointing arrows on any outcomes, raising questions about the added value of humanlike PAs. Furthermore, learner preference for a PA did not interact with PA assignment, contradicting claims that personalization or preference matching improves educational outcomes in PA-enhanced environments. The only consistent effects observed were for PA preference itself, where participants who preferred a PA produced better outcomes for several variables, regardless of which learning aid they received. Together, these findings suggest that the instructional benefits often attributed to PAs might be overstated and that more attention should be given to identifying the specific conditions under which they facilitate learning.

The broader goal of this research is to better understand when and for whom educational technologies, particularly PAs, offer the greatest instructional value. Achieving this goal would be supported by a high-level, literature-guided review to identify the factors that most strongly influence the pedagogical benefits of PAs, as described in the earlier *5.2 Limitations and Future Research* section. Additionally, examinations into how educational technologies interact with learner characteristics and lesson content should be conducted to optimize learning outcomes across all individuals. As highlighted by the present study, preferences for educational technologies vary considerably, with approximately one-third to one-half of participants across all samples reporting that they did not want to learn with a PA. As instruction in all subject areas becomes increasingly digital, it is essential to determine not only whether complex tools like PAs enhance learning, but also when simpler alternatives might be similarly effective.

APPENDIX A. EXCERPT FROM BERNOULLI TRIALS LESSON TRANSCRIPT

In this lesson, I'm going to tell you about a statistical concept called a Bernoulli trial. I'm excited to help you as you take this step towards deepening your mathematical knowledge!

Now, you might be wondering, what exactly is a Bernoulli trial? Well, have you ever flipped a coin? If so, you've already performed a Bernoulli trial without even knowing it! A Bernoulli trial is simply an experiment or process that results in one of two outcomes: a success or a failure. In a Bernoulli trial, a "success" is simply our outcome of interest.

Let's consider the example of flipping a coin. Suppose we want the coin flip to result in heads. In this scenario, we would define heads as a "success" and "tails" as a failure. Imagine you flipped the coin 100 times. In the context of Bernoulli trials, the most important information wouldn't be which of those flips ended in heads, meaning in "success." Instead, the important thing to know is *how many* of the flips ended in heads – meaning, how many "successes" you ended up with.

Let's think about our coin flip example in the context of a mathematical formula. Suppose we flip a coin **n** times. We can use **x** to denote the number of heads, our successes, that we get in **n** coin flips. **p** will represent the probability that we get a heads on any particular coin flip. Then, this equation (show equation) will tell us the probability of getting **x** heads in **n** coin flips:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Don't worry, you don't need to memorize this formula for the test. It'll be given to you when you need it.

Now, let's say that exact thing same in a strictly mathematical context for Bernoulli trials. Instead of talking about "heads", we will talk about "successes", and instead of talking about "coin flips", we will talk about "Bernoulli trials."

The equation for a set of Bernoulli trials, which we just looked at, is used to calculate the probability of getting a certain number of successes over the course of a certain number of trials.

We will let the variable **x** denote the number of successes in **n** Bernoulli trials, where **p** is the probability of success on any particular trial. Therefore, our equation (show equation) tells us the probability of getting **x** successes in **n** trials.

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

The first part of this equation, $\binom{n}{x}$ said as "n choose x", is a statistical procedure called a **combination**. Like we just discussed, in the context of a Bernoulli trial, **n** is the number of trials, and **x** is the number of trials that are successes. You won't need to calculate the value of any combinations during this study, but it's important to know what **n** and **x** represent and what numbers to substitute into the equation for them.

Lastly, the variable **p** in our equation refers to the probability of a particular trial being a success. Depending on the problem, you'll either be given the value or **p**, or you'll have to calculate it.

You've done a great job at learning the basics of Bernoulli trials! Now, let's look at an example!

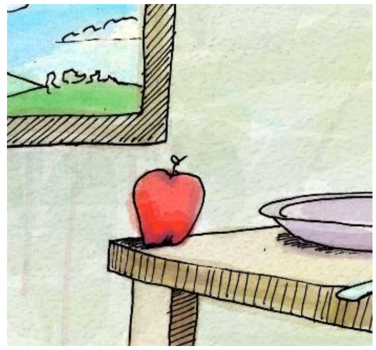
APPENDIX B. EXCERPT FROM PRINCIPLES OF DESIGN LESSON TRANSCRIPT

In this lesson, I’m going to tell you about an artistic concept called the principles of design. I’m excited to help you as you take this step towards deepening your understanding of art!

These principles are key guidelines that many artists use to create visually striking and balanced work. They also help them with justifying certain design decisions. By the end of our session, you'll have a deeper understanding of what goes into making art, a deeper appreciation for the effort artists put into their work, and you'll be equipped with skills to create your own artistic pieces. We will be discussing five principles of design: emphasis, balance, contrast, proportion, and unity. We'll explore each of these in detail, enhancing your art knowledge.

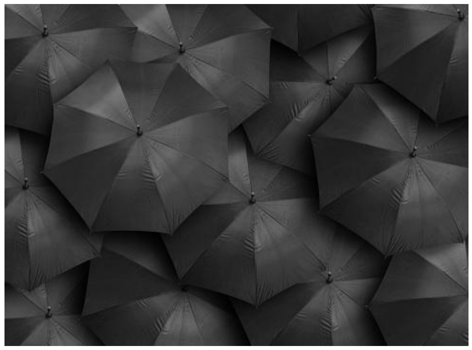
Let's start with the first principle: emphasis. Think of emphasis as the spotlight in an artwork. It's all about drawing your attention to the specific part of the piece that the artist really wants you to focus on.

Artists can emphasize a part of their artwork in several ways. One common method is through placement. For example, in this drawing, what was your attention drawn to first? … It was probably the apple, right? This is because it's positioned in the center of the artwork, making it the focal point.



There are many other ways besides placement that artists can emphasize parts of their work. Let's talk about size. Take a look at this painting. Notice how your eyes are immediately drawn to the tree? That's not a coincidence, it's a thoughtful decision by the artist to emphasize the tree by making it the largest object in the artwork.

Can you think of any other ways that emphasis can be created in art? … If you thought of color, you're correct! Color is a common and powerful way that artists add emphasis to their work. For example, in this photo, do you notice how all of the umbrellas blend together. There isn't a particular one that stands out, is there? But now, let's use color to create emphasis. Wow! The blue umbrella really stands out now, doesn't it? By adding color to just one umbrella, we've created emphasis in this picture.



APPENDIX C. ELEMENT INTERACTIVITY EXAMINATION:

MATH VERSUS ART LESSON

Imagine you have a bag of 8 marbles. If 6 of the marbles are blue and 2 of the marbles are white, then the probability of randomly picking a white marble is 2/8, or 0.25. Now, suppose you want to pick blue marbles out of the bag instead. You pick 4 marbles out of the bag. What is the probability of picking 0 blue marbles, picking 1 blue marble, and picking 2 blue marbles? Assume that you are sampling with replacement. Let's talk about how to approach this problem.

First, let's discuss what the problem means when it says you're "sampling with replacement." In our example, this means that you put each marble back in the bag after you pick it out and before you pick the next one out of the bag. What do you think this means about the probability of picking a certain marble? It means that the probability of getting a certain color of marble does not change depending on how many marbles you're picking out of the bag. In other words, the probabilities associated with each color of marble remain constant throughout your calculations. Now that we've covered sampling with replacement, let's dive into how to solve this problem.

As usual, we'll start by determining the values of our variables. Let's talk about n and p first and hold off on talking about x for a little bit.

The problem tells us that you're picking 4 marbles out of the bag, so n equals 4. It also tells us that the probability of picking a white marble out of the bag is 0.25. But, take another look at the problem. We're interested in the probability of picking a blue marble, not a white marble. How can we find the probability of picking a blue marble using the information provided to us? There're two important things to keep in mind as we figure this out. First, we know that in our bag, we only have two colors of marbles: white and blue. There's nothing else in there, so every marble draw will result in either a white or a blue marble. The second bit of information is a statistical rule: All probabilities must add up to 100%, which is 1.0 as a decimal. This rule means that if we know the probability of one event – in this example, picking a white marble – we can easily find the probability of the opposite event – picking a blue marble – by subtracting from the total probability. Now, let's put this information into action. Knowing that the probability of drawing a white marble is 0.25, we can use this rule to find the probability of the other event – drawing a blue marble. We calculate this by subtracting the white marble's probability, 0.25, from the total probability, 1.0. This subtraction, 1.0 minus 0.25, leaves us with 0.75. So, this means that p, the probability of picking a blue marble, is 0.75. Great job figuring this out!

Now that we know the values of n and p, let's move on to discussing x. The problem asks us to find the probability values associated with three different numbers of successes – 0, 1, and 2 – where success is equal to picking a blue marble. Since we're finding the probabilities associated with three numbers of successes, we're going to have three different x values. What do you think these x values are? If you said, 0, 1, and 2, you're correct! Great job! Since x = 0, 1, and 2, we will need to solve the Bernoulli trials equation three times: once where x=0, once where x=1, and once where x=2. So, now that we know the values of x, n, and p, can you guess what the next step is based on our previous examples? That's right! Next, we plug our values into the Bernoulli trial equation. Let's start with x = 0. Remember that n = 4 and p = 0.75.

$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ Here's our Bernoulli trial equation.

$P(X = 0) = \binom{4}{0} 0.75^0 (1-0.75)^{4-0}$ Since we're start with x = 0, let's plug in 0 for x. We'll also substitute 4 for n and 0.75 for p.

$P(X = 0) = 1 \times 1 \times (0.25)^4$ Now, since the procedure from this point forward is identical to our previous two examples, we don't have to go through each step again.

$P(X = 0) = 1 \times 1 \times 0.00391$

$P(X = 0) = 0.00391$ Let's just look at the answer: 0.00391.

This means that the probability of selecting 0 blue marbles from the bag of 4 marbles is 0.00391 or 0.391%. In our previous examples, this is where we stopped. However, for this example, we need to keep going so that we can determine the probabilities associated with x = 1 and x = 2. Let's move on to x=1. Remember that n=4, and p=0.75.

$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

$P(X = 1) = \binom{4}{1} 0.75^1 (1-0.75)^{4-1}$ Again, let's substitute 4 for n and 0.75 for p. This time, we'll plug in 1 with x. Notice that this means we replace any instance of 0 in the previous equation with 1 in this equation because **x** now equals 1.

$P(X = 1) = 4 \times 0.75 \times (0.25)^3$ Just like last time, the rest of the procedure is identical to our previous examples, so let's skip to the answer.

$P(X = 1) = 4 \times 0.75 \times 0.015625$

$P(X = 1) = 0.04688$ The answer this time, when x = 1, is 0.04688.

So, this means that the probability of selecting 1 blue marble from the bag of 4 marbles is 0.04688 or 4.688%. Our last step would be to solve for x=2.

Figure C1 – Longest segment of math lesson, analyzed for level of element interactivity (see Figure C2).

$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$	This is the equation with no numbers substituted.
$P(X = 2) = \binom{4}{2} 0.75^2 (1-0.75)^{4-2}$	Let's see what the equation looks like when we plug in 4 for n, 0.75 for p, and 2 for x.
$P(X = 2) = 6 \times 0.5625 \times (0.25)^2$	
$P(X = 2) = 6 \times 0.5625 \times 0.0625$	
$P(X = 2) = 0.21094$	Skipping to the answer, when x =2, the answer is 0.21094.

Therefore, the probability of selecting 2 blue marbles from the bag of 4 marbles is 0.21094 or 21.094%. We've now calculated three probabilities associated with this problem. Great job! We know the probability of picking 0 blue marbles, 0.00391, the probability of picking 1 blue marble, 0.04688, and the probability of picking 2 blue marbles, 0.21094, when picking 4 marbles out of a bag containing 6 blue marbles and 2 white marbles.

Figure C1 (continued) – Longest segment of math lesson, analyzed for level of element interactivity (see Figure C2).

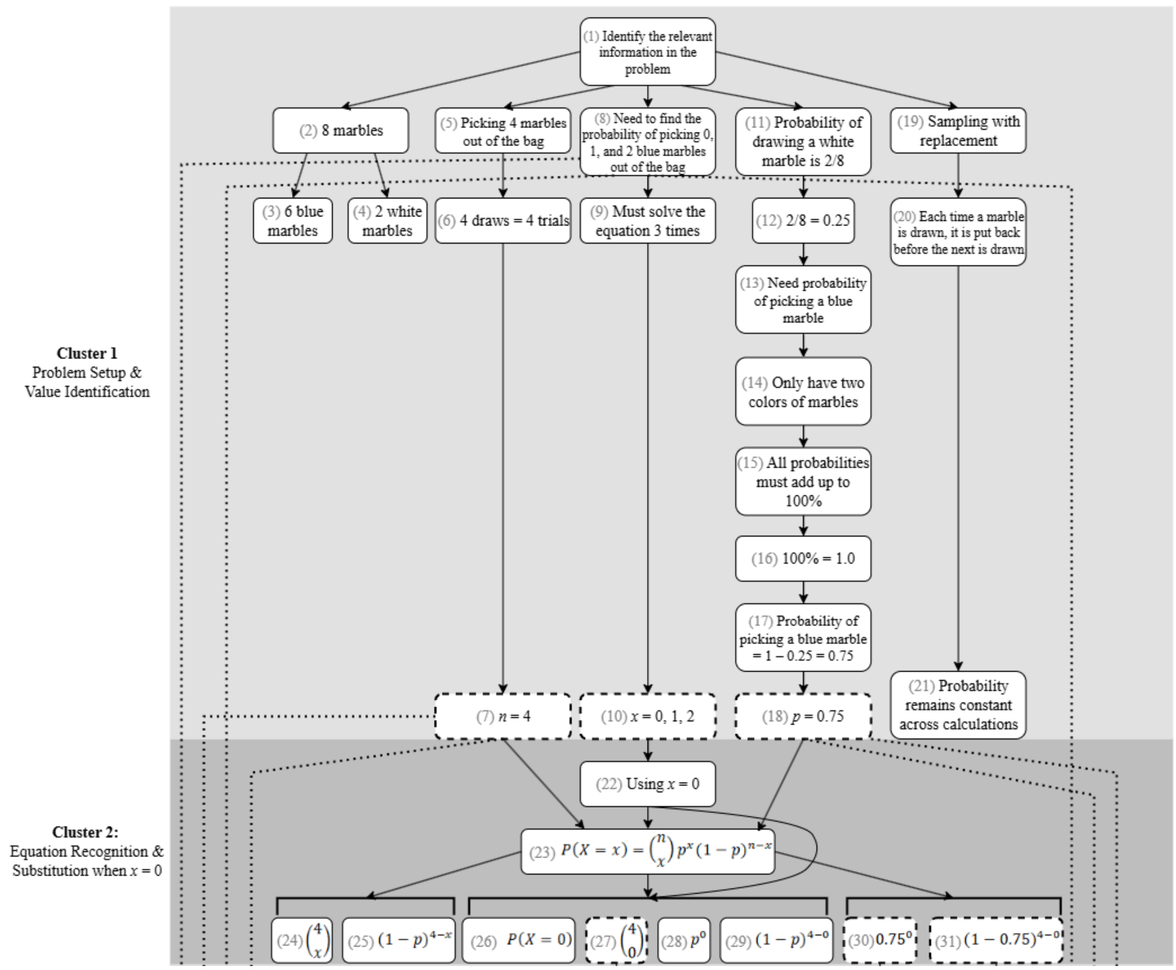


Figure C2 – Visual representation of the 14 element clusters in the math lesson segment shown in Figure C1.

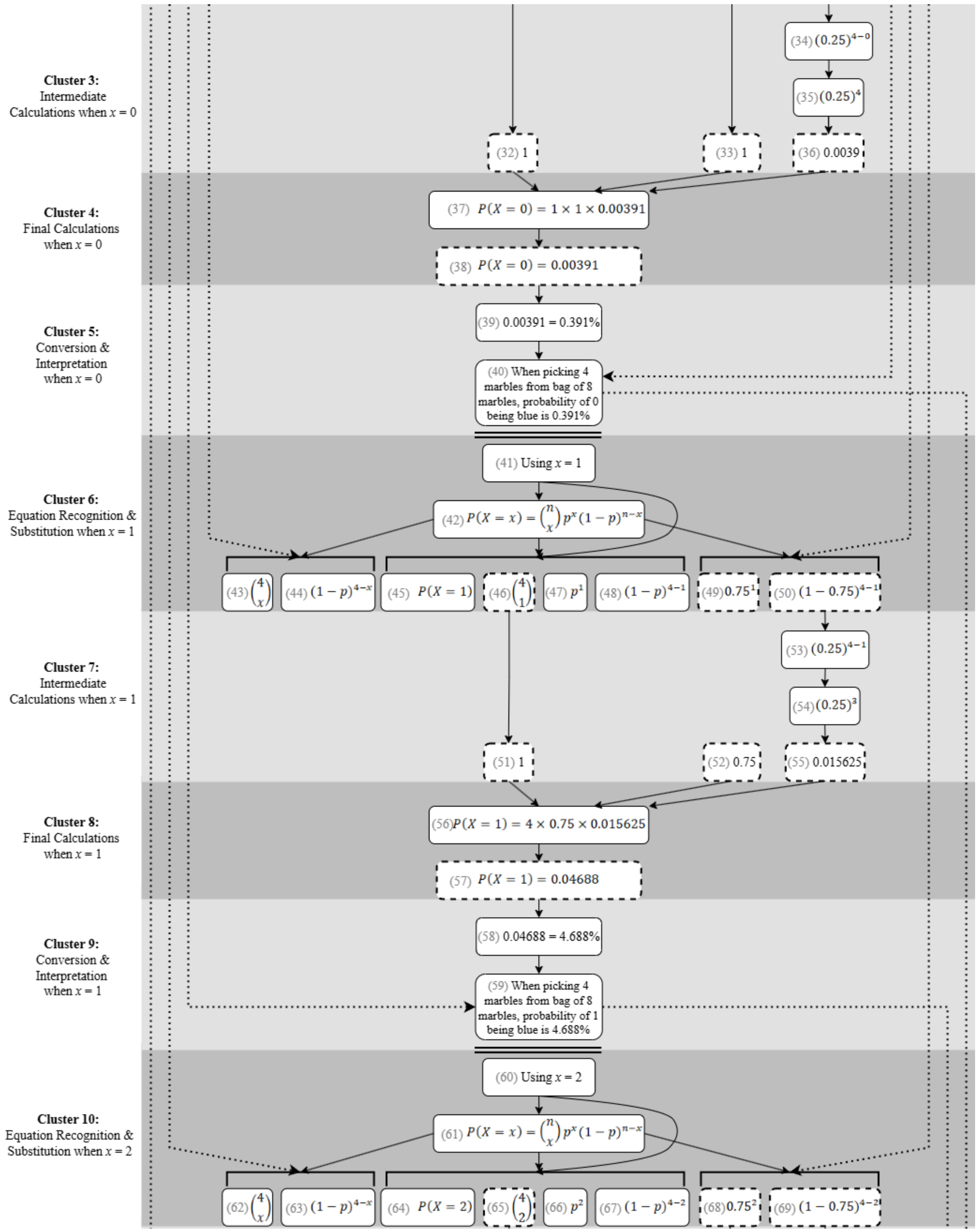


Figure C2 (continued) – Visual representation of the 14 element clusters in the math lesson segment shown in Figure C1.

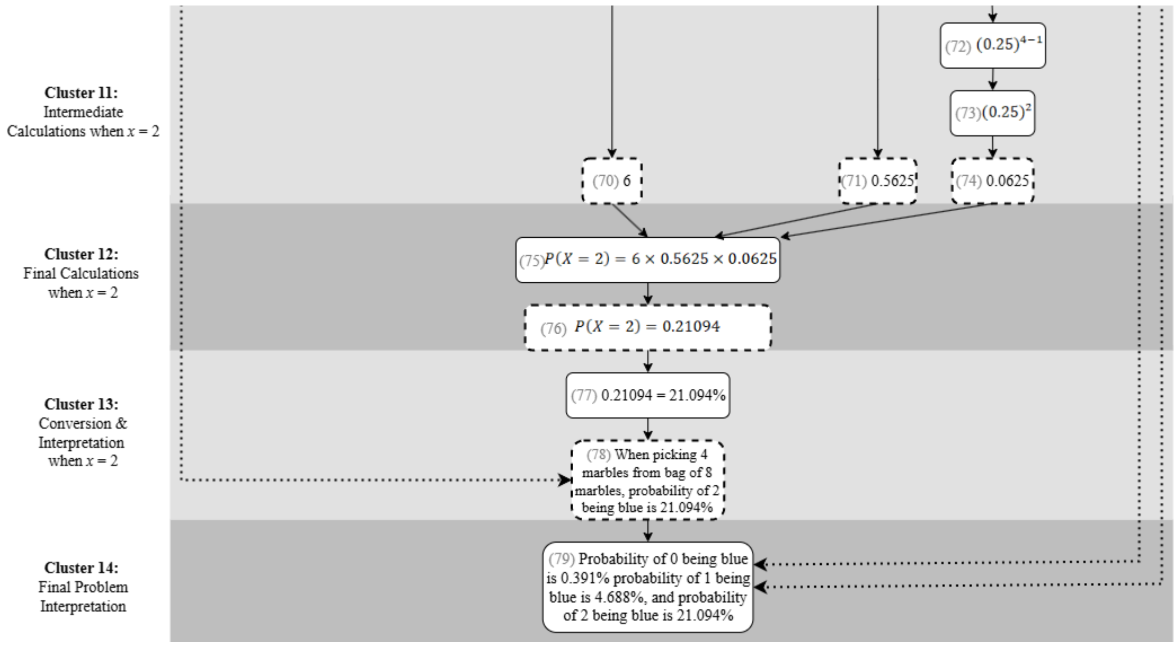
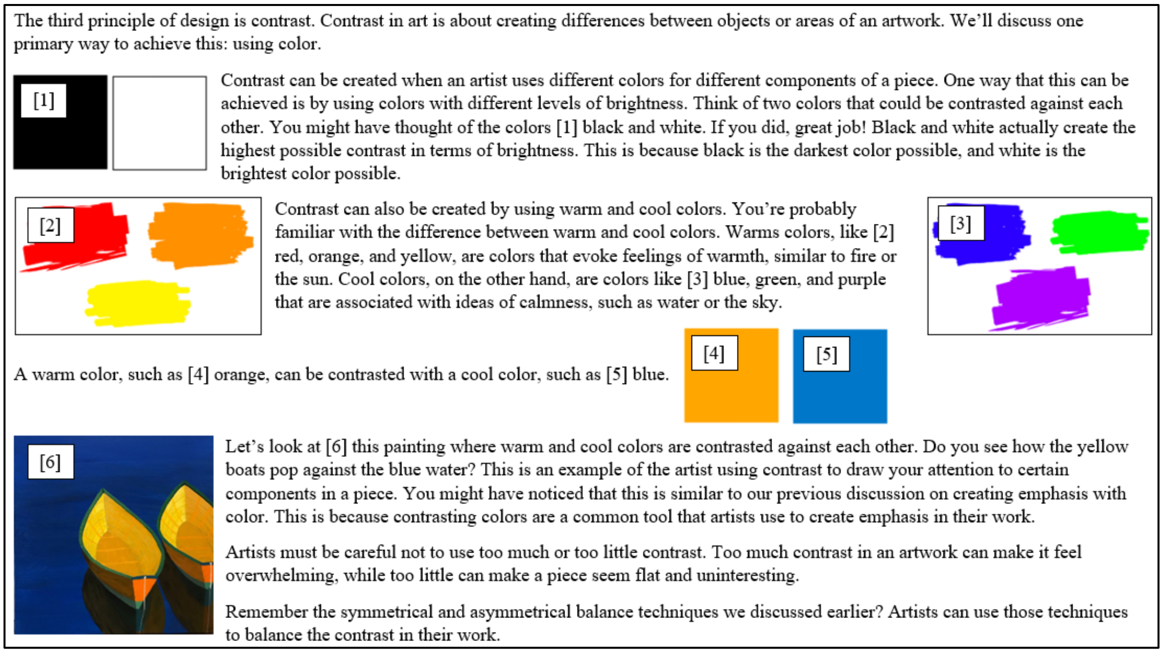


Figure C2 (continued) – Visual representation of the 14 element clusters in the math lesson segment shown in Figure C1.

Note. Alternating shaded regions indicate cluster boundaries. Elements within the same cluster must be processed simultaneously for comprehension of that cluster. Solid-box elements appear in only one cluster, while dashed-box elements are shared between adjacent clusters. Solid arrows show connections within a cluster and between adjacent clusters, while dotted arrows show connections across non-adjacent clusters. Brackets group elements to improve readability when one previous element connects to multiple subsequent elements. Double lines indicate that elements in a block are not connected to elements in the next block.

Full diagram. A non-segmented version of this diagram is available here.

Figure C2 shows a representative diagram of the element interactivity involved in the math lesson segment presented in Figure C1. The elements are divided into fourteen clusters. Each cluster reflects a distinct stage of reasoning, during which simultaneous processing of those elements (i.e., element interactivity) is required for comprehension of that phase of problem solving. Cluster 1 (*Problem Setup & Value Identification*) contains elements 1–21; the learner must identify the relevant information in the scenario, such as the number of marbles, number of trials, target outcome (including that they must solve the equation three times), success probability, and what it means to sample with replacement. Cluster 2 (*Equation Recognition & Substitution when* $x = 0$) contains elements 7, 10, and 18, and 22–31; the learner must recognize the Bernoulli trial equation and substitute the appropriate values for n, x, and p. Cluster 3 (*Intermediate Calculations when* $x = 0$) contains elements 27 and 30–36; the learner must observe the necessary computations on the substituted values, including solving the combination, squaring and cubing values, and simplifying expressions. Cluster 4 (*Final Calculations when* $x = 0$) contains elements 32–33 and 36–38; the learner must perform multiplication on the intermediate results to obtain the final answer. Cluster 5 (*Conversion & Interpretation when* $x = 0$) contains elements 8 and 38–40; the learner must convert the final answer from a decimal to percentage and understand what the answer represents. Clusters 6 and 10 are identical to Cluster 2, except that $x = 1$ and $x = 2$, respectively. The same applies to Clusters 7 and 11 for Cluster 3, Clusters 8 and 12 for Cluster 4, and Clusters 9 and 13 for Cluster 5. A total of 34 elements appear in multiple clusters to reflect their continued cognitive involvement across stages. The total number of interacting elements per cluster is 21, 13, 8, 5, 4, 12, 8, 5, 4, 12, 8, 5, 4, and 4, respectively. These counts, particularly those for Clusters 1–3, 6–7, and 10–11 reflect the high level of element interactivity present in the math lesson segment.



The third principle of design is contrast. Contrast in art is about creating differences between objects or areas of an artwork. We'll discuss one primary way to achieve this: using color.

[1]

Contrast can be created when an artist uses different colors for different components of a piece. One way that this can be achieved is by using colors with different levels of brightness. Think of two colors that could be contrasted against each other. You might have thought of the colors [1] black and white. If you did, great job! Black and white actually create the highest possible contrast in terms of brightness. This is because black is the darkest color possible, and white is the brightest color possible.

[2]

Contrast can also be created by using warm and cool colors. You're probably familiar with the difference between warm and cool colors. Warms colors, like [2] red, orange, and yellow, are colors that evoke feelings of warmth, similar to fire or the sun. Cool colors, on the other hand, are colors like [3] blue, green, and purple that are associated with ideas of calmness, such as water or the sky.

[3]

A warm color, such as [4] orange, can be contrasted with a cool color, such as [5] blue.

[4] [5]

[6]

Let's look at [6] this painting where warm and cool colors are contrasted against each other. Do you see how the yellow boats pop against the blue water? This is an example of the artist using contrast to draw your attention to certain components in a piece. You might have noticed that this is similar to our previous discussion on creating emphasis with color. This is because contrasting colors are a common tool that artists use to create emphasis in their work.

Artists must be careful not to use too much or too little contrast. Too much contrast in an artwork can make it feel overwhelming, while too little can make a piece seem flat and uninteresting.

Remember the symmetrical and asymmetrical balance techniques we discussed earlier? Artists can use those techniques to balance the contrast in their work.

Figure C3 – Longest segment of art lesson, analyzed for level of element interactivity (see Figure C4).

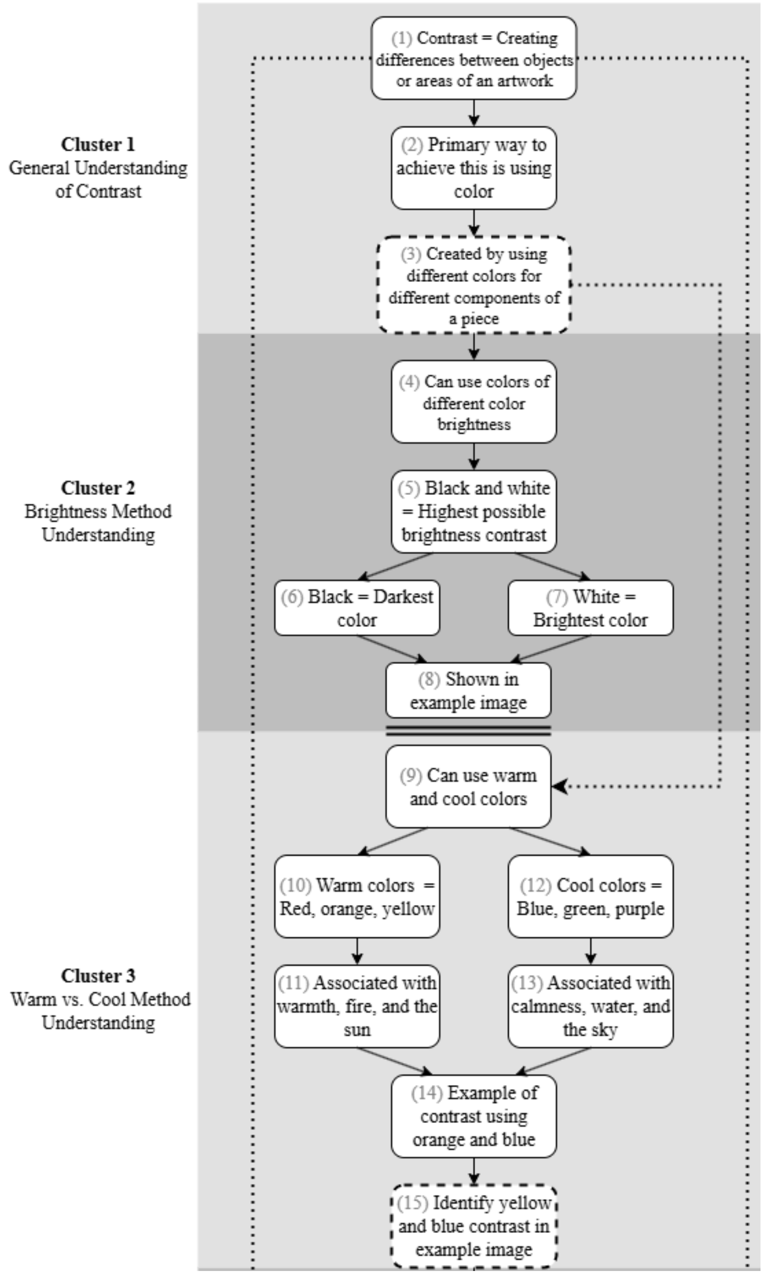


Figure C4 – Visual representation of the five element clusters in the math lesson segment shown in Figure C3.

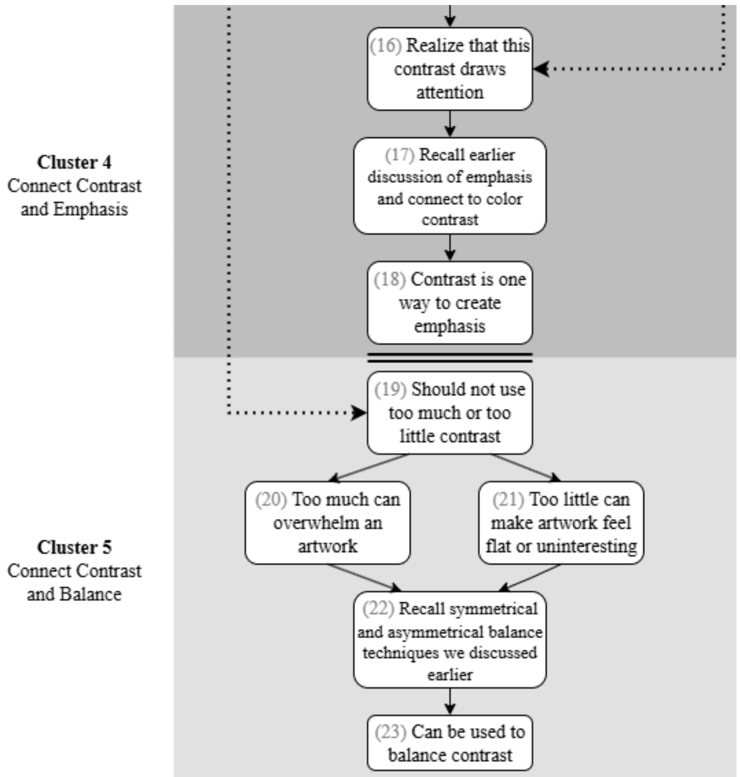


Figure C4 (continued)– Visual representation of the five element clusters in the math lesson segment shown in Figure C3.

Note. Alternating shaded regions indicate cluster boundaries. Elements within the same cluster must be processed simultaneously for comprehension of that cluster. Solid-box elements appear in only one cluster, while dashed-box elements are shared between adjacent clusters. Solid arrows show connections within a cluster and between adjacent clusters, while dotted arrows show connections across non-adjacent clusters. Brackets group elements to improve readability when one previous element connects to multiple subsequent elements. Double lines indicate that elements in a block are not connected to elements in the next block.

Full diagram. A non-segmented version of this diagram is available here.

Figure C4 shows a representative diagram of the element interactivity involved in the art lesson segment presented in Figure C3. The elements are grouped into five clusters, each representing a distinct stage of reasoning. As in the previous diagram, understanding that stage of problem solving requires simultaneous processing of the elements within the cluster (i.e., element interactivity). Cluster 1 (*General Understanding of Contrast*) contains elements 1–3; the learner must understand the definition of contrast, including that the primary way to achieve it is by using different colors for difference parts of an artwork. Cluster 2 (*Brightness Method Understanding*) contains elements 3–8; the learner must understand that contrast can be created by using colors of different brightness levels, recognize that black and white create the highest possible brightness contrast, and recognize the contrast in the example image. Cluster 3 (*Warm vs. Cool Method Understanding*) contains elements 3 and 9–15; the learner must understand that contrast can be created by using warm versus cool colors, recognize examples of warm and cool colors, understand what warm and cool colors are associated with, recognize the contrast of orange and blue in an example image, and identify yellow and blue as contrasting in an example image. Cluster 4 (*Connect Contrast and Emphasis*) contains elements 1 and 15–18; the learner must realize that the contrast in the example image draws attention, recall the earlier discussion of emphasis, and connect emphasis to contrast by realizing that contrast is one way to create emphasis. Cluster 5 (*Connect Contrast and Balance*) contains elements 1 and 19–23; the learner must understand why they should neither use too much nor too little contrast, recall the earlier discussion of symmetrical and asymmetrical balance techniques, and realize that these techniques can be used to balance contrast. Elements 1, 3, and 15 appear in multiple clusters to reflect their continued cognitive involvement across stages. The total number of interacting elements per cluster is 3, 6, 8, 5 and 6, respectively. These counts, particularly for Cluster 1, reflect the low level of element interactivity present in the art lesson.

Given that the art lesson segment contained only five clusters, these will be compared against the first five clusters of the math segment. As shown in Table C1, the math segment has more interacting elements than the art segment in two of the five clusters (Clusters 1 and 2), with large differences in both cases (21 vs. 3 and 13 vs. 6, respectively). The two segments have the same number of interacting elements in Clusters 3 and 4, and the art segment has more in Cluster 5, with the difference being small (6 vs. 4). In addition, more elements are shared between clusters in the math segment than in the art segment (34 vs. 5). Together with the diagrams in Figures C2 and C4, these counts highlight the greater element interactivity in the math segment compared to the art segment. Combined with the matching conclusion from section *3.4.2 Lessons*, and given that the segments examined in section *3.4.2* and this appendix serve as a representative sample, this suggests that the math lesson as a whole has a higher level of element interactivity than the art lesson.

	Number of Interacting Elements													
	Clusters													
Lesson	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Math	21	13	8	5	4	12	8	5	4	12	8	5	4	4
Art	3	6	8	5	6					N/A				

Table C1 – Counts of interacting elements per cluster, organized by Lesson.

Number of Elements Shared between Clusters																				
Cluster Pairs																				
1, 2	1, 3	1, 4	1, 5	1, 6	1, 9	1, 10	1, 13	2, 3	3, 4	4, 5	5, 14	6, 7	7, 8	8, 9	9, 14	10, 11	11, 12	12, 13	13, 14	Total
3	0	0	1	2	1	2	1	3	3	1	1	3	3	1	1	3	3	1	1	**34**
1	1	1	1		N/A			0	1	0						N/A				**5**

Note. Only cluster pairs in which at least one lesson contained an instance of a shared element are shown. All other possible pairs had a count of 0.

Table C2 – Counts of elements shared between clusters and total shared elements, organized by Lesson.

APPENDIX D. DEFINITIONS & EXAMPLES OF QUESTION TYPES

Question Type	Recognition	Recall	Transfer		
			Near Transfer	Medium Transfer	Far Transfer
Defintion	Correctly identify information from the lesson when provided multiple options	Provide information from the lesson without the aid of options	Apply the information from the lesson to examples or scenarios that are very similar to those in the lesson	Apply the information to examples or scenarios that share similarities with those in the lesson but include minor variations or additional considerations	Apply the information from the lesson to examples or scenarios that are conceptually related to, yet significantly different from, the examples or scenarios in the lesson

Table D1 – Definitions for the recognition, recall, and transfer (near, medium, far) question types.

Note. These definitions apply to the post-tests and the retention tests for both math and art.

Bernoulli Trials					
Question Type	**Recognition**	**Recall**	**Transfer**		
			Near Transfer	**Medium Transfer**	**Far Transfer**
Example	**What does x represent in the Bernoulli trial equation?** A.) The total number of Bernoulli trials B.) The number of successful trials C.) The probability of an unsuccessful trial D.) The probability of a successful trial	**How would you define a "failure" in the context of flipping a coin for heads? Why?**	You're playing a game with your family. During your turn, you must roll 5 six-sided dies. Each die has a 16% chance of rolling a 4. You are hoping that exactly 3 of the die will roll a 4. What is the probability of this occurring? **What is the value or values of x in this problem? If there are multiple values, separate each value with a comma.**	You're managing a small orchard and decide to plant 8 apple trees. Each tree has a 62% chance of bearing fruit. You need 5 of the trees to bear fruit for an upcoming order. What is the probability that the exact number of trees needed to fulfill the order will bear fruit? **The preceding steps to solve this problem have been solved for you. Given the current status shown below, what is the next step?** **$P(X = x) = 56 \times 0.09161 \times 0.05487$**	A new game show features a challenge called "The Lucky Strike". In this challenge, a contestant is faced with 10 identical boxes, out of which 3 contain a gold coin and the rest contain a silver coin. The contestant wins a prize if they select exactly 2 gold coins out of 5 boxes they choose to open, with each choice being independent of the others. **What does the fact that each choice is independent of the others imply about the problem's p value?**
Answer	B.) The number of successful trials	The coin landing on tails because this is the opposite of the desired outcome	2	Multiply the 3 numbers together	This implies that the probability of getting a success does not change depending on which box is opened/is the same for all boxes
Number of Question Type on Test	8	8	8	14	12

Table D2 – Sample questions and answers from the Bernoulli trials post-test.

Note. Organized by question type.

Principles of Design					
Question Type	Recognition	Recall	Transfer		
			Near Transfer	Medium Transfer	Far Transfer
Example	**What is the primary purpose of using emphasis?** A.) To draw the viewer's attention to a specific part of an artwork B.) To make some elements visually heavier than others C.) To create an artwork with two visually identical halves D.) To create a sense of harmony in an artwork	**In any order, list the principles of design that were discussed in the lesson. Separate responses with a comma.**	 **What technique was primarily used to create emphasis in this image?**	 **Consider the colors used in this painting. How is contrast being created?**	**Imagine that you are designing a flyer for an event. You want to add contrast to a portion of the flyer to draw attention to it. Unfortunately, your printer cannot print in color. Are you still able to achieve contrast in this situation? If so, how?**
Answer	A.) To draw the viewer's attention to a specific part of an artwork	Emphasis, balance, contrast, proportion, unity	Color	Warm color (red) versus cool color (green)	Yes, by using black and white/light gray or colors with different levels of brightness
Number of Question Type on Test	8	8	7	17	10

Table D3 – Sample questions and answers from the principles of design post-test.

Note. Organized by question type.

APPENDIX E. INSTRUCTIONS & SAMPLE QUESTIONS FROM BERNOULLI TRIALS PRE-TEST

Note: Correct answers in green.

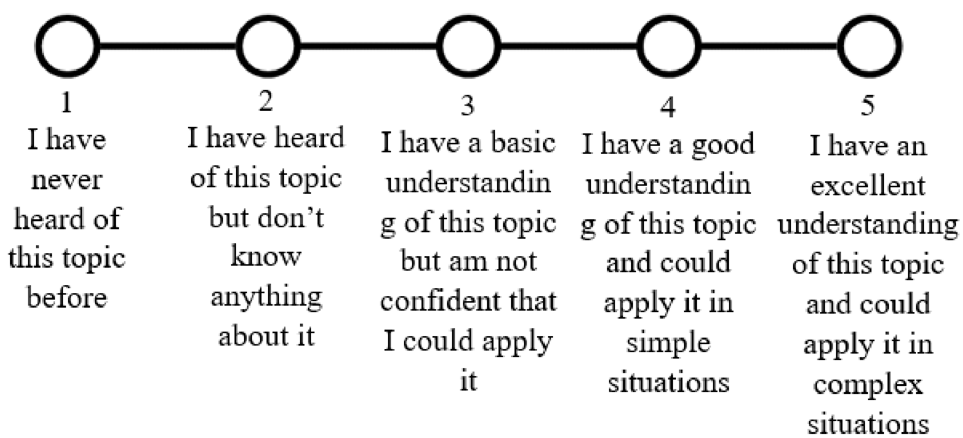
Please respond to the following two questions. You will not be able to return to these questions after clicking “Next”.

1.) Do you know what a Bernoulli trial is?

A. Yes

B. No

2.) Rate your knowledge about Bernoulli trials on a scale from 1 to 5.



Answer the following 12 questions to the best of your ability. If you don’t know the answer to a question, respond with the option “E. No response”. After you have finished a question, click “Next” to move to the next question. After clicking “Next”, you will not be able to return to any previous questions.

1.) How many possible outcomes are there in a Bernoulli trial?

A.) 2

B.) 3

C.) 4

D.) It depends on the problem

E.) No response

7.) What do x, n, and p represent in the Bernoulli trial equation?

A.) Number of a specific outcome, number of trials, and probability of a specific outcome

B.) A continuous variable, sample size, and population size ratio

C.) Number of items, number of ways to choose the items, and probability of choosing one item

D.) Number of events that occur during a certain timeframe, length of timeframe, and average rate of occurrence

E.) No response

8.) If x=2, n=5, and p=0.5 are in the equation for a Bernoulli trial, what does this mean?

A.) The variable x, which represents a continuous variable, is set to 2, the sample size is 5, and 5 represents 50% of the total population size.

B.) There are 2 distinct items, there are 5 different ways to choose these items, and the probability of picking any single item during each selection is 0.5.

C.) 2 independent events occur during a 5-minute interval, and the average rate of occurrence is 0.5 times per minute.

D.) The probability of 2 outcomes occurring in a series of 5 experiments is calculated, where there is a 50% probability of the outcome in each experiment.

E.) No response

Refer to the problem below for questions 9-11.

Consider the following scenario: Suppose you have a kickball that is red on one side and blue on the other side. A study of 15 trials found that the chance of kicking the ball and it stopping with the red side upwards is 0.5. If you were to kick the ball 10 times, how likely is it to stop with the blue side upwards 7 of those times?

10.) What do x, n, and p equal in this scenario, in that order?

A.) 10, 7, 0,5

B.) 0.5, 10, 7

C.) 7, 10, 0.5

D.) 7, 15, 0.5

E.) No response

11.) If the answer to the problem is 0.117, what does this mean?

A.) The probability of the ball landing with the red side upwards 7 out of 15 times is 11.7%.

B.) The probability of the ball landing with the blue side upwards 11.7% of the time is 50%.

C.) The probability of the ball landing with the red side upwards 7 out of 10 times is 11.7%.

D.) The probability of the ball landing with the blue side upwards 7 out of 10 times is 11.7%.

E.) No response

APPENDIX F. INSTRUCTIONS & SAMPLE QUESTIONS FROM PRINCIPLES OF DESIGN PRE-TEST

Note: Correct answers in green.

Please respond to the following two questions. You will not be able to return to these questions after clicking “Next”.

1.) Do you know what the principles of design are?

A. Yes

B. No

2.) Rate your knowledge about the principles of design on a scale from 1 to 5.

1	2	3	4	5
I have never heard of this topic before	I have heard of this topic but don’t know anything about it	I have a basic understandin g of this topic but am not confident that I could apply it	I have a good understandin g of this topic and could apply it in simple situations	I have an excellent understanding of this topic and could apply it in complex situations

Answer the following 11 questions to the best of your ability. If you don’t know the answer to a question, respond with the option “E. No response”. After you have finished a question, click “Next” to move to the next question. After clicking “Next”, you will not be able to return to any previous questions.

1.) Which of the options accurately lists principles of design?

A.) Placement, harmony, dimension, mood, inspiration

B.) Dimension, flow, symmetry, harmony, contrast

C.) Tone, placement, depth, dimension, pattern

D.) Balance, emphasis, unity, proportion, contrast

E.) No response

3.) What are three methods that can be used to create emphasis in art?

A.) Texture, symmetry, frame

B.) Placement, size, color

C.) Motion, transparency, shadow

D.) Pattern, brightness, gradient

E.) No response

4.) Which color would be considered visually heavier: orange or red?

A.) Orange

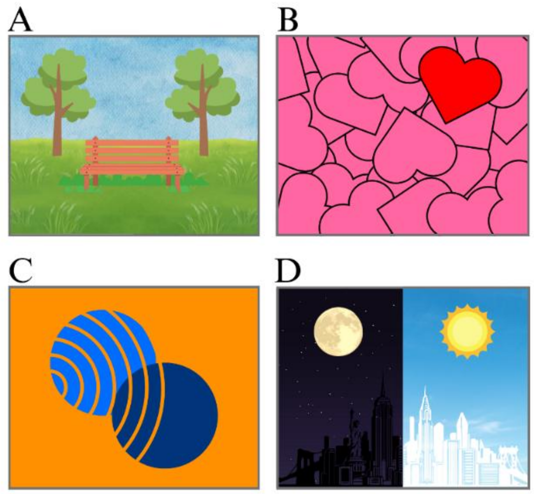
B.) Red

C.) Neither because they weigh the same

D.) Neither because colors are not capable of having visual weight

E.) No response

For questions 11-12, refer to the image below.



11.) Which principles of design do the images for A, B, C, D, and E represent, in that order?

A.) Balance, emphasis, contrast, proportion

B.) Harmony, pattern, movement, balance

C.) Tone, unity, color, emphasis

D.) Symmetry, color, movement, brightness

E.) No response

12.) If an artist used all four of these principles in one painting, what would this be most likely to do?

A.) This question is irrelevant because an artist could not use all four of these principles at once

B.) Invoke the principle of design called unity

C.) Create feelings of visual chaos, making the painting feel overwhelming

D.) Enhance the depth of the painting

E.) No response

APPENDIX G. REDUCED INSTRUCTIONAL MATERIALS MOTIVATION SURVEY (RIMMS; LOORBACH ET AL., 2015)

5-point Likert Scale: 1 (or A) = Not true; 2 (or B) = Slightly true; 3 (or C) = Moderately true; 4 (or D) = Mostly true; 5 (or E) = Very true

Attention

1. The quality of the writing helped to hold my attention.
2. The way the information is arranged on the pages helped keep my attention.
3. The variety of exercises helped keep my attention on the lesson.

Relevance

4. It is clear to me how the content of this material is related to things I already know.
5. The content and style of writing in this lesson convey the impression that its content is worth knowing.
6. The content of this lesson will be useful to me.

Confidence

7. As I worked on this lesson, I was confident that I could learn the content.
8. After working on this lesson for a while, I was confident that I would be able to pass a test on it.
9. The good organization of the content helped me be confident that I would learn this material.

Satisfaction

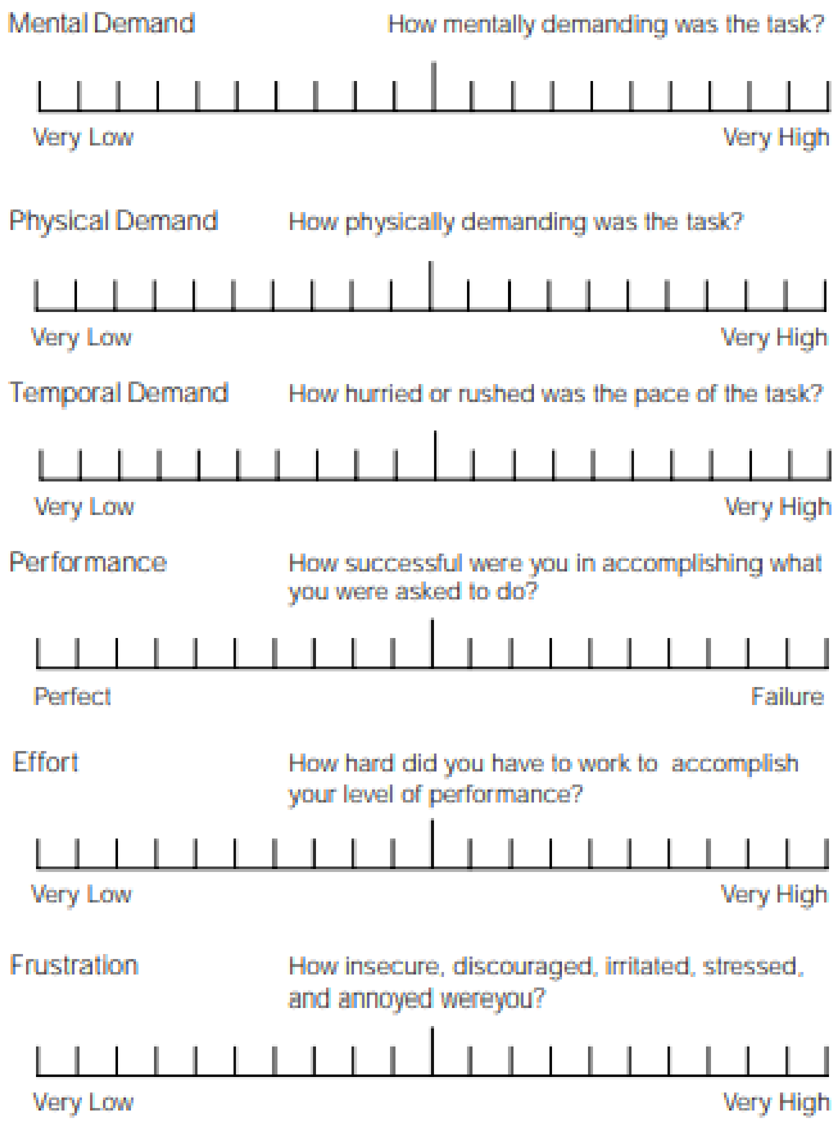
10. I enjoyed this lesson so much that I would like to know more about this topic.

11. I really enjoyed studying this lesson.

12. It was a pleasure to work on such a well-designed lesson.

Note. Questions organized by subscale.

APPENDIX H. NASA-TASK LOAD INDEX (NASA-TLX; HART & STAVELAND, 1988)



Mental Demand How mentally demanding was the task?

Very Low Very High

Physical Demand How physically demanding was the task?

Very Low Very High

Temporal Demand How hurried or rushed was the pace of the task?

Very Low Very High

Performance How successful were you in accomplishing what you were asked to do?

Perfect Failure

Effort How hard did you have to work to accomplish your level of performance?

Very Low Very High

Frustration How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low Very High

Note. Questions organized by subscale.

APPENDIX I. REVISED AGENT PERSONA INSTRUMENT (API-R; SCHROEDER ET AL., 2018)

All items should be presented with a 5-point Likert scale, ranging from 1=Strongly disagree to 5=Strongly agree.

Facilitating Learning (10 items)

The agent led me to think more deeply about the presentation.

The agent made the instruction interesting.

The agent encouraged me to reflect what I was learning.

The agent kept my attention.

The agent communicated the main ideas clearly.

The agent helped me to concentrate on the presentation.

The agent focused me on the relevant information.

The agent helped me learn the material.

The agent was good at teaching.

The agent was easy to learn from.

Credible (5 items)

The agent was knowledgeable.

The agent was intelligent.

The agent was useful.

The agent was helpful.

The agent was an effective teacher.

Humanlike (5 items)

The agent has a personality.

The agent’s emotion was natural.

The agent was humanlike.

The agent’s movement was natural.

The agent showed emotion.

Engaging (5 Items)

The agent was engaging.

The agent was enthusiastic.

The agent was entertaining.

The agent was motivating.

The agent was easy to connect with.

Note. Questions organized by subscale.

APPENDIX J. SAMPLE QUESTIONS FROM OPINION-BASED SURVEY

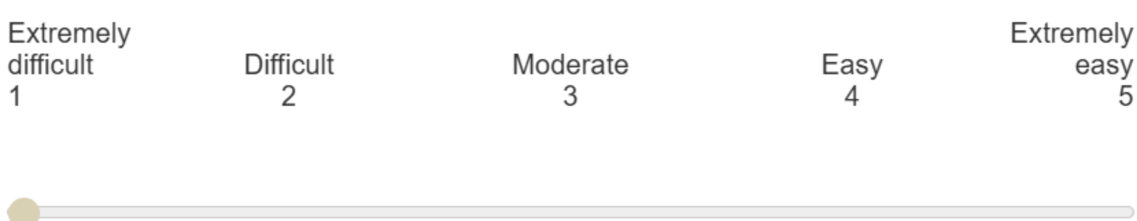
Note: Survey 1 provides sample questions from a survey asking about the lesson content. This includes sample instructional text, Likert scale question (question 1), multiple choice questions (questions 2-3), and text-entry question (question 3 follow-up). Participants in both the With-PA and No-PA conditions will receive these questions. The instructions will be modified to mention “the arrows” rather than “the pedagogical agent” in the No-PA condition.

<u>Survey 1 of 5: Lesson Content</u>

For the following questions, please consider only the lesson content. Do NOT consider the pedagogical agent or design of the lesson (background, font, etc.).

Some questions will require you to use a slider to respond. To ensure your responses are recorded, move the slider on these questions, even if you wish to choose the default position. After moving it, you can return it to the default position if that's your intended response.

1.) On a scale from 1 to 5, how would you rate the difficulty of the lesson?



2.) Did the lesson provide you with enough examples to understand the material?

A.) Yes

B.) No

3.) Was there any part of the lesson that you felt was not explained well enough or left you feeling confused?

A.) Yes

B.) No

If you responded with "Yes" on question 3, feel free to provide example(s) below.

Note: Survey 5 provides sample questions from a survey asking about general aspects of the pedagogical agent or arrows. This includes sample instructional text, multiple choice question (question 1), and Likert scale question (question 2). Participants in both the With-PA and No-PA conditions will receive these questions. The instructions will be modified to mention “the arrows” rather than “the pedagogical agent” in the No-PA condition.

Survey 5 of 5: Agent in General

For the following questions, please consider the pedagogical agent in general.

The pedagogical agent refers to the virtual character that was positioned on the left side of the whiteboard in the video.

Some questions will require you to use a slider to respond. To ensure your responses are recorded, move the slider on these questions, even if you wish to choose the default position. After moving it, you can return it to the default position if that's your intended response.

1.) In general, did you find the pedagogical agent to be distracting?

A.) Yes

B.) No

2.) On a scale from 1 to 5, how and to what degree do you think the pedagogical agent impacted your learning?

Significantly harmed my learning	Sort of harmed my learning	Did not harm or enhance my learning	Sort of enhanced my learning	Significantly enhanced my learning
1	2	3	4	5

APPENDIX K. GEORGIA TECH DEMOGRAPHIC SURVEY

The following are demographic questions. Please answer these questions to the best of your ability. Your responses will be kept confidential.

1. What gender do you identify as?
   - ☐ Female
   - ☐ Male
   - ☐ Non-Binary
   - ☐ Other
   - ☐ Prefer not to answer
2. What is your current age in years?
3. What year of college are you in?
   - ☐ Freshman
   - ☐ Sophomore
   - ☐ Junior
   - ☐ Senior
   - ☐ Other (please specify)
4. What is your major? If you have more than one, separate your responses with a comma.
5. Do you have a minor?
   - ☐ Yes
   - ☐ No
6. Please list your minor(s) below. If you have more than one, separate your responses with commas.

7. Have you previously taken or are you currently enrolled in a statistics course?
   - ☐ Yes
   - ☐ No

8. To the best of your memory, please list the **course name and number and the university** where you took the statistics course(s). If you don't know some of this information, please just include the information that you do know. Type separate responses on different lines.

9. Have you previously taken or are you currently enrolled in an art or design course?
   - ☐ Yes
   - ☐ No

10. To the best of your memory, please list the course name and number and the university where you took the art and/or design course(s). If you don't know some of this information, please just include the information that you do know. Type separate responses on different lines.

Note. Question 9 was not asked in Experiment 2. Questions 6, 8, and 10 only appeared if the immediately preceding question was answered with “Yes”.

APPENDIX L. ATTENTION CHECKS & MESSAGES FOR PROLIFIC PARTICIPANTS

Before you begin, please ensure you can meet the following requirements:

1. Complete this study on a **computer** (rather than a smartphone or tablet).
2. Use **headphones or speakers**, as the study includes **audio**. Make sure you are in a quiet environment where you can listen without distractions.
3. Complete this study in **one sitting**. Brief interruptions (e.g., for water or a restroom break) are completely understandable, but **extended breaks should be avoided**.

Figure L1 – First Page of Qualtrics Survey (Study Requirements)

For this question, please select "1".

1

2

3

Figure L2 – Attention Check: During Educational Technology Preferences Survey

13.) For this question, please move the slider to "4 - Mostly true".

Not true	Slightly true	Moderately true	Mostly true	Very true
1	2	3	4	5

Figure L3 – Attention Check: At End of RIMMS

For this question, please select "A.) I agree".

A.) I agree

B.) No opinion

C.) I do not agree

Figure L4 – Attention Check: During Post-Test

You did not follow the instructions on one of the previous questions. Please read all questions carefully. If you fail any additional attention checks, you will be removed from the study and will not receive compensation.

Figure L5 – Warning Message After First Failed Attention Check

Thank you for your interest in this study.

Based on your responses, you failed multiple attention checks that were included to ensure data quality. As a result, your participation has ended early, and you are not eligible to receive compensation for this study.

Please **return** your submission by clicking the "Return Submission" button in your Prolific dashboard. Returning your submission ensures that your approval rating will not be affected.

If your submission is not returned within 48 hours, it may be rejected.

Thank you for your time.

Figure L6 – Exclusion Message After Second Failed Attention Check

APPENDIX M. PROLIFIC INFORMED CONSENT DOCUMENT

You are being asked to be a volunteer in a research study. The purpose of this study is to investigate the effectiveness of educational technology at facilitating learning. During this study, you will be asked to respond to a demographic survey and a survey on your preferences regarding several types of educational technology. Then, you will take a pre-test, watch a video lesson, take a post-test, and respond to several surveys. The duration of this study will be approximately **1 hour and 15 minutes** (1.25 hours). You reserve the right to stop participating at any time.

Your data will be kept confidential using a code/ID system. The data itself will be stored digitally on password encrypted computers located behind locked doors. The risks involved are no greater than those involved in daily activities. We will comply with any applicable laws and regulations regarding confidentiality.

You will be compensated with $12 through Prolific for the successful completion of this study. Upon successful completion, you will be redirected to Prolific via a URL. Your submission will be reviewed within 48 hours.

It is possible that you may be removed from the study early based on your responses to the educational technology preferences survey, due to the maximum number of participants in your assigned group already being reached. If this occurs, you will be asked to return your submission. Within 48 hours of your submission being returned, you will receive a partial payment of $4 via Prolific’s Bonus Payment feature, in recognition of your participation up to that point.

It is also possible that you may be removed from the study early if your responses indicate attention check failure. If this occurs, you will be asked to return your submission and will not receive payment. If your submission is not returned within 72 hours, it may be rejected.

Only participants located in the U.S. are permitted to participate in this study. Participants located in China, the EU, and/or the UK are excluded from participation.

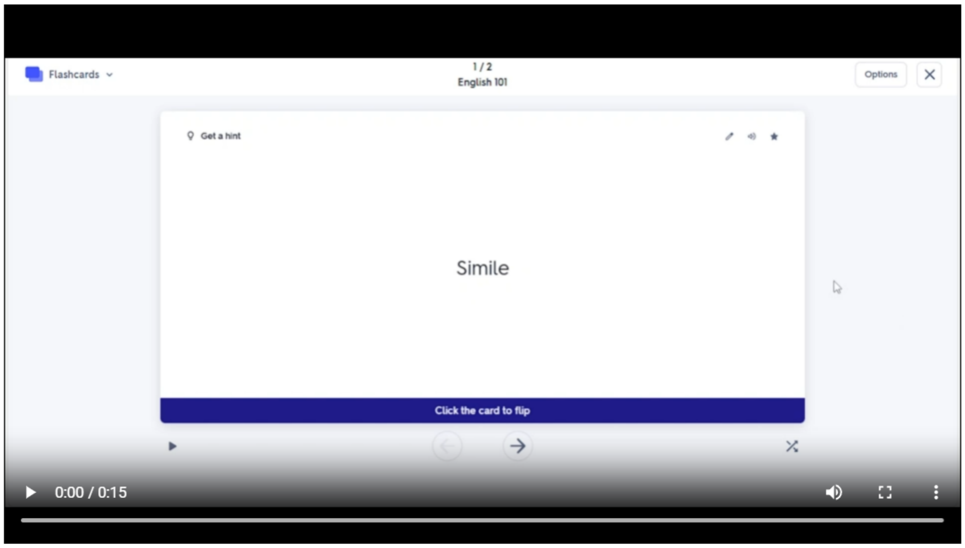
To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB may review study records. The Office of Human Research Protections may also look at study records. If you have any questions about the study, you may contact Dr. Richard Catrambone (Principal Investigator) at telephone 404-894-2680. If you have any questions about your rights as a research subject, you may contact Georgia Institute of Technology Office of Research Integrity Assurance at IRB@gatech.edu. Thank you for participating in this study.

If you consent to participating in this study, please select “I agree”. If you do not consent, select “I do not agree”.

I agree

I do not agree

APPENDIX N. SAMPLE QUESTIONS FROM EDUCATIONAL TECHNOLOGY PREFERENCES SURVEY



The video above depicts an example of **digital flashcards**. Please watch the video before answering the following questions.

If you were given the choice, would you want to learn with **digital flashcards**?

Yes

No

How strongly do you feel about your response to the previous question?

Very weak	Weak	Moderate	Strong	Very strong
1	2	3	4	5

Figure N1 – Filler Question Example

The video above depicts an example of a **virtual instructor**. Please watch the video before answering the following questions.

If you were given the choice, would you want to learn with a **virtual instructor**?

Yes

No

How strongly do you feel about your response to the previous question?

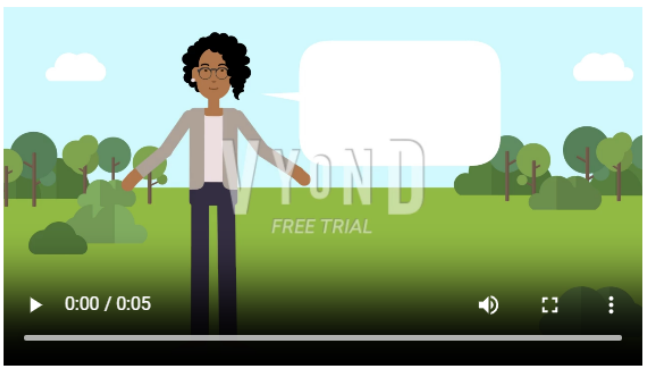
Very weak	Weak	Moderate	Strong	Very strong
1	2	3	4	5

Figure N2 – Target Question on Study's PA

APPENDIX O. POST-TASK PEDAGOGICAL AGENT PREFERENCE QUESTION

Final Question

As a reminder, a pedagogical agent (PA) is a digital character included in a computer-based learning or training video to enhance the user's learning experience through roles such as guiding the learner. PAs can vary widely in design, including differences in appearance (e.g., human or non-human), gender, race or ethnicity, level of realism, voice, and subject matter. The two example PAs from the initial survey are shown below as a reference. These videos represent only two of the infinite possible PA designs.



Considering only the function of a PA in a video lesson (and not its design, appearance, voice, or other characteristics), would you prefer to learn with or without a PA?

I would prefer to learn with a PA

I would prefer to learn without a PA

APPENDIX P. PROLIFIC DEMOGRAPHIC SURVEY

The following are demographic questions. Please answer these questions to the best of your ability. Your responses will be kept confidential.

1. What gender do you identify as?
   - ☐ Female
   - ☐ Male
   - ☐ Non-Binary
   - ☐ Other
   - ☐ Prefer not to answer

2. What is your current age in years?

   [ ]

3. What is the highest level of education you have completed?
   - ☐ Less than high school
   - ☐ High school diploma or GED
   - ☐ Associate degree
   - ☐ Bachelor's degree
   - ☐ Other (please specify)

     [ ]

4. Are you currently enrolled as an undergraduate student at a college or university (e.g., Associate's or Bachelor's program)?

   It is okay if you are not actively taking classes, such as if you are currently on a break between semesters. If this is the case, respond with "Yes."

   Additionally, if you graduated college in the Spring 2025 semester, please select "Yes."
   - ☐ Yes
   - ☐ No

5. What type of institution are you enrolled in?
   - ☐ 2-year college (e.g., community college)
   - ☐ 4-year college or university (undergraduate)
   - ☐ Other (please specify)

     [ ]

6. What year of college are you currently in?
   - ☐ Freshman
   - ☐ Sophomore
   - ☐ Junior
   - ☐ Senior
   - ☐ Other (please specify)

7. What is your major or program area? If you have more than one, separate your responses with a comma.

8. Do you have a minor?
   - ☐ Yes
   - ☐ No

9. Please list your minor(s). If you have more than one, separate your responses with a comma.

10. Have you previously taken or are you currently enrolled in a statistics course?

    Your response to this question will not affect your eligibility to participate in the study. Please answer honestly.
   - ☐ Yes
   - ☐ No

11. To the best of your memory, please list the course name and number and the university where you took the statistics course(s). If you don't know some of this information, please just include the information that you do know. Type separate responses on different lines.

Note. Used only in Experiment 2 for Prolific participants. Questions 5-9 appeared only if participants answered “Yes” to question 4. Questions 9 and 11 appeared only if the immediately preceding question was answered with “Yes”.

REFERENCES

Andre, A.D., & Wickens, C.D. (1995). When users want what's not best for them. *Ergonomics in Design, 3*(4), 10-14. https://doi.org/10.1177/106480469500300403

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology, 94*(2), 416-427. https://doi.org/10.1037/0022-0663.94.2.416

Baddeley, M. (2006). [Review of Efficiency in learning: Evidence-based guidelines to manage cognitive load, by R. C. Clark, F. Nguyen, & J. Sweller]. *Performance Improvement, 45*(9), 46-47. https://doi.org/10.1002/pfi.4930450920

Barabasz, A. F. (1968). A study of recall and retention of accelerated lecture presentation. *Journal of Communication, 18*(3), 283-287. https://doi.org/10.1111/j.1460-2466.1968.tb00077.x

Baylor, A. L., & Kim, S. (2009). Designing nonverbal communication for pedagogical agents: When less is more. *Computers in Human Behavior, 25*(2), 450-457. https://doi.org/10.1016/j.chb.2008.10.008

Baylor, A. L., & Kim, Y. (2004). Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In J. C. Lester, R. M. Vicari, & F. Paraguaçu (Eds.), *Intelligent Tutoring Systems. ITS 2004. Lecture Notes in Computer Science* (*Vol. 3220*, pp. 192-201). Springer. https://doi.org/10.1007/978-3-540-30139-4_56

Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education, 15*(2), 95-115. https://link.springer.com/journal/40593

Blanchard, H. E., Lewis, S. H., Ross, D., and Cataldo, G. (1993). User performance and preference for alphabetic entry from 10-key pads: Where to put Q and Z? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 37*(3), 225-229. https://doi.org/10.1177/154193129303700303

Brachten, F., Brunker, F., Frick, N.R.J., Ross, B., & Stieglitz, S. (2020). On the ability of virtual agents to decrease cognitive load: An experimental study. *Information*

Systems and e-Business Management, 18, 187-207. https://doi.org/10.1007/s10257-020-00471-7

Castro-Alonso, J. C., Wong, R. M., Adesope, O. O., & Paas, F. (2021). Effectiveness of multimedia pedagogical agents predicted by diverse theories: A meta-analysis. *Educational Psychology Review, 33*(3), 989-1015. https://doi.org/10.1007/s10648-020-09587-1

Chen, O., Paas, F., & Sweller, J. (2023). A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educational Psychology Review, 35*(63). https://doi.org/10.1007/s10648-023-09782-w

Chen, Z.H. (2012). We care about you: Incorporating pet characteristics with educational agents through reciprocal caring approach. *Computers & Education, 59*, 1081-1088. https://doi.org/10.1016/j.compedu.2012.05.015

Choi, S., & Clark, R. E. (2006). Cognitive and affective benefits of an animated pedagogical agent for learning English as a second language. *Journal of Educational Computing Research, 34*(4), 441-466. https://doi.org/10.2190/A064-U776-4208-N145

Cinkara, E., & Bagceci, B. (2013). Learners' attitudes towards online language learning; And corresponding success rates. *Turkish Online Journal of Distance Education, 14*(2). https://dergipark.org.tr/en/pub/tojde

Clark, R. E., & Choi, S. (2005). Five design principles for experiments on the effects of animated pedagogical agents. *Journal of Educational Computing Research, 32*(3), 209-225. Scopus. https://doi.org/10.2190/7LRM-3BR2-44GW-9QQY

Clark, R. E., & Choi, S. (2007). The questionable benefits of pedagogical agents: Response to Veletsianos. *Journal of Educational Computing Research, 36*(4), 379-381. https://doi.org/10.2190/2781-3471-67MG-5033

Clark, R.C., Nguyen, F., & Sweller, J. (2005). Efficiency in learning: Evidence-based guidelines to manage cognitive load. Pfeiffer. www.pfeiffer.com

Craig, S. D., & Schroeder, N. L. (2017). Reconsidering the voice effect when learning from a virtual human. *Computers & Education, 114*, 193-205. https://doi.org/10.1016/j.compedu.2017.07.003

Craig, S. D., & Schroeder, N. L. (2019). Text-to-Speech Software and Learning: Investigating the Relevancy of the Voice Effect. *Journal of Educational Computing Research, 57*(6), 1534-1548. https://doi.org/10.1177/0735633118802877

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features and redundancy. *Journal of Educational Psychology, 94*(2), 428-434. https://doi.org/10.1037/0022-0663.94.2.428

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dai, L., Jung, M. M., Postma, M., & Louwerse, M. M. (2022). A systematic review of pedagogical agent research: Similarities, differences and unexplored aspects. *Computers & Education, 190*, 104607. https://doi.org/10.1016/j.compedu.2022.104607

Davis, R. O. (2018). The impact of pedagogical agent gesturing in multimedia learning environments: A meta-analysis. *Educational Research Review, 24*, 193-209. https://doi.org/10.1016/j.edurev.2018.05.002

Davis, R. O., Park, T., & Vincent, J. (2022). A meta-analytic review on embodied pedagogical agent design and testing formats. *Journal of Educational Computing Research, 61*(1), 30-67. https://doi.org/10.1177/07356331221100556

de Melo, C.M., Kim, K., Norouzi, N., Bruder, G., & Welch, G. (2020). Reducing cognitive load and improving warfighter problem solving with intelligent virtual assistants. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.554706

Dincer, S., & Doganay, A. (2017). The effects of multiple-pedagogical agents on learners' academic success, motivation, and cognitive load. *Computers & Education, 111*, 74-100. https://doi.org/10.1016/j.compedu.2017.04.005

Dirkin, K. H., Mishra, P., & Altermatt, E. (2005). All or nothing: Levels of sociability of a pedagogical software agent and its impact on student perceptions and learning. *Journal of Educational Multimedia and Hypermedia, 14*(2), 113-127. https://aace.org/pubs/jemh/

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/bf03193146

Frechette, C., & Moreno, R. (2010). The roles of animated pedagogical agents' presence and nonverbal communication in multimedia learning environments. *Journal of Media Psychology, 22*(2), 61-72. https://doi.org/10.1027/1864-1105/a000009

Freitag, E. T., & Sullivan, H. J. (1995). Matching learner preference to amount of instruction: An alternative form of learner control. *Educational Technology Research and Development, 43*(2), 5-14. https://doi.org/10.1007/BF02300469

Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, *48*(4), 612–618. https://doi.org/10.1109/TE.2005.856149

Grice, H.P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), Syntax and Semantics (Vol. 3, pp. 41-58). Academic Press. https://booksite.elsevier.com/brochures/academicpress/

Gulz, A. (2004). Benefits of virtual characters in computer based learning environments: Claims and evidence. *International Journal of Artificial Intelligence in Education, 14*(3), 313-334. https://iaied.org/

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139-183. https://doi.org/10.1016/S0166-4115(08)62386-9

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? A review of empirical research. *Educational Research Review, 6*, 27-54. https://doi.org/10.1016/j.edurev.2010.07.004

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70.

Hone, K., Axelrod, L., & Parekh, B. (2005). Development and evaluation of an empathic tutoring agent. In *Proceedings of the Joint Symposium on Virtual Social Agents* (pp. 103–108). Retrieved from https://ris.utwente.nl/ws/portalfiles/portal/232516446/Proceedings_AISB_2005.pdf#page=116

Hook, K., Persson, P., & Sjölinder, M. (2000). Evaluating users' experience of a character-enhanced information space. *AI Communications, 13*(3), 195-212. https://api.semanticscholar.org/CorpusID:6126998

Jeffrey, T. E., and Beck, F. J. (1972, November). *Intelligence information from total optical color imagery* (Research Memorandum 72-4). U.S. Army Behavior and Systems Research Laboratory.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). The Guilford Press.

Keller, J. M. (1987). Development and use of the ARCS Model of instructional design. *Journal of Instructional Development, 10*(3), 2-10. https://doi.org/10.1007/BF02905780

Keller, J.M. (2010). Motivational design for learning and performance: The ARCS Model approach. Springer. https://doi.org/10.1007/978-1-4419-1250-3

Keyson, D. K., and Parsons, K. C. (1990). Designing the user interface using rapid prototyping. *Applied Ergonomics, 21*(3), 207-213. https://doi.org/10.1016/0003-6870(90)90004-H

Kim, K., de Melo, C.M., Norouzi, N., Bruder, G., & Welch, G.F. (2020). Reducing task load with an embodied intelligent virtual assistant for improved performance in collaborative decision making. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), 529-538*. https://doi.org/10.1109/VR46266.2020.00074

Kim, Y., Baylor, A.L., & Shen, E. (2007). Pedagogical agents as learning companions: The impact of agent emotion and gender. *Journal of Computer Assisted Learning, 23*(3), 220-234. https://doi.org/10.1111/j.1365-2729.2006.00210.x

Kizilkaya, G., & Askar, P. (2008). The effect of an embedded pedagogical agent on the students' science achievement. *Interactive Technology and Smart Education, 5*(4), 208-216. https://doi.org/10.1108/17415650810930893

Krämer, N. C., Simons, N., & Kopp, S. (2007). The effects of an embodied conversational agent's nonverbal behavior on users' evaluation and behavioral mimicry. In C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, & D. Pelé (Eds.), *Intelligent Virtual Agents* (pp. 238–251). Springer. https://doi.org/10.1007/978-3-540-74997-4_22

Li, W., Wang, F., Mayer, R. E., & Liu, H. (2019). Getting the point: Which kinds of gestures by pedagogical agents improve multimedia learning? *Journal of Educational Psychology, 111*, 1382-1395. https://doi.org/10.1037/edu0000352

Liew, T. W., Tan, S.-M., & Jayothisa, C. (2013). The effects of peer-like and expert-like pedagogical agents on learners' agent perceptions, task-related attitudes, and learning achievement. *Educational Technology and Society, 16*, 275-286. https://www.jstor.org/stable/jeductechsoci.16.4.275

Lin, L., Atkinson, R.K., Christopherson, R.M., Joseph, S.S., & Harrison, C.J. (2013). Animated agents and learning: Does the type of verbal feedback they provide matter? *Computers & Education, 67*, 239-249. http://dx.doi.org/10.1016/j.compedu.2013.04.017

Loorbach, N., Peters, O., Karreman, J., & Steehouder, M. (2015). Validation of the Instructional Materials Motivation Survey (IMMS) in a self-directed instructional setting aimed at working with technology: Validation of the IMMS. *British Journal of Educational Technology, 46*(1), 204-218. https://doi.org/10.1111/bjet.12138

Mayer, R. E. (2014). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (*2nd ed*., pp. 345-370). *Cambridge University Press*. https://doi.org/10.1017/CBO9781139547369

Mayer, R. E. (2023). Research-based principles for designing multimedia instruction. In C. E. Overson, C. M. Hakala, L. L. Kordonowy, & V. A. Benassi (Eds.), *In their own words: What scholars and teachers want you to know about why and how to apply the science of learning in your academic setting* (pp. 143-157). Society for the Teaching of Psychology. https://teachpsych.org/ebooks/itow

Mayer, R. E., & DaPra, C. S. (2012). An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied, 18*(3), 239–252. https://doi.org/10.1037/a0028616

Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology, 100*(2), 380-386. https://doi.org/10.1037/0022-0663.100.2.380

Mayer, R. E., Dow, G. T., & Mayer, S. (2003a). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds. *Journal of Educational Psychology, 4*, 806-813. https://doi.org/10.1037/0022-0663.95.4.806

Mayer, R. E., Fennell, S., Farmer, L., & Campbell, J. (2004). A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology, 96*(2), 389-395. https://doi.org/10.1037/0022-0663.96.2.389

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187-198. https://doi.org/10.1037/0022-0663.93.1.187

Mayer, R.E., Sobko, K., & Mautone, P.D. (2003b). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology, 95*(2), 419-425. https://doi.org/10.1037/0022-0663.95.2.419

Moreno, R., Mayer, R. E., & Lester, J. C. (2000). Life-like pedagogical agents in constructivist multimedia environments: Cognitive consequences of their interaction. In J. Bourdeau & R. Heller (Eds.), *Proceedings of ED-MEDIA 2000. World Conference on Educational Multimedia, Hypermedia & Telecommunications* (pp. 776-781). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/16160/

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction, 19*(2), 177-213. https://doi.org/10.1207/S1532690XCI1902_02

Moreno, R., Reislein, M., & Ozogul, G. (2010). Using virtual peers to guide visual attention during learning: A test of the persona hypothesis. *Journal of Media Psychology, 22*(2), 52-60. https://doi.org/10.1027/1864-1105/a000008

Moundridou, M., & Virvou, M. (2002). Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning, 18*, 253-261. https://doi.org/10.1046/j.0266-4909.2001.00237.x

National Center for Education Statistics. (2023). *Characteristics of public school teachers. Condition of Education.* U.S. Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/clr

Nobles III, W. (2024, July 4). *Morehouse College using AI teaching assistants to power virtual classrooms*. Axios Atlanta. https://www.axios.com/local/atlanta/2024/07/04/morehouse-ai-teaching-assistants

Novick, D., Afravi, M., Camacho, A., Rodriguez, A., & Hinojos, L. (2019). Pedagogical-agent learning companions in a virtual reality educational experience. In P. Zaphiris & A. Ioannou (Eds.), *Learning and collaboration technologies: Ubiquitous and virtual environments for learning and collaboration. hcii 2019. lecture notes in computer science* (*Vol. 11591*, pp. 193-203). Springer. https://doi.org/10.1007/978-3-030-21817-1_15

Ozogul, G., Johnson, A. M., Atkinson, R. K., & Reisslein, M. (2013). Investigating the impact of pedagogical agent gender matching and learner choice on learning outcomes and perceptions. *Computers & Education, 67*, 36-50. https://doi.org/10.1016/j.compedu.2013.02.006

Paas, F., & van Merriënboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*(4), 737-743. https://doi.org/10.1177/001872089303500412

Pastore, R. (2015). Time-compressed instruction: What compression speeds do learners prefer? *International Journal of Instructional Technology and Distance Learning, 12*(6), 3-20. https://www.itdl.org/

Pastore, R., & Ritzhaupt, A.D. (2015). Using time-compression to make multimedia learning more efficient: Current research and practice. *TechTrends, 59*(2), 66-74. https://doi.org/10.1007/s11528-015-0841-2

Plant, E. A., Baylor, A. L., Doerr, C. E., & Rosenberg-Kima, R. B. (2009). Changing middle-school students' attitudes and performance regarding engineering with computer-based social models. *Computers & Education, 53*(2), 209-215. https://doi.org/10.1016/j.compedu.2009.01.013

Prolific. (2024). Prolific (Version used: May 2025) [Online software]. https://www.prolific.com

Reeves, B., & Nass, C. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press.

Reid, R. H. (1968). Grammatical complexity and comprehension of compressed speech. *Journal of Communication, 18*(3), 236-242. https://doi.org/10.1111/j.1460-2466.1968.tb00073.x

Ritzhaupt, A. D., & Barron, A. (2008). Effects of time-compressed narration and representational adjunct images on cued-recall, content recognition, and learner satisfaction. *Journal of Educational Computing Research, 39*(2), 161-184. https://doi.org/10.2190/EC.39.2.d

Ritzhaupt, A.D., Pastore, R., & Davis, R. (2015). Effects of captions and time-compressed video on learner performance and satisfaction. *Computers in Human Behavior, 45*, 222-227. http://dx.doi.org/10.1016/j.chb.2014.12.020

Ryu, J., & Baylor, A. (2005). The psychometric structure of pedagogical agent persona. *Cognition and Learning, 2*, 291-314. https://www.researchgate.net/publication/228356035

Ryu, J., & Ke, F. (2018). Increasing persona effects: Does it matter the voice and appearance of animated pedagogical agent. *Educational Technology International, 19*(1), 61-91. https://educationaltechnology.net/international-journal-educational-technology/

Schneider, S., Krieglstein, F., Beege, M., & Rey, G. D. (2022). The impact of video lecturers' nonverbal communication on learning - An experiment on gestures and facial expressions of pedagogical agents. *Computers & Education, 176*, 1-12. https://doi.org/10.1016/j.compedu.2021.104350

Schroeder, B.L., Fraulini, N.W., Van Buskirk, Wendi L., & Miller, R. (2022). Assessing the social agency of pedagogical agents in adaptive training systems. In R. A. Sottilare & J. Schwarz (Eds.), *Adaptive Instructional Systems. HCII 2022. Lecture Notes in Computer Science* (*Vol. 13332*, pp. 302-313). Springer. https://doi.org/10.1007/978-3-031-05887-5_21

Schroeder, N. L. (2017). The influence of a pedagogical agent on learners' cognitive load. *Journal of Educational Technology & Society, 20*(4), 138-147. https://www.jstor.org/stable/26229212

Schroeder, N. L., & Adesope, O. O. (2014). A systematic review of pedagogical agents' persona, motivation, and cognitive load implications for learners. *Journal of Research on Technology in Education, 46*(3), 229-251. https://doi.org/10.1080/15391523.2014.888265

Schroeder, N. L., & Adesope, O. O. (2015). Impacts of pedagogical agent gender in an accessible learning environment. *Educational Technology & Society, 18*(4), 401-411. https://www.jstor.org/stable/jeductechsoci.18.4.401

Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research, 49*(1), 1-39. https://doi.org/10.2190/EC.49.1.a

Schroeder, N. L., Yang, F., Banerjee, T., Romine, W. L., & Craig, S. D. (2018). The influence of learners' perceptions of virtual humans on learning transfer. *Computers & Education, 126*, 170-182. https://doi.org/10.1016/j.compedu.2018.07.005

Schroeder, N.L., Romine, W.L., & Craig, S.D. (2017). Measuring pedagogical agent persona and the influence of agent persona on learning. *Computers & Education, 109*, 176-186. http://dx.doi.org/10.1016/j.compedu.2017.02.015

Shah, S.J. (2020). *Measuring the additive effects of multimedia social cue principles on learners' cognitive load, emotions, attitude, and learning* (Publication No. 27957128) [Doctoral dissertation, Old Dominion University]. *ProQuest Dissertations & Theses Global*. https://www.proquest.com/dissertations-theses/measuring-additive-effects-multimedia-social-cue/docview/2421423933/se-2

Siegle, R. F., Schroeder, N. L., Lane, H. C., & Craig, S. D. (2023). Twenty-five years of learning with pedagogical agents: History, barriers, and opportunities. *TechTrends, 67*, 851-864. https://doi.org/10.1007/s11528-023-00869-3

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123-138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J. (2011). Cognitive Load Theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of Learning and Motivation* (*Vol. 55*, pp. 37-76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296. https://psycnet.apa.org/doi/10.1023/A:1022193728205

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. Educational Psychology Review, 31(2), 261-292. https://doi.org/10.1007/s10648-019-09465-5

Tognazzini, B. (1992, September). Violence in software. *Sun World*, 105-107.

van der Meij, H., van der Meij, J., & Harmsen, R. (2015). Animated pedagogical agents effects on enhancing student motivation and learning in a science inquiry learning environment. *Educational Technology Research and Development, 63*(3), 381-403. https://doi.org/10.1007/s11423-015-9378-5

van Mulken, S., André, E., & Müller, J. (1998). The persona effect: How substantial is it? In H. Johnson, L. Nigay, L., & C. Roast (Eds.) *People and Computers XIII*. (pp. 53-66). Springer. https://doi.org/10.1007/978-1-4471-3605-7_4

Yung, H. I. (2009). Effects of animated agent with instructional strategies in facilitating student achievement of different educational objectives in multimedia learning. *Journal of Educational Multimedia and Hypermedia, 18*(4), 453-466. https://www.learntechlib.org/primary/p/26124/

Yung, H. I., & Paas, F. (2015). Effects of cueing by a pedagogical agent in an instructional animation: A Cognitive load approach. *Educational Technology & Society, 18*(3), 153-160. https://www.jstor.org/stable/jeductechsoci.18.3.153

Yusoff, M. D. M., Yusof, A. M., Jamaludin, N. S. A., & Ruzzaman, M. S. K. (2023). Cognitive load in pedagogical agent research. *Proceedings of International Academic Forum on Creative Contents [IAFCC]*.