Motor and Delay-based Information Access Costs and their Impact on Behaviour and Learning Outcomes in a VR-based Category Learning Experiment

by

Robin C. A. Barrett

Bachelor of Arts (Psychology), Simon Fraser University, 2018

Diploma of Arts (Psychology), Langara College, 2016

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Arts

in the
Department of Psychology
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY
Spring 2024

Declaration of Committee

Name:	**Robin C. A. Barrett**
Degree:	**Master of Arts (Psychology)**
Title:	**Motor and Delay-based Information Access Costs and their Impact on Behaviour and Learning Outcomes in a VR-based Category Learning Experiment**
Committee:	**Chair: Thomas M. Spalek** Professor, Psychology **Mark R. Blair** Supervisor Associate Professor, Psychology **Steve DiPaola** Committee Member Professor, Interactive Arts and Technology **Bernhard Riecke** Examiner Professor, Interactive Arts and Technology

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Abstract

Immersive virtual reality offers category learning researchers the ability to present a wider variety of interactable and maneuverable 3D stimuli which more closely resemble scenarios encountered in the real world. In this study, information access costs, either requiring increased motor movement or time-based delays, are implemented into a VR-based category learning experiment to see whether the predicted impacts of information access costs on learning and attention-related behaviour are contingent on the type and intensity of the cost. Possible predictors at the individual level that might explain differences in learning outcomes between participants are also investigated. Delay costs impacted attention-related behaviours more than motor costs, causing participants to be more economical in their use of attentional resources. Frequency of video game play had a small impact on learning outcomes. This work concludes with a discussion of the limitations, future directions, and possible applications of the results.

Keywords: Categorization; Virtual Reality; Access Cost; Individual Differences; Attention

Dedication

To my loving Fiancé, Michelle, for her constant encouragement and understanding as we both pursue our aspirations together. As we enter the next chapter of our lives together, I am excited to further deepen our bond while we learn to navigate whatever challenges lie ahead of us in our personal and professional journeys. I love you.

In loving memory of my Beloved Friend, Cee. Our love for one another has meant so much to me over this past year. Although they're gone, the passion she always brought to learning new skills will continue to motivate me in reaching new heights as an educator and as a learner. I know you'd be proud of me.

And to my friends and family, who have been there for me throughout this project and continue to support me in my studies and personal growth. Thank you.

Acknowledgements

There are so many people whose contributions and thoughts have influenced and enabled the completion of this work. First and foremost is my supervisor, Dr. Mark Blair, who believed in me from the very beginning and gave me the freedom to pursue my own research goals while providing critical guidance at all parts of the process. We have been collaborating on projects for nearly a decade now, and this document is a testament to the growth in our relationship over those years. This project is also heavily indebted to the hard work and assistance of several members of the SFU Cognitive Science Lab, with a special mention to Justin O'Camb, Dr. Caitlyn McColeman, Dr. Joseph Thompson, Jordan Barnes, Rajan Hayre, Cal Woodruf, and all the rest of my co-authors, whose hard work and guidance over the years has been instrumental to my growth as a developer, as a researcher, and as a leader.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

The process of categorization is commonly encountered throughout daily life. Learning our left hand from our right, the letter ‘b’ from the letter ‘d’, and several other mundane tasks depend on categorization. Even navigating our environment requires being able to categorize the sorts of mediums we can traverse from the ones we cannot. Although differentiating a wall from an empty doorway seems trivial, the processes involved in doing this are in fact complex acts of categorization. Because the study of categorization provides insights into such a wide variety of cognitive phenomena, it is a rich field for researchers to engage with a variety of topics.

1.1. Category Learning Research

Categorization research has been an active field in psychological research for over a century (Ashby & Maddox, 2005; Fisher, 1916; Hughes & Thomas, 2021; Hull, 1920; Medin & Schaffer, 1978; Trabasso & Bower, 1964), but the study of attentional processes in use during these tasks is still relatively new, with less than 20 years of research actually measuring the eye movements of participants while learning to categorize stimuli (Rehder & Hoffman, 2005a, 2005b). Although the definition of attention as being selectivity in perception is not limited to the use of eye movements, several studies have found that eye movements are a strong indicator of overt visual attention (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Kowler et al., 1995). While covert attention is possible, it is not the default state, and while attention can be shifted without an eye movement, it is not possible to make an eye movement without also moving attention to that location at the same time (Rayner et al., 1978; Shepherd et al., 1986). Studies like these allow for researchers to generate useful insights into how our attention changes as a function of learning, as well as how changes in the environment can influence our learning related behaviours. Generally, these studies have observed fairly reliable trends in attention related behaviours over the course of learning, with the transition from novice to expert being fairly predictable across experiments (Ashby & Maddox, 2005; Kruschke, 1992; McColeman et al., 2014). Specifically, in addition to getting faster and more accurate in their responses (Rehder & Hoffman, 2005a, 2005b), learners tend to spend less time examining irrelevant information (Rehder & Hoffman, 2005a), make fewer eye movements to irrelevant information (McColeman et al., 2014), and rely less on corrective feedback as learning progresses (Dolguikh et al., 2021; Watson & Blair, 2008). With eye movements being closely tied to decision making (Orquin & Mueller Loose, 2013), explaining and predicting changes in eye movements can, consequently, be useful in differentiating experts from novices in their fields and help to better understand how people at different levels of skill make decisions during complex tasks.

1.1.1. Information Access Costs

Previous categorization research has observed that making information difficult to access can sometimes encourage more efficient learning related behaviours. For example, one experiment reported in Meier and Blair (2013) covered each stimulus feature with masks which were only removed after a participant had fixated their eyes on the feature for a certain duration. Participants faced with longer delays were quicker to determine which features offered the highest utility in determining the stimulus category, avoiding less useful features. Rajsic et al. (2018) found similar results, as participants tended to use more efficient visual search strategies when the cost of accessing information was increased by using gaze-contingent feature masks. In a follow-up experiment, Rajsic et al. (2018) required participants to make mouse movements to the stimulus features to reveal information, resulting in even more efficient patterns of information access behaviours. The authors argued that this change in efficiency was due to the increased number of muscle movements involved in moving the mouse compared to moving one's eyes. Lastly, Morgan et al. (2010) combined mouse movements with a delay cost and found that longer information access costs made learners more resilient to interruptions during the learning task. Taken together, these findings demonstrate that information access costs often encourage more efficient allocation of attentional resources.

Not every study has reported benefits; Yang et al. (2013; 2015) both used a 2x2 design to explore how information access costs and test expectations might interact during a junior doctor's patient assessments. In both experiments, Junior medical residents were given 4 different case files to study and assess. The doctors were either told that their assessments would remain confidential or that they would be evaluated by another professional on their choices. Each doctor was tested on their recognition and recall of the case and their confidence in their answers. Afterwards, a second test was administered where they were told they could refer to the patient record to refresh their memory as needed. Information access cost was manipulated here by the positioning of the computer with the patient record on it. Either the computer was immediately available next to the participant, or the computer was at another desk 5 meters away, requiring the participant to physically walk across the room to access the information. When the computer was farther away, participants referred to the patient record less than when it was immediately available and answered the test questions much more quickly. However, having a higher information access cost also resulted in a greater number of errors in assessment overall. These studies suggest there is a point at which increasing the cost of accessing information by too much may have an inverse effect on learning. As this study used real doctors being asked to provide medical examinations, this shows that the impact of information access costs could have drastic effects on health-related outcomes. As such, investigations are needed to better understand what kinds and intensities of information access cost will lead to the best outcomes during training and practice.

In reviewing these studies, two distinct forms of information access cost seem to be apparent: motor costs and delay costs. Motor costs involve situations where a physical action is required to access information, with studies using low motor costs such as eye movements (Meier & Blair, 2013), medium costs like hand movements with a computer mouse (Rajsic et al., 2018), and high costs such as walking across a room (Yang et al., 2013, 2015). Delay costs involve some kind of timed delay, whether it be a loading screen on a webpage, or a 3-second stimulus mask (Meier & Blair, 2013). Waiting for a delay and activating muscle movements are qualitatively different activities, yet attention and learning research has not distinguished them explicitly to determine if implementing these different kinds of costs would be more or less likely to produce the increased learning efficiency observed. Certainly, any additional movement will likely require additional time to perform, and so the two costs might sometimes be conflated in terms of time spent accessing information. However, these costs are strongly differentiated in how motor costs require additional muscle movement to navigate, whereas delay costs are navigated simply by waiting. Studies have also neglected to explore the point at which the intensity of these costs becomes a detriment to learning instead of a benefit.

1.1.2. Non-Learners in Category Learning Experiments

While category learning experiments attempt to describe how people make category judgements, a common feature of these kinds of experiments is that many participants are unable to reach high levels of accuracy over the course of the study. Participants that do not reach a certain criterion point are typically excluded from the final analysis and labelled as “non-learners”. This can result in studies being underpowered, as it is common to throw out close to 25-50% of sample data for not reaching the learning criterion (Feldman, 2021; Medin & Schaffer, 1978; Pérez-Gay Juárez et al., 2017). Knowing they may lose up to half their data, some researchers may opt to run twice as many participants to offset the expected exclusion rate. This impact on wasted time and resources is undesirable enough on its own, but excluding so many participants from analysis brings forward a key theoretical implication for the field in that models of learning based on these studies are based only on the subset of participants who managed to reach the criterion, limiting the generalizability of their findings. Although some researchers have argued that the category structures are simply difficult to learn (Medin & Schaffer, 1978; D. J. Smith & Minda, 2000), this ignores a vast multitude of possible explanations as to why a particular individual struggle with category structure while other participants have no issue with it. This explanation also seems to overlook the fact that participants who do reach the criterion points typically reach peak accuracy quite quickly (Barrett et al., 2022; McColeman et al., 2014). If the task were difficult, one would not expect such fast-rising learning curves. Of course, some participants may simply not be trying their best, while others misunderstand the task completely, but it is unlikely that this explanation accounts for all the so-called “non-learners” in these studies.

Some researchers have used non-learner rates as a way to study strategy selection among participants. In work by Mathews et al. (1984), participants engaged in a simplified form of the Bouthilet keyword matching task. In this task, participants are shown a keyword such as “heroism”, and two words to pick from such as “moss” and “help”. In one condition the correct word would contain only letters that were also present in the key word (“moss” in this example). In another condition, the correct word contained none of the letters from the keyword (“help” in this example), and in the last condition, the rule alternated between the “all” and “none” conditions after every two errors made by the participant. After reaching a criterion of 10 correct trials in a row or going through all 200 trials, participants were given a number of test trials with no feedback, and then asked how they decided on what word to select next, with their response being used to indicate whether the participant had figured out the rule or not. Based on their descriptions, participants were separated into three groups: ones that had reached the criterion and could state the rule, ones that had reached criterion but could not state the rule, and those that did not reach criterion. Their findings suggest that participants who had reached criterion but could not state the rule tended to use memory-based strategies, simply memorizing the previously presented stimuli. This strategy often worked to achieve criterion but did not extend perfectly to novel stimuli used in the post-criterion training. This strategy was especially ineffective in the condition with a fluctuating rule, with almost half the participants being non-learners in this group while the proportion of learners who could state the rule remained the same across the conditions (Mathews et al., 1984). Rehder & Hoffman (2005a) also reported that different participants appeared to employ different strategies in their learning, and that the choice to use a different strategy could severely affect the rate of learning depending on the complexity of the category structure used. While these studies describe differences in how learners perform during the experiment, none of the literature reviewed attempted to understand why these differences in strategy use exist. Examining individual differences between learners and non-learners is therefore necessary to determine if the high levels of attrition in category learning experiments can perhaps be explained by different traits, attributes, or beliefs each participant may bring to the study.

1.1.3. Possible Individual Factors Influencing Learning Outcomes in Categorization Research

Although the question of non-learner rates in category learning studies has not yet been explicitly addressed, there are several individual differences that could plausibly act as predictive factors in this context due to their associations with learning outcomes in other fields of research. These include a participants age, sex, their working memory, self-efficacy, whether they have ADHD, and their tendency to adopt a growth mindset when faced with difficulties.

Self-Efficacy is the idea that achieving desired outcomes in learning situations are, in part, contingent on the learner’s beliefs and expectations about their ability to achieve that outcome (Bandura, 1977, 1986). Someone who believes themselves to be capable of achieving a certain goal will be more likely to take actions which support the attainment of that goal. Applied to learning scenarios, it has been found that higher self-efficacy results in better learning outcomes (Makransky & Petersen, 2021; Pajares & Schunk, 2002; Usher & Pajares, 2008), especially in cases where learning is self-regulated (Panadero et al., 2017; Wang et al., 2013). In the context of category learning, this suggests that if a learner feels that they are incapable of learning the categories, then they will be more likely to resort to random guessing, whereas learners with higher self-efficacy ought to persist in applying different strategies until they find the correct one over the course of the experiment.

Related to the idea of Self Efficacy is the work of Dr. Carol Dweck on the Growth Mindset (Blackwell et al., 2007; Dweck & Leggett, 1988) which has been explicitly connected to learning outcomes at a classroom level. The central claim of this theory is that there are two main implicit mindsets which guide task-oriented behaviour during skill acquisition in a given domain. Someone with a fixed mindset believes that their ability to perform certain tasks is non-alterable and unchanging, while those with growth mindsets believe that their abilities are not innate, and that with practice, they can improve in that domain. A meta-analysis examining this theory (Burnette et al., 2023) finds that the presence of a growth mindset has been highly predictive of how people engage with training tasks and subsequently, how well they perform on final tests. They also found that this effect was stronger when learners were faced with feedback pointing out their failures. In category learning, this variable may influence learning outcomes as the stimulus sets are novel to participants, and so it is expected that a lot of mistakes will be made during earlier trials. Participants with fixed mindsets may interpret these initial failures as being reflective of a personal inability to learn the categories and give up, while those with a growth mindset may interpret the situation positively, putting more effort into the learning of stimulus categories.

Smith and Minda (2000), using a 5-4 category structure like the one used in Medin & Schaffer (1978), point out that when the number of possible exemplars in a category is small, or when the categories are poorly structured, learners will avoid specific rule-based strategies in favour of just memorizing the features of each exemplar. Blair and Homa (2003), also using the 5-4 category structure, likewise found that when the categories only had a few exemplars each, participants often relied on memorization strategies as the small number of exemplars made memorization just as efficient as learning the rule. Additionally, studies which directly asked participants to state their categorization strategy have found that memorization is often used in combination with other strategies (Gouravajhala et al., 2020; Wahlheim et al., 2016). Subsequently, participants found to have a larger working memory capacity may be expected to have a better likelihood of achieving higher levels of accuracy during a category learning task.

Much work has been done to investigate the impact of aging on categorization (Bowman et al., 2022, 2023; Filoteo & Maddox, 2004; Gouravajhala et al., 2020; Wahlheim et al., 2016). Participants in these studies tend to be either in their 20's and 70's, reflecting that these researchers are mainly concerned with comparing younger adults to much older adults. Very few studies have explored whether there are any differences that appear at younger age groups (Casadevante et al., 2019; Reetzke et al., 2016). Reetzke et al. (2016), using an auditory category learning task, found that adults aged 20-23 years old outperformed children aged 13-19 years old both in terms of rate of learning and overall accuracy in category learning. The uneven age interval sizes make it difficult to know if this difference would apply to a university aged population of 18-25 years old however, and so it is worth exploring if this effect would be observed while treating age as a continuous variable instead. Work by Thompson et al. (2014) suggests that cognitive decline may begin at 24, but it is unclear whether this decline, observed in a complex esports environment, would apply to much simpler category learning tasks.

Lastly, Attention deficit disorders impact approximately 2-7% of adults in Canada (Espinet et al., 2022), and can have small to large impacts on decision-making, learning, inhibitory control, self-esteem, and working memory, in addition to the typical attention-related impairments that define the disorder (For a comprehensive review, see Faraone et al., 2021). In category learning tasks, participants with ADHD have been observed to have slower learning rates and spend more time engaging with irrelevant stimulus features than non-ADHD controls (Huang-Pollock et al., 2014). These effects are exacerbated by the presence of corrective feedback as, when using trial-by-trial corrective feedback, participants with ADHD reliably underperformed neurotypical participants on various category learning tasks (Gabay & Goldfarb, 2017). Knowing that around 1 in every twenty participants are likely to have ADHD, and that this population has been found to underperform on category learning tasks in the past, it is worthwhile to explore whether prevalence of ADHD, so rarely recorded in category learning research, may account for some of the non-learner rates across these experiments.

1.2. Virtual Reality as a new Medium for Research

Immersive Virtual Reality (VR) has become increasingly affordable and high fidelity in the last decade, resulting in this technology getting a lot of attention by both industry and academic professionals alike who are interested in harnessing the potential for this technology to improve workplace training and educational programming. Many companies have already started using VR to train employees (Hou et al., 2017; Likens & Mower, n.d.), and educators at all levels have been beginning to try and leverage the presumed impacts of this new technology in their classrooms ("Fisk University, HTC VIVE, T-Mobile and VictoryXR Launch 5G-Powered VR Human Cadaver Lab," 2021; Virtual Reality for Schools, n.d.; Makransky et al., 2021; Meyer et al., 2019). While the enthusiasm is certainly worth celebrating, there is much debate about the measurable impact this technology has had so far on actual learning outcomes. Reviews and meta-analyses have been largely positive towards virtual reality (Angel-Urdinola et al., 2021; Di Natale et al., 2020; Jensen & Konradsen, 2017; Muller Queiroz et al., 2018), but even within these reviews, reported effect sizes for learning outcomes are highly variable. In Angel-Urdinola et al. (2020), more than half of the studies considered in their meta-analysis reported a neutral-positive finding; that is to say, VR had no advantage over traditional non-immersive modalities such as pen-and-paper or 2D Desktop screens. Similarly, Barrett et al. (2022) found only a few meaningful differences in performance outcomes between participants learning to categorize stimuli, regardless of whether they were working on a 2D flat-screen or rotating the stimulus with their arms in VR. Work by other labs has corroborated these findings in other learning tasks, finding no substantial differences between immersive VR and video-based teaching (Oser & Fraser, 2015; Parong & Mayer, 2018).

Rather than discourage the use of VR in an educational setting, these findings only reinforce the need for research that identifies the key design factors that contribute to improved learning outcomes. In Richard Clark's (1994) commentary, boldly titled "Media will never influence learning outcomes", he points to how studies exploring the impact of computer-based instruction on learning outcomes often failed to find any meaningful impact of the technology tested. Furthermore, those studies that did find positive impacts were often confounded by the fact that the computer-based instruction was qualitatively different from the control conditions. Clark points out that not all teachers use the same pedagogical approach within traditional classroom environments. Likewise, computer programs are each embedded with their own instructional design process and pedagogical foundation. It is reductionist to say that computers have an impact on learning outcomes when it is the instructional methods embedded in those programs that make the difference. Yet, it is rare that studies on the use of computer-based instructional methods report their theoretical underpinnings (Mikropoulos & Natsis, 2011). Some programs and modalities may make it easier to implement certain teaching strategies, but again, it is the instructional strategy that matters, and not the technology itself (Clark, 1994). The same argument applies to designing VR programs to deliver instructional materials where critical reviews have also found that well-designed content is the primary driver of learning outcomes, and not the technology alone (Di Natale et al., 2020; Jensen & Konradsen, 2017; Radianti et al., 2020). Research in this area should instead focus on these design factors, and work to identify which of these provide the most meaningful impacts on student learning.

Some work has already been done to identify which design affordances, unique to VR would be worthwhile exploring (Radianti et al., 2020), with recent reviews pointing also to design factors that have already been shown to influence learning in VR more than others (Jensen & Konradsen, 2017; Pellas et al., 2020). Makransky and colleagues, cited commonly in these reviews, have conducted studies showing how learning outcomes within the VR modality can be improved through the integration of evidence-based pedagogical practices such as those recommended by Mayer (2014). Among these practices, they have explored techniques such as generative learning (Klingenberg et al., 2020; Petersen et al., 2023), enactment (Andreasen et al., 2019), retrieval practice (Parong & Mayer, 2018), and pre-training (Petersen et al., 2020). These researchers are optimistic about the future of immersive technology in the classroom but emphasize that the question must shift from simplistic inquiries as “Is learning possible in VR?” to more nuanced problems like “When is VR best suited for specific learning outcomes?” and “What features must be present for a VR-based training program to be most effective?”. The findings from these researchers point to several guidelines for designing VR learning environments, but their work has so far been restricted to testing the recommendations that come from educational research. In this work, we opt to focus on exploring the findings from the category learning research discussed above as a source of additional guidelines for the design of virtual reality learning environments.

1.2.1. Prior Use of VR in Cognitive Psychology

Within academic research, VR technology also offers up the opportunity to conduct increasingly ecologically valid research through convincing simulated environments, while allowing the researcher to maintain experimental control over the stimuli presented. Much research has already been done to transpose standard research methods into VR, with the hopes that findings from the lab would be generalizable to the virtual world. Some examples of this include replicating Milgram's obedience to authority experiment (Slater et al., 2006), the Trolley Problem (McDonald et al., 2017), and tests for a wide variety of cognitive phenomena (Corriveau Lecavalier et al., 2018; Li et al., 2020; S. A. Smith, 2019; Soranzo et al., 2013; Soranzo & Wilson, 2014). Within neurological research, researchers are also optimistic about using VR to present more convincing stimuli than are currently possible with the 2D screens used in most brain-imaging experiments (Bohil et al., 2011; Kourtesis et al., 2020; Parsons, 2015). Many of these studies report that participants in VR respond similarly to stimuli while immersed in the virtual environment as when using a 2D screen. This is desirable because it means that predictions of behavioural patterns emerging from studies of participants interacting with flat screens are roughly generalizable to how people interact with stimuli in a 3D environment as well. However, there are still a large variety of stimuli and scenarios that cannot be represented well on flat screens that would benefit from a more immersive presentation format, such as social situations and scenarios involving more complex motor movements.

In categorization research, stimulus have typically been simple flat objects, presented on a flat screen or on paper. While exceptions to this exist with the use of 3D stimuli (Barnhart et al., 2018; Gauthier & Tarr, 1997; Hammer et al., 2012), even these exceptions present the stimuli to participants on a flat computer screen, with no option to rotate the object. Transfer of learning from one task to another is notoriously difficult to predict (Blume et al., 2010; Thompson et al., 2023), with recent research finding that switching from 2D to 3D stimuli can sometimes lead to different learning outcomes on tasks designed to measure core cognitive abilities (Alvarez & Cavanagh, 2004; Nejati, 2021; Neuburger et al., 2015). These findings give us reason to infer that models of categorization which can successfully predict information access behaviours for 3D stimuli will be better suited for generalization to the real world than those which only successfully predict information access behaviours for 2D stimuli. Consequently, conducting categorization research in VR is desirable to ensure that findings from prior categorization studies can be scaled up to 3D objects and more immersive scenarios.

1.2.2. Category-VR: Transposing Category Learning Experiments into VR

Prior work by our lab (Barrett et al., 2022) transposed a category learning task into both VR, and into a desktop-based implementation using 3D stimuli. In both conditions, participants rotated a 3D virtual cube, either while using hand-controllers in immersive VR, or with a standard gaming controller on a desktop computer. Data from these conditions was compared to results from a conceptually equivalent experiment using a 2D categorization task on a flat computer screen where only eye movements were necessary to access stimulus features. Trends for measures of Accuracy, Optimization, Fixation Durations and Counts, as well as response times all followed the anticipated qualitative changes as predicted by general trends from previous category learning research (McColeman et al., 2014, 2020). Some important differences did arise however, as although the direction of change was identical across all groups, the degree of change was different. Specifically, it was thought that because making arm movements to access stimulus features takes more metabolic effort, the associated information access cost would have elicited the types of learning patterns seen in Meier and Blair (2012). However, by the end of the trials, participants in the VR group still made more fixations per trial than participants in other conditions with less information access cost. In many ways, this was contrary to expectations, as higher costs had been previously associated with improvements to learning. Consequently, it is unclear under what conditions the benefits of information access costs may appear.

1.2.3. Possible Individual Factors Influencing Learning in VR Research

Spatial ability, as described by Carroll (1993), is defined as “an ability in manipulating visual patterns, as indicated by the level of difficulty and complexity in visual stimulus material that can be handled successfully without regard to the speed of task solution" (p.362). Because VR programs often involve the manipulation of 3D stimuli with one's hands, it is unsurprising that spatial ability should be a relevant individual factor here that could determine performance. In line with this idea, there is an ongoing debate between two hypotheses introduced by Mayer and Sims (1994), predicting who would benefit more from digital learning environments: people with lower spatial ability (referred to as the *ability-as-compensator* hypothesis), or whether people with higher spatial ability should receive the highest benefit from using digital environments (referred to as the *ability-as-enhancer* hypothesis). In support of the ability-as-compensator hypothesis, some studies have found that learners with low spatial ability benefit more from operating 3D learning materials than students with high spatial ability (Höffler & Leutner, 2011; Lee & Wong, 2014; Weng et al., 2019). These studies argue that the affordances of the 3D environment allow the learner to manipulate the stimulus in the same way that a learner with higher spatial ability would manipulate the object using mental visualization techniques. In contrast, studies supporting the ability-as-enhancer hypothesis have also shown that people with higher spatial ability do better in learning situations with animated features because of their superior ability to make better use of these visual cues (Chikha et al., 2021; Duffy et al., 2018; Epler-Ruths et al., 2020). Although these hypotheses have been treated as mutually exclusive, some researchers have begun to explore the interaction effects that may allow for both approaches to be applicable depending on the situation (Gittinger & Wiesche, 2023; Kühl et al., 2022). Applied to VR Learning simulations and information access costs, it is unclear if people with low spatial ability will be better catered to or if their low spatial ability will hinder their ability to learn to identify 3D stimuli. Additionally, it is unknown if these learners will be disproportionally impacted by information access costs, as while the ability to freely rotate a stimulus to see its features might be helpful to some degree, higher information access costs might nullify any benefits this group may get from VR-based learning.

1.3. The Current Project

Using the same category learning stimuli as Barrett et al. (2022), the current project aims to explore whether the way information costs are implemented can change how people learn to categorize stimuli, and whether any individual differences between participants might predict performance on this task.

There was an information access cost present in Barrett et al. (2022), but this cost took the form of a small "motor cost" rather than a "delay cost". While needing to rotate the stimulus cube did take more time than moving one's eyes, with fixations in VR lasting a full second longer on average compared to the eyetracking condition, no mask was present on the stimulus features, so the participants were only delayed by their own physical ability to rotate the cube. In the VR-based version of the experiment, the predicted finding that information access cost would have an impact on learning behaviours was not observed, as learning and attention-related behaviours in all conditions mostly followed the same patterns, despite there being different motor costs to access the stimulus features. It is unknown if the cause for these conflicting findings has to do with the use of VR, or the type of information access cost incurred: delay cost vs motor cost.

To investigate the discrepancy between Barrett et al (2022) and other papers, the current project reused the stimulus from Barrett et al. (2022) using only the VR-based implementation to focus on the type of information access cost used in the experiment. By separating the impact of motor costs from the impact of increased delay costs, their potentially unique contributions to information access behaviours and learning outcomes can be more directly observed. The current project aims to resolve two main issues in this line of inquiry. Using an immersive VR-based category learning experiment, delay and motor costs are tested to try and replicate the previously observed benefits of information access costs. Uncovering these effects may help to explain why Barrett et al. (2022) and others did not observe these benefits. Lastly, I explore possible reasons as to why some participants do better on categorization tasks than others. Exploring whether certain individual differences predict rates of learning in this type of experiment may have important implications for designing and implementing more accessible learning environments.

As such, this project has two main research questions:

1. How do different types and intensities of information access cost influence changes in learning outcomes and attention-related behaviours inside a VR learning environment?

2. Which individual differences, if any, predict the likelihood of a participant being able to successfully learn to categorize stimuli during training in a VR environment?

Chapter 2. Methods

2.1. Participants

130 undergraduate students were recruited from Simon Fraser University in Canada. Each participant received course credit for their participation. The study was approved by the Simon Fraser University Office of Research Ethics and was deemed to be minimal risk (Study Number: 30000750). All participants gave written informed consent prior to completing the survey. Each participant was given instruction on how to withdraw from the experiment if they experienced any discomfort or sickness at any point during the experiment.

2.2. Exclusions

2 participants withdrew from the study, both giving no reason for their withdrawal. 1 participant withdrew after just a few trials citing that they felt dizzy. 4 participants reported mild discomfort at the end of the experiment, though none of these chose to withdraw. Since their data does not appear to be impacted by this experience, their data was retained for analysis. All the same, this indicates that roughly one out of every twenty participants experienced some symptoms of VR related sickness despite it being a relatively low-intensity, seated experience. Moreover, not all participants completed all 96 trials of the experiment. Any participant who had less than 2 bins of trials (24 trials per bin) was excluded from all analyses. In total, 11 participants were excluded for this reason, resulting in a sample of 119 viable participants (30 males; 87 females; 1 intersex; 1 prefer not to say) with sufficient data to include in the analyses. Ages of these participants ranged from 17 to 32. 61 reported that they wore glasses or corrective lenses, though very few wore their glasses during the VR program.

For the first research question, as in Barrett et al. (2022), participants who did not reach a threshold of 24 trials in a row were excluded from the analysis of learning and attention-related behaviours, excluding an additional 72 participants. After accounting for these exclusions, the final sample size for this research question was 47 participants. Research question two was designed to investigate possible reasons for most participants not reaching the criterion point, and so analyses of this research question will include all 119 viable participants.

2.3. Design

The experiment followed a 2x2 design with four conditions (Table 2.1.). Conditions varied to what degree participants would experience delay and/or motor costs when attempting to access the stimulus features on each side of the cube. A video showcasing each of the four conditions can be found at the following link: https://summit.sfu.ca/item/38111

Table 2.1. The 2x2 structure of the experiment, demonstrating all possible conditions.

	Motor Costs	
Delay Costs	Low Motor Cost 1 Second Delay Cost	High Motor Cost 1 Second Delay Cost
	Low Motor Cost 5 Second Delay Cost	High Motor Cost 5 Second Delay Cost

To increase the motor cost while in VR, the rotational-drag of the virtual object was adjusted. In the low motor cost condition, participants could use their hands to quickly rotate the cube and see its different sides, while in the high motor cost group, the cube moved much more slowly, forcing the participant to move their arms much more slowly and sometimes make multiple adjustments to bring each side of the cube into view. This slowness would require the participant to spend more effort moving their arms through the air, activating more of their muscles over the course of the 100 trials.

To increase delay costs, timed masks similar to the ones used in previous research (McColeman et al., 2014; Meier & Blair, 2013) were used to cover stimulus features. When the participant rotated the cube to view a particular side, they would have to stare at the mask for a set amount of time before the feature beneath was made visible to them. Each mask used a visual countdown to indicate how much longer a participant would have to wait for the information to be visible. In the present project, a one-second delay was used for the low delay cost group, and a five-second delay was used in the high delay cost group.

2.4. Stimuli and Materials

The stimuli in this experiment were nearly identical to the stimuli used in Barrett et al. (2022), with the only modifications being the addition of the delay and motor costs described above. Consequently, this stimulus set is functionally identical to the stimuli used in previous research as well (Meier & Blair, 2012). See Figure 2.1. for an example of how to identify a category under this system.

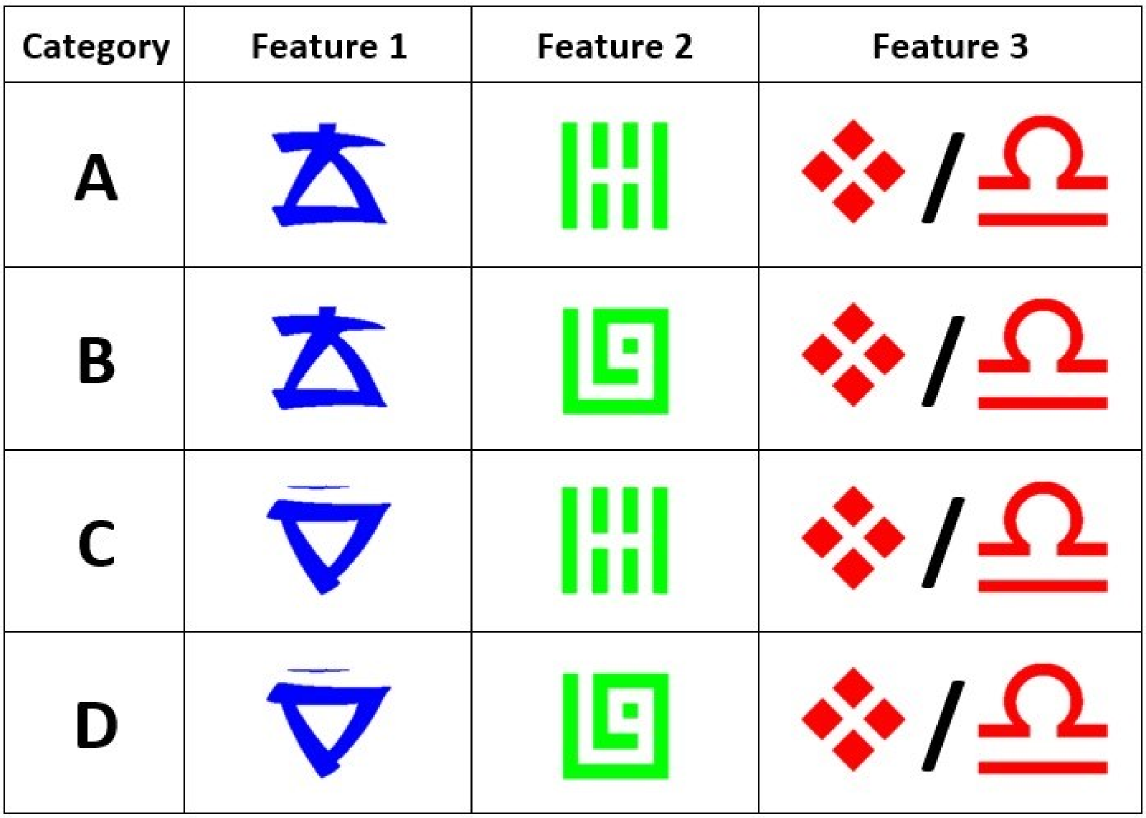


Figure 2.1. An example of the category structure used in Barrett et al. (2022) and the current experiment. Note how only two features are necessary to correctly identify the category group, while the third feature provides no relevant information.

Using an immersive virtual reality program built using Unity (Unity Technologies, 2020), participants in all conditions were shown 3D cubes belonging to one of four categories which could be identified by the markings on the sides of the cube. An example stimulus can be seen in Figure 2.2. below. The sides of each cube were indented with deep wells, so that only one marking could be visible at one time. To view the feature in each well, the cube would have to be rotated to at least 56 degrees relative to the plane of the participant’s field of view as measured from the centre of the VR headset. Each cube had three unique markings that could be used to identify what category it belonged to. As can be seen in Figure 2.2., on initial presentation, the cube was positioned on its corner so that no side was visible to the participants at the beginning of each trial. Markings were repeated on the opposite side of the cube to fill up all 6 sides of the cube. Referring back to Figure 2.1., the 3 markings were binary coded to only appear in one of two possible states. This allows for a total of 8 unique cubes for each participant.

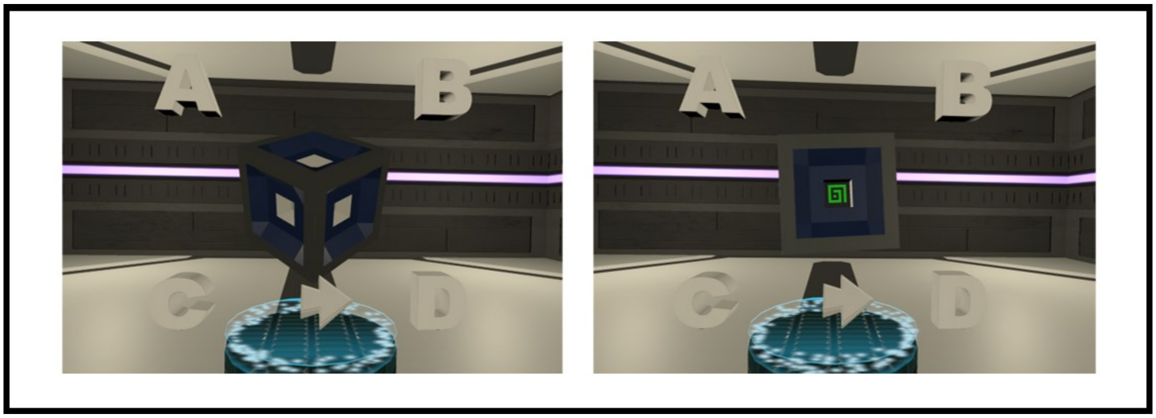


Figure 2.2. **Screenshots from the experiment showing the stimulus cube. On the left, the trial begins with the cube positioned with its corner facing the participant so that no side is immediately visible to them. On the right, the cube has been rotated so that the participant can see the feature on that side.**

After examining the cube, participants reached out to one of four choice buttons appearing as floating letters around the cube, and feedback was presented by changing the colour of the letters to indicate the correct choice (Figure 2.3.).

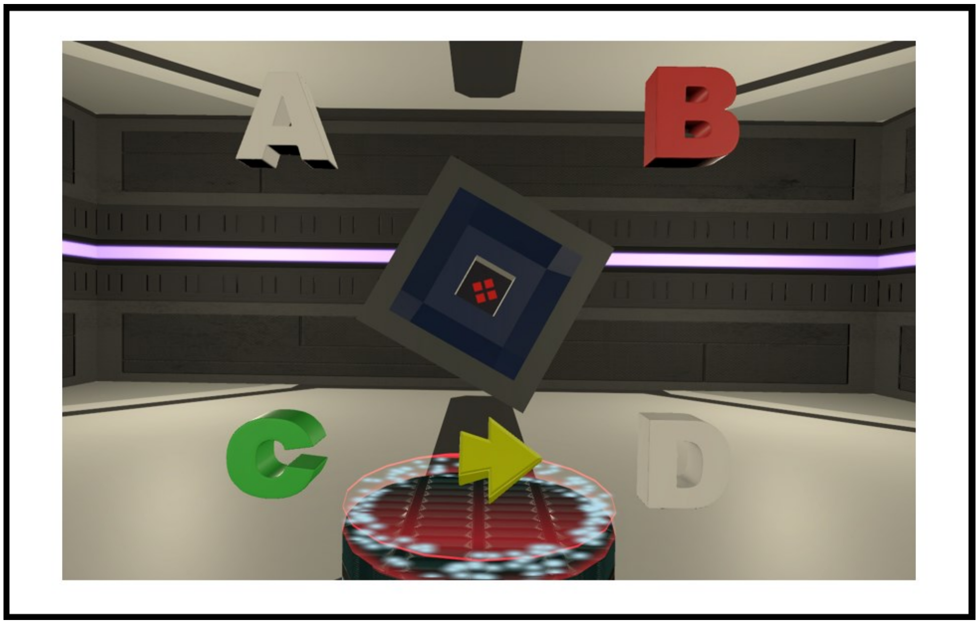


Figure 2.3. A screenshot showing the participant view after they have made a choice of category. The participant's choice (B) is highlighted in red, indicating that this choice was incorrect. The correct choice for this trial (C) is highlighted in green. The participant can spend as much time as they like during this phase looking over the cube again before moving to the next trial.

All possible cube sets were featured across the four conditions, shuffling what features would be irrelevant or relevant, for a total of 6 cube sets, as well as what position on the cube each feature would inhabit when the cube was displayed (top, left, right) for a total of 24 possible arrangements per cube set. Counterbalancing was used to ensure that nearly every possible arrangement of stimulus features in each cube set was used with at least one participant, and that a roughly equal number of the six cube sets were distributed across the conditions.

2.5. Variables Used to Study Information Access Costs

There are a variety of variables typically measured in category learning experiments. Each one offers a different perspective on the way in which different conditions impact learning related behaviours. Refer to Table 2.2. for a summary of these measures.

Accuracy is the percentage of correct responses scored in each trial block. This is expected to improve for all groups over the course of the experiment, but by examining the rate at which it increases, it can be seen if any groups reach peak accuracy faster or slower than the others.

Response Time is the time it takes to make a response during each trial and represents how quickly participants identify stimuli. This will be analyzed to indicate any differences in speed across the experimental conditions. However, response times in this experiment are confounded by the fact that participants face different levels and combinations of an information access cost. For example, each fixation in the high delay cost condition adds an additional 5 seconds to the trial that is uncontrolled by the participant, and those with higher motor costs can be expected to have longer response times simply because the stimulus cube is rotated more slowly. To evaluate the impact of each condition on response times while controlling for the actual conditions themselves, an **adjusted response time** is calculated for each trial. The adjusted response time removes all time spent looking at the countdown timer, as well as all time spent on the corners of the cubes, leaving us with just the total time spent actually looking at the stimulus features themselves before making a choice.

Attentional Optimization gives us a ratio of how much time a participant spends looking at relevant information vs irrelevant information in each trial for the current block. This is used to track how participants learn to prioritize important information over time during learning, ranging from -1 (only looking at irrelevant information) to 1 (only looking at relevant information). This measure is calculated using only the time in which a feature is actually visible to the participant and does not include time spent waiting for the feature cover to disappear.

Following Dolguikh et al. (2021), studying the amount of **time spent on feedback** after each guess by participants can help to see if information access costs incurred during the trial influence how long participants engage with feedback before choosing to continue to the next trial.

The average **fixation duration** for each time participants spend fixating on the sides of the cube, as well as the **number of fixations** per trial, when measured together, provide a more nuanced picture of how efficient learners become over time under different conditions (Chen et al., 2013; McColeman et al., 2014). Fixation Durations in this case only includes the amount of time looking at the feature information itself and does not include time spent looking at the feature masks during the delay period. They are calculated in-game using C# code to track when the features were visible and save fixation information to a file as each fixation took place, including what features were being looked at. Likewise, a fixation is only counted in each trial if the participant meets a minimum threshold of 75ms after the stimulus mask is lifted, similar to the Area of Interest based fixation detection algorithm described in Salvucci & Goldberg (2000). This approach has drawbacks when working with eye movements and can sometimes overestimate fixation durations. However, having predetermined areas of interest, and feature masks that ensure the participant is actually looking in the direction of the stimulus feature for it to become visible makes this a reasonable way to predict attention without actually having eyetracking equipment built into the headset. It is possible that a participant makes multiple eye movements while examining a stimulus feature, but the purpose of these fixations will be generally the same: to examine the stimulus feature in view. The direction of attention towards each feature is of more importance to the research question than the actual eye movements themselves, and so this approach is ideal for our goals.

Table 2.2. A summary of the variables of interest measured in the category learning task.

Categorization Task: Variables of Interest	
Variable	**Definition**
Accuracy	Proportion of correct trials in each block
Response Time	Average time between pressing the “start trial” button and making a category choice for each block
Adjusted Response Time	Average time between pressing the “start trial” button and making a category choice for each block, subtracted by the amount of time where stimulus features were not visible
Attentional Optimization	The average ratio of relevant to irrelevant feature viewing time during the response phase of a trial for that block. Ranges from -1 (only looking at irrelevant features) to 1 (only looking at relevant features).
Time Spent Viewing Feedback	The average amount of time a participant spends during the feedback phase of a trial before pressing the “Next Trial” button for each block.
Fixation Duration	Time spent on individual feature fixations during the response phase of a trial on average
Fixation Count	The average number of fixations made within the response phase of a trial

2.6. Variables Used to Study Individual Differences and Non-Learners

To measure learning, we used a learning criterion of 24 correct trials in a row to define learners and non-learners, as well as overall accuracy in the experiment for all participants. The criterion point was used to better compare the results to previous categorization research, using the same criterion of 24 correct trials in a row which was used in Barrett et al. (2022). According to the discussion in Smith and Minda (2000), this criterion is greater than the 9 errorless trials used by Medin and Schaffer (1978), but is fewer than the criterion of 36, 70, and 90 correct trials used in other studies (Hartley & Homa, 1981; Homa et al., 1979, 1981). Fundamentally, the arbitrary nature of these learning thresholds results in a biased estimate of learning and reduces learning to a binary true/false indicator, which fails to credit learners for partial successes. To avoid this issue, we will use overall accuracy of the experiment as our outcome measure for all participants when considering how their individual differences impact learning outcomes.

In this way, we can better differentiate between participants who reached high accuracy, those who were maybe close to reaching criteria but gave up, slow learners, and those who simply guessed randomly. A variety of attributes and participant traits have been proposed as being potentially relevant to category learning, summarized in Table 2.3., and the full text and protocol for each measure can be found in Appendix A.

Table 2.3. A summary of the variables to be used in investigating the individual differences possibly correlated with performance in category learning

Individual Differences: Variables of Interest	
Variable	**Definition**
Self-Efficacy Score	A continuous scale reporting the degree to which a participant reports feeling capable in their ability to learn new skills.
Growth/Fixed Mindset	A semi-continuous scale reporting the degree to which participants report attitudes which correlate with responses congruent with a growth or fixed mindset approaches to problem solving.
Mental Rotation Ability	A continuous scale reporting a participant's achieved score on a test of their ability to mentally rotate objects in mental rotation task.
Working Memory Ability	How many items a participant can hold in short term memory. Measured with the Complex Span Task.
Demographics	Age, sex, frequency of video game play, frequency of VR use, and ADHD or other attention-related neurodivergence.

Participants had their spatial ability measured using redrawn version of the Vandenberg & Kuse Mental Rotation Task (Peters et al., 1995; Vandenberg & Kuse, 1978). For each problem in the test, participants were shown a set of four 3D Tetris block-like images and asked to compare these to a target image. Two images in each set of four could be mentally rotated to match the target image, and participants had to get both correct to get a point for that problem. The structure of the test begins with a brief tutorial, followed by two sets of 12 problems, with 4-minutes to go through each set, separated by a small break in between.

The Corsi block-tapping Task as described by Corsi (1972) and Kessels et al. (2000) was used to measure how much information participants could hold in working memory during a visuo-spatial task. In this task, participants were shown a set of blocks which light up in sequence. Participants were then asked to repeat the sequence. After every correct sequence, the next trial increases the sequence length by one until a mistake is made. If two mistakes are made in a row, the test ends, and the maximum sequence length completed successfully is recorded as their Corsi Span. Kessels et al. (2000) reports a span of 5-6 being considered average for “normal” human subjects.

To measure self-efficacy, we used the short form of Bandura’s General Self-Efficacy Scale developed by Romppel et al. (2013), and a person’s alignment with the Growth and Fixed mindset tendencies was determined using the adjusted Growth Mindset Scale validated by Midkiff et al. (2018). Due to a technical error, one of the growth mindset questions was excluded from the survey, and so the score here is based on seven questions instead of eight.

Additional information about the participants was collected as well, including their sex, age, experience with VR and video games. Based on feedback from participants during the first round of data collection, an amendment to the running procedures was made to add a question asking if participants had ADHD or any other attention related neurodivergence.

As discussed already, each of these variables has some plausible linkage to learning outcomes in categorization tasks as well as in virtual reality learning tasks. Using the ‘lme4’ R package developed by Bates et al. (2014), we performed linear modeling to identify which of the variables in Table 2.3. were predictive of a participant’s accuracy.

2.7. Procedures

Participants filled out a computer-based questionnaire built using the PsyToolkit software package (Stoet, 2010, 2017). In this survey, participants provided basic information about themselves, including age, sex, whether they had ADHD or any other attention related neurodivergence, how often they play video games or use virtual reality, as well as any vision differences such as whether they have colour blindness, glasses, or contact lenses. This was followed by the General Self Efficacy Assessment Tool, the Growth Mindset Questionnaire, and finally the visual Corsi task to assess working memory. When this part of the experiment was finished, the experimenter administered the paper-based Vanderburg Mental Rotation Test, using the 4-minute variation described by Peters et al. (1995).

After these measures were collected, the participant was introduced to the VR headset, given a brief tutorial on how to use the controllers, and shown how to adjust the headset for optimal visibility. They were warned that some people experience discomfort or sickness while in VR and were instructed to withdraw by notifying the experimenter if they experienced any of these feelings. Following this, the researcher opened the Unity program to begin the experiment, giving the participant a chance to ask any remaining questions they might have before starting the program.

The experimental program consisted of two main sections: a tutorial, and the main experimental trials. During the tutorial, the participant was taught to rotate a stimulus cube and make guesses as to its category. They were then shown an example of trial feedback where they could compare the stimulus cube to the feedback presented. Transitioning into the experimental trials, each participant underwent at least 96 trials. 96 trials were chosen based on the fact that in Barrett et al. (2022), all participants who reached the criterion point reached it prior to the midpoint of the experiment. After 96 trials, or approximately 40 minutes of experiment runtime, the experiment ended, and the participant was asked briefly about their experience by the experimenter as they cleaned up the equipment. This conversation was aimed at probing for possible bugs or other issues with the experiment to address. The participant was then given a debrief slip with information about the study and the researcher’s contact information in case they had additional questions about the research.

Chapter 3. Results

3.1. Analysis of Learning Outcomes and Attention Related Behaviours

Using R Statistical Software (v4.0.0; R Core Team, 2020), and the linear mixed effects regression package developed by Bates et al. (2014), we investigated how each of the different experimental factors contributed to changes on the outcome variables for all participants included in this analysis (N=47). Initially, the base models used trail bin (bins 1-4) as a fixed effect, and Participant ID as a random intercept to account for individual variation in the data. Each factor was introduced as a fixed effect individual before introducing the interaction of both terms. Inspection of histograms of residuals and qq-plots for each dependent variable indicated that normality was generally observed, with LMERs being robust to most deviations from normality. As well, there were no visible indicators that homoscedasticity was greatly violated in any of the variables. Likelihood ratio tests were then used to compare models with and without the individual factors to see if adding that term would result in better model predictions. Modelling factors in this way allows us to see if they have an effect on the outcome variable, while also allowing us to look for interaction effects between them. The largest model we explored for each dependent variable can be denoted through the formula: DV ~ Bin + Bin^2 + MotorCost + DelayCost + MotorCost:DelayCost.

For every outcome variable, a practice effect was obvious, with participant scores changing across the training period as they learned to categorize the stimulus groups. Visual inspection of the graphs in Figure 3.1. indicates more of a curvilinear progression for each outcome variable across the training period. Testing this, a quadratic term was added to each base model and was found to be significant for all variables: accuracy ($\chi^2(1) = 36.99$, $p<0.001$), response time ($\chi^2(1) = 54.81$, $p<0.001$), adjusted response time ($\chi^2(1) = 58.54$, $p<0.001$), optimization ($\chi^2(1) = 6.12$, $p=0.013$), feedback duration ($\chi^2(1) = 41.65$, $p<0.001$), fixation duration ($\chi^2(1) = 39.689$, $p<0.001$), and fixation count ($\chi^2(1) = 35.387$, $p<0.001$). Every dependent variable having a significant improvement in model fit after the inclusion of a quadratic effect indicates that in addition to there being an effect of practice on scores, that relationship is curvilinear nature, showing that the largest changes in learning occurred earlier in the experiment.

Next, each group was examined to see how many participants reached the learning criterion (N=47) and how many failed to reach the criterion required to be included as a "learner" (N=72). Using Pearson's Chi-squared Goodness of Fit test, counts of learners were compared first along each of the main groups, motor and delay costs, and then with all four conditions treated separately to detect any interaction effects. Starting with the impact of delays, we observed that having an increased delay cost did not change the likelihood that participant would succeed in learning the categories ($\chi^2(1, N=119) = 2.604, p=0.107$). Likewise, an increased motor cost did not change the likelihood of reaching the criterion point ($\chi^2(1, N=119) = 0.11, p=0.740$). Treating each condition individually also yielded no differences in the proportions of participants who reached criterion in this experiment. ($\chi^2(1, N=119) = 2.773, p=0.428$). All groups had a roughly equal ratio of learners to non-learners, and no single condition produced a disproportionate number of non-learners.

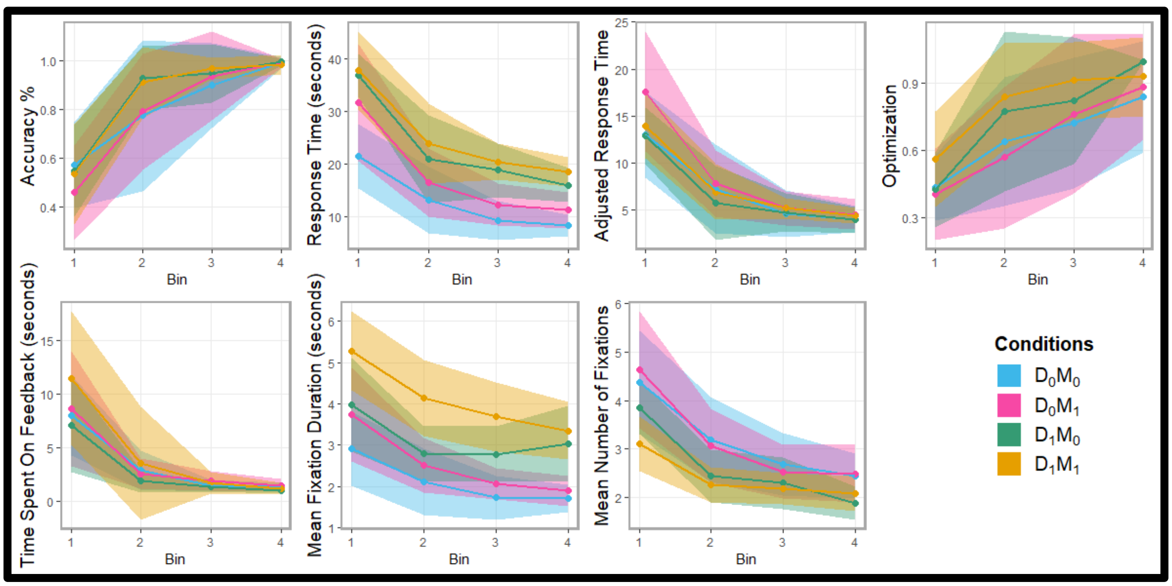


Figure 3.1. Means and standard deviations for all learning outcomes and attention-related behaviours, plotted by bin for each condition. In the legend in the bottom right corner, D and M refer to the delay and motor costs respectively, with the subscript indicating whether the cost was low (0) or high (1).

3.1.1. Learning Outcomes

For each analysis, Figure 3.1. showcases the data for each of the four conditions individually, while Figure 3.2. and Figure 3.4. display the main of effects for delay and motor costs, respectively. Starting with accuracy, there was no effect of motor cost ($\chi^2(1) = 0.176$, p=0.675), or of delay cost ($\chi^2(1) = 2.058$, p=0.151). No interaction was found between the two factors either ($\chi^2(2) = 2.195$, p=0.533). All participants who reached the learning criterion point were able to reach that point at relatively the same trial number no matter their condition.

Response time for each group, as pictured in Figure 3.1., was also relatively uneventful. Motor cost was found to have no effect on response times ($\chi^2(1) = 3.321$, p=0.068). As can be seen in Figure 3.2. however, delay cost impacted response times ($\chi^2(1) = 27.601$, p<0.001), as the higher delay conditions forced participants to wait 5 seconds on each fixation before they continue on. Modelling the interaction effect was found to improve the model fit compared to the delay cost on its own ($\chi^2(2) = 9.182$, p=0.007), showing again that design of the conditions had forced participants to go slower on each trial, whether they are slowed by how much effort it takes to access information, or how long they have to wait for access to it. While this finding is somewhat confounded by the nature of the experimental manipulation, on its own, this reminds us that any potential benefits of information access costs must be weighed against the trade-off of longer response times.

Using the adjusted response times which control for the delays imposed by the experimental manipulations, I examined how removing the time added by the conditions themselves could potentially isolate any difference in time on task by the participant spent in addition to the time necessary to navigate the information access costs. Looking at both costs through this perspective finds no impact of motor ($\chi^2(1) = 3.587$, p=0.058) or delay ($\chi^2(1) = 1.313$, p=0.252) costs. Testing for a possible interaction between these effects also yielded no results ($\chi^2(1) = 3.782$, p=0.151). These analyses show that although participants did take longer to navigate each trial overall, no additional time per trial was added beyond that which resulted from the nature of the conditions themselves.

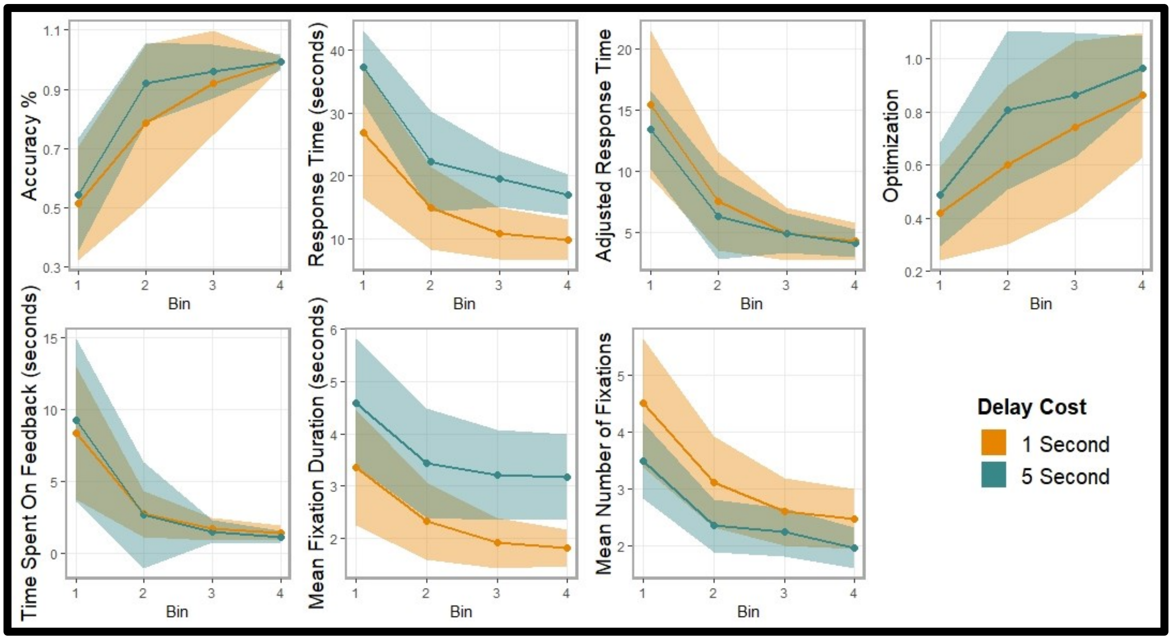


Figure 3.2. Means and standard deviations for each learning outcome and attention-related behaviour, grouped by Delay Cost.

In a more applied setting, such as in education, time to learn is not measured in trials, but in the actual time spent on task. It is plausible that accuracy may have increased at the same rate across groups in terms of trials, but the longer response times may indicate that the accuracy over time might progress faster for groups with lower information access costs who can go through more trials per minute. To see if learning outcomes over time might be impacted by this, the trials were binned into five-minute chunks and the analyses of accuracy were redone using the new, time-based bins. Modeling the trials in this way produced the same results. Again, adding a quadratic model to the base model was found to be a better fit than the base linear model ($\chi^2(1) = 36.999$, $p<0.001$). For each condition, neither the motor cost ($\chi^2(1) = 0.176$, $p=0.675$) nor the delay cost ($\chi^2(1) = 2.058$, $p=0.151$) improved the model fit, indicating that neither condition reliably impacted the overall rate of learning in participants. A visual inspection of the graph for this analysis (Figure 3.3.) indicates that the group with no delay cost and no motor cost may appear to be improving slightly more quickly at first, but that this advantage disappears after the first twenty minutes of learning.

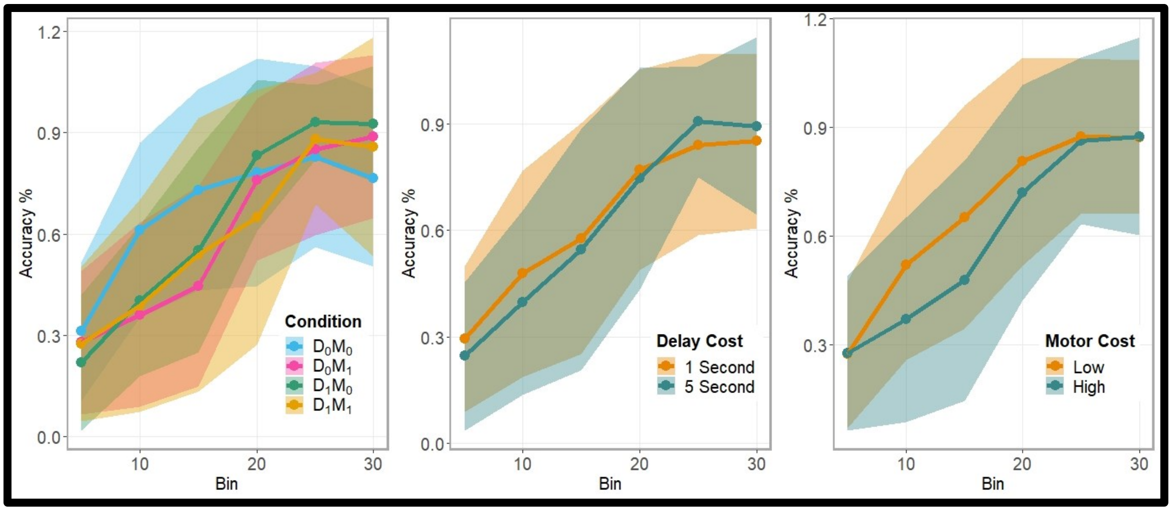


Figure 3.3. Accuracy scores, binned into five-minute chunks. The far left graph shows accuracy over time for all four groups, while the two figures to the right show accuracy over time for each of the main effects.

3.1.2. Information Access Behaviours

After making their choice for each trial, participants had the option to review the cube and feedback together so as to examine the stimulus features with the correct answer in view. Neither motor cost ($\chi^2(1) = 2.897$, p=0.089) nor delay cost ($\chi^2(1) = 0.031$, p=0.861) were found to have any impact on how long people viewed the feedback for. There was no interaction between the two factors either ($\chi^2(2) = 5.493$, p=0.139). Participants examine the feedback for 5-15 seconds during the first bin but spend less than 5 seconds by bins two and onward.

Participants learned to prioritize important information and ignore irrelevant features as they examined the stimulus cubes, with all learners improving their optimization gradually over the course of the experiment. However, increased motor costs did not impact the how quickly participants learned to prioritize important information ($\chi^2(1) = 0.036$, p=0.85). Increasing the delay costs did result in higher optimization scores ($\chi^2(1) = 4.115$, p=0.043), showing that increased delay costs motivated participants to avoid irrelevant information more quickly. There was no interaction effect between the motor and delay costs ($\chi^2(2) = 0.442$, p=0.802), showing that the impact of the delay cost was not impacted by increases in motor cost.

Focusing only on the response times fails to account for time spent looking at objects other than the stimulus and is unable to tell us about the attentional patterns of the learner. By examining the number of fixations and their respective durations, we can work to see how each condition may have influenced the way participants viewed task relevant information. Starting with fixation durations, motor cost ($\chi^2(1) = 4.821$, p=0.028) and delay costs ($\chi^2(1) = 26.806$, p<0.001) were both found to have a significant impact on the model, as participants facing a higher information cost tended to spend more time looking at each feature, regardless of the type of cost they experienced. Modeling an interaction effect between these factors and comparing it to the impact of the delay cost on its own was also found to be significant ($\chi^2(2) = 14.013$, p<0.001), showing how the delay cost, when combined with the motor cost was more impactful on how long participants fixated on each feature.

The number of fixations per trial was not impacted by having a higher motor cost in this experiment ($\chi^2(1) = 0.056$, p<0.814). However, an increased delay cost did make a significant impact on the model ($\chi^2(1) = 17.477$, p<0.001), showing that when faced with a higher delay cost, participants made fewer fixations per trial on average across all bins. No interaction effect was found between the two factors ($\chi^2(2) = 1.052$, p<0.591), suggesting that the combination of motor and delay costs had no impact on the number of fixations for participants who reached the criterion point.

Across these analyses, the findings can be summarized as follows: All groups learned at roughly the same rate, and a proportionally equivalent number of participants reached criterion in all four conditions, showing that our manipulations did not impact the essential learning outcomes of this task, reaching criterion, and the rate at which participants learned to identify the stimuli. Although a statistical impact of condition was observed in participant response times, this is accounted for by the additional time required to access information in those conditions and cannot be easily attributed to the experimental manipulations. Information access patterns were also impacted by the conditions in this experiment. During trials, the effect of having a higher delay cost was seen on three measures, improving how participants chose to prioritize relevant information over irrelevant information, increasing the amount of time they linger on each feature before moving to the next fixation – especially when combined with increased motor costs – and lastly, participants facing higher delay costs made much fewer fixations per trial on average.

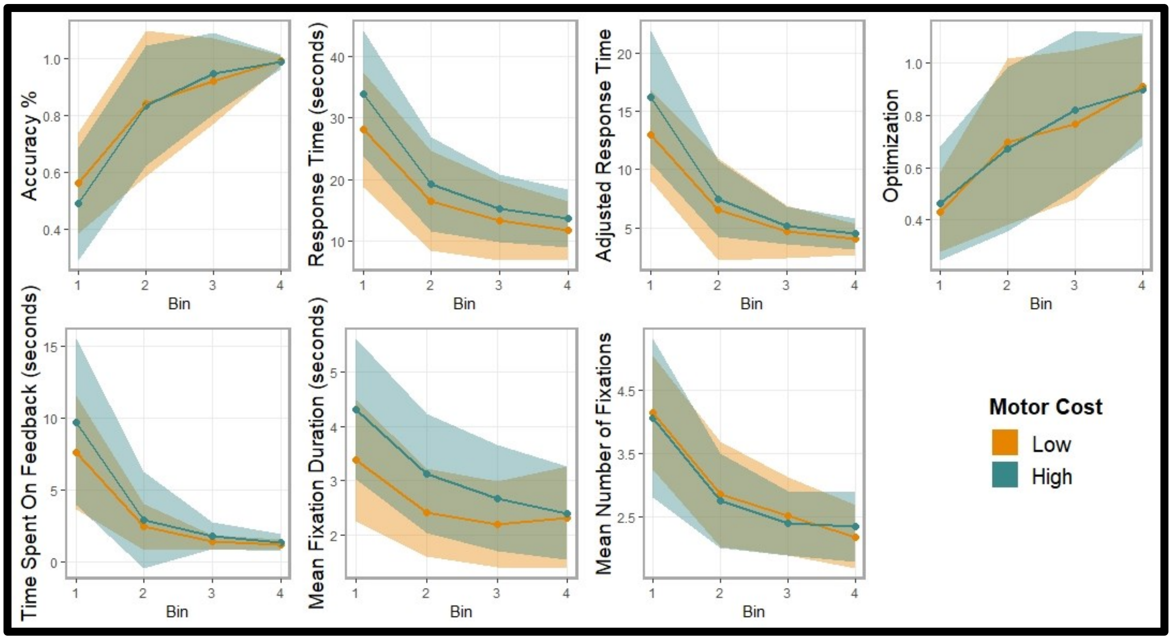


Figure 3.4. **Mean and Standard Deviation across the four bins for each attention-related behaviour and learning outcome, grouped by whether they faced a high or low motor cost when manipulating the stimulus cube.**

3.1.3. Replicating Barrett et al. 2022

In many ways, the group with the low motor cost and low delay cost was intended to act as a pseudo-replication of the VR group from Barrett et al. (2022). To determine whether this group successfully replicated the findings of Barrett et al. (2022), additional analyses were made using just the first four bins of the VR group in Barrett et al. (2022) and the Low Motor Low Delay group from the current experiment. Initially, the base models used trial bin (bins 1-4) as a fixed effect, and Participant ID as a random intercept to account for individual variation in the data. Each factor was introduced as a fixed effect individually before introducing the interaction of both terms. The largest model explored for each outcome can be described with the formula: DV ~ Bin + Bin2 + Experiment + Bin2:Experiment.

As before, each outcome variable was found to improve with practice. A quadratic term was added to each base model and was found to be significant for all variables: accuracy ($\chi^2(1) = 19.291$, $p<0.001$), response time ($\chi^2(1) = 17.695$, $p<0.001$), adjusted response time ($\chi^2(1) = 17.70$, $p<0.001$), optimization ($\chi^2(1) = 4.930$, $p=0.026$), feedback duration ($\chi^2(1) = 27.628$, $p<0.001$), fixation duration ($\chi^2(1) = 12.62$, $p<0.001$), and fixation count ($\chi^2(1) = 5.30$, $p<0.021$). This shows that learning occurred in both experiments and that most learning occurred in the earlier parts of the experiment.

Starting again with accuracy, Figure 8 shows that there was no substantial difference across the experiments in how quickly each grouped learned to categorize the stimuli in this experiment ($\chi^2(1) = 3.158$, $p=0.076$). No interaction was found between the experiment and bin either ($\chi^2(2) = 3.111$, $p=0.211$), showing that participants in both experiments learned at roughly the same rate.

Response times were found to be different between the two experiments ($\chi^2(1) = 12.522$, $p<0.001$), suggesting that the 1 second delay used in the current experiment may have resulted in much slower response times by participants. As can be seen in Figure 3.5., an interaction effect was observed between the conditions and the bin number ($\chi^2(2) = 17.5$, $p<0.001$), showing that the shape of the curve for reaction times was different as the group from the current experiment started with much slower response times, but had mostly caught up with the other experiment by bin 3.

The adjusted response times in this case only control for the impact of the 1 second delay cost, leaving in the time between fixations as the motor cost across these experiments was identical and therefore unnecessary to factor out. After controlling for the delay cost, response times were no longer found to be different across the experiments ($\chi^2(1) = 1.75$, $p=0.187$), suggesting that the increased reaction times observed in later trials was an artefact of the changes made to the current iteration of the experiment. However, an interaction effect was still observed between experiment and bin ($\chi^2(2) = 9.094$, $p=0.011$), showing how although the response times were generally similar across the experiments, participants in the current study still started with slower response times, again catching up completely by bin 3.

Optimization scores across the two experiments were also found to be different ($\chi^2(1) = 4.529$, $p=0.033$), with participants in the current experiment starting with higher optimization scores and maintaining that advantage. No interaction effect was found ($\chi^2(2) = 5.700$, $p=0.127$), showing that the advantage for participants with the 1 second delay was consistent across the bins.

Participants in both experiments spent roughly the same amount of time on feedback across the four bins ($\chi^2(1) = 2.209$, p=0.137), but an interaction effect was observed ($\chi^2(2) = 11.253$, p=0.01), as the shape of the corresponding curve in Figure 3.5. shows that participants in the current experiment spent slightly more time on feedback in the first bin compared to the participants in Barrett et al. (2022), but that this difference was not observed for the rest of the experiment.

For fixation durations, it is clear that participants in the present experiment spent much more time looking at each stimulus feature ($\chi^2(1) = 16.532$, p<0.001). An interaction effect was also observed ($\chi^2(2) = 19.136$, p<0.001), showing that while participants in Barrett et al. (2022) started with shorter fixations and stayed relatively short, participants in the current experiment started off spending much more time per fixation and improving over the course of the experiment, though not as quick as the subjects in Barrett et al. (2022).

Lastly, participants in the current experiment made much fewer fixations compared to the participants in the previous experiment ($\chi^2(1) = 12.344$, p<0.001), and inspection of the associated graph in Figure 8 suggests that subjects in the current experiment were much more predictable in how many fixations they would make while those in Barrett et al. (2022) were much more unpredictable in addition to having more fixations overall in their first four bins. No interaction effect was observed ($\chi^2(2) = 2.024$, p=0.363), suggesting that this difference was consistent across all four bins.

In summary, the present experiment was often a good comparison to the previous experiments in terms of learning outcomes. Several important differences in optimization, time spent on feedback, fixation durations and number of fixations per trial were observed however, suggesting that even a one second delay cost was enough to trigger several important shifts towards more economical use of attentional resources.

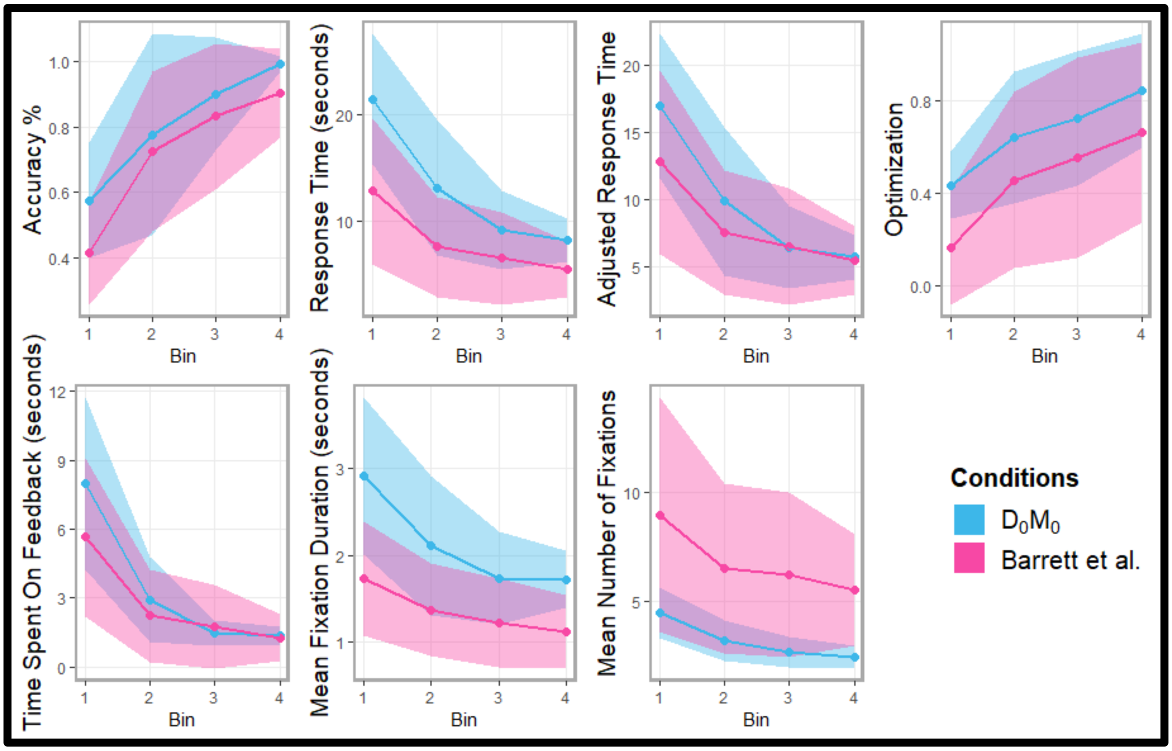


Figure 3.5. Means and standard deviations for all learning outcomes and attention-related behaviours, plotted by bin for both the Low Motor Low Delay Group from the current experiment and the first 4 bins from the VR group documented in Barrett et al. (2022).

3.2. Exploration of Individual Differences

To explore the impact of individual differences on overall accuracy, a forward stepwise linear regression was used to ascertain which survey instruments would be potential predictors of accuracy in the experiment. Sex, age, video games per week, VR usage level, self-efficacy, fixed mindset tendencies, growth mindset tendencies, working memory, and mental rotation scores were all included in the stepwise regression as possible candidates as predictors of accuracy. Missing data in any variable was dealt with by excluding that participant, resulting in a total sample size of 106 participants for this analysis. ADHD was excluded from the regression model building to be examined independently for this reason, so as not to diminish the statistical power by removing an additional 32 participants. For each part of the process, variables were evaluated according to their p-values in conjunction with the AIC to narrow down a list of recommended variables to use in the final model. Of the ten variables included in the model, age, video games per week, and mental rotation were found to be possible predictors of accuracy. Coefficients for these variables suggest that each increase in the level of video games played per week should be associated with a 4.8% increase in accuracy, each point earned in the mental rotation test is predictive of a 0.9% increase in accuracy, and each year of age should predict a -2% drop in accuracy. In all cases, these effect sizes are small, and should be explored more thoroughly to confirm whether they are reliably predictive of performance.

Fitting a linear model to predict accuracy with age, video games per week, and mental rotation. The model was able to explain 11.3% of the variance in accuracy and was statistically significant, F(3, 105) = 4.461, p = .005). Though the model itself was found to be significant, none of the individual predictors had statistical significance with alpha set to the .05 level. The effect of Age was statistically non-significant (beta = -0.02, 95% CI [-0.04, 1.50e-03], t(105) = -1.85, p = 0.068; Std. beta = -0.17, 95% CI [-0.36, 0.01]). Next, the effect of video games per week was statistically non-significant (beta = 0.05, 95% CI [-4.35e-03, 0.10], t(105) = 1.82, p = 0.072; Std. beta = 0.18, 95% CI [-0.02, 0.37]). The effect of mental rotation score was also statistically non-significant (beta = 9.11e-03, 95% CI [-5.51e-04, 0.02], t(105) = 1.87, p = 0.064; Std. beta = 0.18, 95% CI [-0.01, 0.37]). Given this information, any relationship suggested between these predictors and the outcome variable must be interpreted cautiously as the model itself not a reliable predictor of the data. Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p-values were computed using a Wald t-distribution approximation.

Since the predictors identified by the stepwise linear regression were found to be nonsignificant when combined into a model featuring all three predictors, the effects of each predictor were explored individualy, as combining the predictors into a larger model may have reduced sensitivity to the effects of each predictor. A baseline model was built using just accuracy as an outcome variable, again using the ‘lme4’ package in R (Bates et al., 2015), with each factor introduced as a fixed effect one at a time to identify which predictor variables seemed to have the most impact on accuracy. Each model described in table 3.1. was constructed as a linear model predicting Accuracy as a function of the survey outcome being examined. With the exception of the ADHD model, which was built with only 74 data points, each model in this analysis includes the data of 108 subjects.

Table 3.1. A summary of the statistical analyses performed for each of the individual differences investigated in the survey.

Variable	Estimate	Std. Error	R^2	t-value	df	p-value
Sex (male)	0.031	0.057	-0.006	0.552	1,107	0.582
Age	-0.023	0.011	0.023	-1.884	1,107	0.06
Age (<25)	0.073	0.081	-0.003	0.901	1,105	0.37
Video Games Played Per Week	0.067	0.025	0.052	2.636	1,107	0.01*
VR Usage	0.027	0.032	-0.003	0.836	1,107	0.405
ADHD	0.073	0.081	-0.002	0.301	1,73	0.371
General Self Efficacy	-0.003	0.006	-0.008	-0.420	1,107	0.675
Fixed Mindset	0.024	0.032	-0.004	0.741	1,107	0.405
Growth Mindset	-0.047	0.036	0.007	-1.307	1,107	0.194
Working Memory (Corsi Span)	0.019	0.020	-0.001	0.957	1,107	0.351
Spatial Ability (Mental Rotation Score)	0.011	0.005	0.037	2.278	1,107	0.025*

Note: * $p<0.05$

Examining each of the factors recommended by the stepwise regression in turn, age was found to have a statistically non-significant effect on accuracy ($F(1,107) = 3.548$, $p=0.062$). Had this effect been found to be statistically significant, it would have indicated only a small 2.3% drop in accuracy for every year of age. In addition to this effect being quite small, it is apparent that this effect seems to be driven by the inclusion of two participants over the age of 25 who had particularly low accuracy scores. Further analysis does show that these participants were in fact outliers, and rerunning the analysis without these two participants reduced the effect to below threshold to be considered reliable ($F(1,105) = 0.985$, $p=0.323$). The difference in outcome here based on the removal of just two participants demonstrates that the effect suggested in the stepwise regression was not a stable relationship and should be disregarded until further evidence with more subjects in the higher age categories can be collected.

ADHD was not found to have any impact on accuracy ($F(1,73)=0.812$, $p=0.371$). Of the models which exceeded a threshold of statistical significance of alpha = 0.05, experience with video games was found to be associated with an improvement in accuracy ($F(1,107)=6.949$, $p=0.009$). Each level of weekly use in the survey was associated with an approximate 6.7% increase in accuracy in the experiment. Adjusted R-squared was only 0.052 however, indicating that this effect is highly volatile, only accounting for 5.2% of variability in accuracy scores overall. As well, mental rotation ability was found to have a slight effect on overall accuracy ($F(1,107)=5.191$, $p=0.025$). This effect was extremely minor however, with each point of mental rotational ability only be associated with a 1.1% increase in accuracy, and a R-squared of only 0.037 indicating that this predictor accounts for only 3.7% of variability in accuracy scores. Taken together, this information shows us that these individual differences have very low predictive power on accuracy.

Not all participants reached the final trial however, as several ran out of time despite having learned the categories. In cases, it can be assumed that they would have gotten 100% accuracy on subsequent trials had they been allowed to continue, resulting in a higher overall accuracy score than observed. To counteract this artificial deflation of scores, just the first two bins were modelled with each predictor again, as all participants have data for these bins. By using just the first two bins, all participants have an equal number of trials in the analysis. The impact of video game experience was retained in this analysis ($F(1,107)=7.331$, $p=0.008$), with an expected increase of 6.8% in accuracy for each level of weekly video game play. However, the effect of mental rotation was not observed in this sub-setting of the data ($F(1.107)=3.2$, $p=0.076$), now predicting only an 0.8% increase in accuracy for each point of increase in mental rotation scores, giving further evidence to the unreliability of this factor.

In summary, only weekly video game usage seemed to have any impact on how well participants did during the experiment, suggesting that playing more video games per week was associated with slightly better learning outcomes. A small effect of mental rotation score was observed, but it was small and unstable, being completely undetectable when examining just the first two bins of data. Age was suggested as an important predictor early in the analysis, but further investigation found this to be unreliable as well, likely driven by outlier data points. In short, almost no individual differences measured had any impact on accuracy scores in this experiment, indicating that other factors must contribute to non-learner rates in these kinds of experiments.

Chapter 4. Discussion

In the first research question, the impacts of different information access costs were tracked across different learning outcomes and attention related behaviours as participants learned to categorize stimuli into different groups. It was found that participants faced with increased delay costs would prioritize important information more than participants with no delay cost, spending less time on irrelevant features over the course of the experiment. As well, both information access costs – delay and motor – were found to increase the amount of time each participant would spend looking at individual features. Having both costs present increased that time even further, showing that participants seemed to be more cautious when examining each feature. Having increased delay costs also led to participants making fewer fixations per trial over the course of the experiment.

Comparing data from the control group of the current experiment to the data from Barrett et al. (2022) revealed that while this data was generally able to replicate the learning outcome trends of that previous work, the one-second delay used in this experiment was enough to trigger the type of changes in attention predicted by previous work on information access costs (Morgan et al., 2010; Rajsic et al., 2018). That even a one second delay could make such a difference is surprising as this group had been intended as a pseudo-replication of the previous study. Some differences exist between the studies, with the current project having the addition of surveys and other pre-measures and an updated game program, but the cube stimulus was directly transferred from the previous project into this one, making the stimuli mostly identical at minimum. A follow-up experiment explicitly comparing these delays would be ideal to confirm this finding.

Higher delay costs did seem to make participants more economical in their information access behaviours, making fewer fixations as they went through the trials and spending more time on each fixation. This was found both in the current experiment where the difference in delay was a lengthy 4 seconds, and in the comparison between the control group of this study and the VR group from Barrett et al. (2022) where there was no delay experienced when accessing stimulus features. Even the one second difference introduced in the current experiment seemed to be enough to impact information access patterns, showing that even small differences in the design of experimental and educational environments can make significant impacts on how people allocate their attention. Lingering on each fixation for a longer period of time could be motivated by a desire to further encode the stimulus feature into memory, as forgetting the feature would result in needing to return to that side of the stimulus, wasting additional time. In contrast, participants with shorter or no delays are not impeded as much if an additional fixation is needed to refresh their memory, so less importance is given to each fixation. This finding bears some resemblance to research investigating the impact of delays on temporal discounting in pigeons (Grace et al., 2012), where it was found that small differences in delay closer to 0 had much more impact on temporal discounting compared to proportional differences that were not close to zero. Likewise, the impact of the 1 second delay compared to the 0 second delay was far greater than the difference between the 1 and 5 second delay groups, despite being a larger difference in delay. This suggests that the impacts of information access costs on attention related behaviour may be tightly linked with the impact of temporal discounting on decision making. Experimenting with a variety of delay costs may be useful to further define the nature of this relationship.

It is also possible that participants experiencing longer delays would be more inclined to learn the categorization rules while those with less delay may opt to make a third fixation because they are using a memorization strategy. The number of fixations on each trial gives us some indication that this is the case, as participants with higher delays tended to make fewer fixations on average, with most participants in the 5 second delay group only needing two fixations to correctly identify the stimulus; the minimum needed to do so. In contrast, those with lower or no delay cost as in Barrett et al. (2022), are still found to make more than two fixations, suggesting that participants with lower delay costs still examine all the features before making a category choice, even after reaching criterion, relying on a memorization strategy rather than learning the categorization rule. Blair and Homa (2003) found that when a small number of unique stimuli were available, participants would often use memorization rather than learn the categorization rule, and that this strategy was often just as effective as rule-based approaches. As such, given that there were only 8 unique stimuli used in this experiment, memorization is just as easy as learning the rule, but having a higher delay cost motivates the learner to seek out the more efficient rule-based strategy.

Memorization strategies can produce successful results, but it’s been observed that when this approach is used during training, learners have difficulty transferring their knowledge to novel cases (Little & McDaniel, 2015). When the goal is to teach learners to generalize knowledge using more rule-based features, adding sufficient information access cost seems to be a viable way to ensure that attention is paid to the actual pattern instead of undermining the transferability of their learning by relying on memorization.

Motor costs did not seem to have a great effect in this experiment compared to prior work (Yang et al., 2013, 2015). While some effects were observed, such as longer fixation times, these effects were not as pronounced when compared to the impact of the delay cost. A possible explanation for this is that the motor cost was too small to be of much impact here. Despite the rotational drag of the stimulus requiring participants to make slower arm movements to rotate the cube, still only a single arm movement was necessary to change to a different side of the cube. In contrast, participants in Yang (2015) were required to walk across a room to access information, a much greater cost than a simple arm movement. Potentially, had the rotational drag of the cube in this experiment required multiple arm movements to access each side, or participants required to walk around the virtual space more, an observable effect might have been elicited. As well, the doctor’s who participated in Yang et al (2013, 2015) were not given a lot of opportunity to improve their performance over time, while participants in this experiment were given nearly a hundred trials. Perhaps training doctors with a higher information access cost would reduce initial training performance but improve later patient care following training. Future research should investigate at what scale motor costs need to be to produce an observable impact on information access patterns. At minimum, it is clear that delay costs are far more potent than motor costs, with only a second’s difference being sufficient to produce reliable differences in attentional patterns.

Lastly, when the delay cost was implemented, this resulted in consistently longer fixation durations. Although the presumed time it would take to perceive the stimulus feature would be the same, participants still spend longer looking at it, proportional to the delay cost used in that condition. Participants with a 1 second delay spent longer than those with no delay, and those with a five second delay spent even longer on each fixation. One explanation has already been discussed, suggesting that this is to ensure deeper encoding of the information being presented, but another explanation is that the delay itself slows down the speed of thought in the participant. Esports athletes often engage in seemingly useless actions during performance (“Actions per Minute,” 2024). At higher levels of skill, these actions help players maintain speed so that they can react quickly to the situation as it unfolds. In esports, this has led to a differentiation between “Actions per Minute”, and “Effective Actions Per Minute” which filters out redundant actions taken to maintain a higher APM. In this task, participants were forced to engage in slower actions by the delay cost, limiting their ability to build up any speed during performance. It is possible that being forced to go slower also reduced the rate of processing and thought, leading to longer fixations.

In my second research question, I explored possible reasons as to why so many people failed to reach the criterion point to be considered learners in these kinds of experiments. Nothing substantial was found, with only people’s level of weekly video game play being found to predict only a slight increase in accuracy. Although seemingly uneventful, these findings have some important implications for thinking about category learning research.

The effect of weekly video game usage observed here makes sense through the lens of cognitive load theory (Sweller, 1988, 1994; Sweller et al., 2011). This theory suggests that each person possesses a certain capacity for how much task complexity they can handle at any given moment. As well, each learning environment has multiple aspects of difficulty that add to that complexity. There is the intrinsic difficulty inherent to the material to be learned, as well as extrinsic difficulties which, while irrelevant to the actual material being learned, still make learning that material more difficult. For example, a loud room may make it more difficult to concentrate on learning math problems. In this experiment, participants not only needed to learn to identify the categories of the stimuli presented, but to do this, they also needed to learn to navigate a novel immersive virtual reality user interface. Consequently, participants with more experience playing video games are likely to have some knowledge of the mechanics of digital games that would transfer to the VR game, allowing them to focus more on learning the categories. Experience with VR was not found to have a significant effect, which is somewhat contradictory to this claim, but so few people reported having more experience with VR games that it is difficult to say if this effect was missed due to not having many people with more experience in VR or not. To verify this claim, measurement of cognitive load during the experiment would be necessary to confirm whether those with less experience were in fact more heavily burdened by the additional challenge of learning to use the VR interface compared to those with some familiarity with video games. As well, future research should be more specific about what kinds of video games are played by participants, as different kinds of games may have different impacts on attention and learning (Bediou et al., 2018).

The fact that, for the most part, no main effects of individual differences were found raises the question that, if the attributes of the participants themselves could not predict whether participants reached learner status or not, then what is driving this high frequency of non-learners across these experiments? One other source of information comes from the notes made by the researcher while observing participants and in conversation with them after each appointment. When asked what they thought the pattern might have been, some participants describe thinking there was some sequential pattern to the stimulus categories, among other suggestions. They would suggest "if the last trial was an A with a green spiral, then it seemed like the next trial would always be B". On the surface, these suggestions seem strange, but they may actually indicate that participants had picked up a subtle regularity to how stimuli were presented. When presenting stimuli, each of the 8 possible cubes was selected in random order without replacement until all 8 cubes were presented before starting again every 8 trials, like drawing cards from a small deck. In this way, the randomness of the presentation order was still within some limits. For example, if the first 2 trials happened to be from group A, then the next 6 trials must be either B, C, or D. If the third trial we're to be group B, then there would only be a 1/5 chance that trial 4 would also be group B, while groups C and D would now have a 2/5 chance of being next. Participants may have picked up on this sampling pattern without realizing it, leading them to believe that some explicit ordering pattern existed where there was none.

Another insight from the running log notes reports that some participants described trying to see if the rotation of the feature mattered, while others did not realize one of the features had an alternate form at all. Participants with these kinds of responses were all non-learners for the most part, suggesting that their confusion as to what to look for may have impeded them from focusing on the important feature characteristics. Misunderstandings of the experiment protocols such as these directed me to examine the design of the experiments themselves. Going back through the literature, a number of papers provided possible leads for future investigation. Although their methods section makes no reference to this, Rehder & Hoffman (2005b) reveal in their discussion that participants were explicitly made aware of the fact that each stimulus feature possessed binary characteristics and were shown all possible symbols for each feature. Following the experimenter scripts of previous research (Barrett et al., 2022; McColeman et al., 2014; D. J. Smith & Minda, 2000), we did not show participants examples of each stimulus feature during the tutorial.

In both of these cases, not knowing the possible feature states or believing there was a sequence to the features, indicates that the participant might be unable to learn the categories not for any lack of intelligence, but because they have misidentified the parameters of the task. Future research should take additional steps to ensure that understanding of experiment protocols is confirmed, potentially by having participants write out the instructions of the task to confirm their understanding prior to being shown the first trials so the researcher can correct any misunderstandings. As well, it would be interesting to follow the example of older works and record participants as they talk out loud during trials (Fisher, 1916; Williams, 1971). In using talk-aloud methods, researchers can confirm whether the participants fully understand the task and affordances of the stimuli provided. Future research using talk-aloud methods would allow researchers to better record novel hypotheses used by participants, and in the case of those who report thinking there was some sequence to the data, talk-aloud methods could also gather participant reasonings midtrial to see if there is any connection between the sequences guessed at and the probability of the sequences presented to the participant. Following Wahlheim et al. (2016), future research might get participants to write out what they think the best strategy is after each trial bin, which might avoid the positive but confounding impact of concurrent self-explanation on performance in problem-solving (Berry, 1983). Explicit ordering could even be used as an experimental manipulation to investigate this ability more thoroughly, testing whether pseudo-random trends in stimulus presentation order can impact performance. Instructional design is an expansive field of research, with several findings demonstrating that even the slightest change to instructional materials can make a large difference in student understanding of content (Khalil & Elkhider, 2016; Massen et al., 2009; Mayer, 2014; Sweller, 1994). Because of this, researchers should be aware that their instructional methods follow evidence-based practices to avoid causing impacts to their results that are due to differences in protocol rather than experimental manipulations.

With respect to industry application, a few key insights are offered by this work. First, the implementation of an information access cost during learning in virtual reality was generally ineffective at impacting learning outcomes. Yes, participants were more economical in their use of fixations, but whether evaluated in terms of the number of trials completed or the amount of time spent during training regardless of trial number, information access costs made no impact on how quickly people learned to categorize stimuli. While information access costs do not appear to be a useful tool in reducing the amount if time it takes to learn, these findings do demonstrate that even a small delay cost may be useful during training to encourage the adoption of more efficient and transferable information gathering strategies. In this way, learners will be more prepared to apply their training to novel situations that are not identical to the examples presented during their training.

Chapter 5. Conclusions

This study adds further fuel to the growing body of literature showing that choice of media and modality does not greatly impact learning outcomes. To improve learning outcomes, more direct intervention may be necessary on the part of the educational designer to ensure that instructional materials take advantage of the best practices in educational design. In this experiment, two kinds of information access cost were implemented to explore whether learning outcomes and changes in patterns of attention could be influenced by increasing the amount of time or physical effort it takes to access information. Although learning outcomes were unaffected by these changes, increased costs, especially delay costs, had a substantial impact on learning-related changes to patterns of attentional allocation. Even a small delay cost of a second was enough to produce a substantial change in participant behaviour, showing how even the smallest alterations to the design of digital learning objects can change how learners engage with the material.

References

Actions per minute. (2024). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Actions_per_minute&oldid=1195425424

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Andreasen, N. K., Baceviciute, S., Pande, P., & Makransky, G. (2019). Virtual Reality Instruction Followed by Enactment Can Increase Procedural Knowledge in a Science Lesson. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. https://doi.org/10.1109/vr.2019.8797755

Angel-Urdinola, D., Castillo-Castro, C., & Hoyos, A. (2021). *Meta-analysis assessing the effects of virtual reality training on student learning and skills development* (Policy Research Working Papers, pp. 958–7). The World Bank, Education Global Practice. https://openknowledge.worldbank.org/bitstream/handle/10986/35299/Meta-Analysis-Assessing-the-Effects-of-Virtual-Reality-Training-on-Student-Learning-and-Skills-Development.pdf?sequence=1&isAllowed=y

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology*, *4*(3), 359–373. https://doi.org/10.1521/jscp.1986.4.3.359

Barnhart, W. R., Rivera, S., & Robinson, C. W. (2018). Effects of Linguistic Labels on Visual Attention in Children and Young Adults. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00358

Barrett, R. C. A., Poe, R., O’Camb, J. W., Woodruff, C., Harrison, S. M., Dolguikh, K., Chuong, C., Klassen, A. D., Zhang, R., Joseph, R. B., & Blair, M. R. (2022). Comparing virtual reality, desktop-based 3D, and 2D versions of a category learning experiment. *PLOS ONE*, *17*(10), e0275119. https://doi.org/10.1371/journal.pone.0275119

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, *144*(1), 77–110. https://doi.org/10.1037/bul0000130

Berry, D. C. (1983). Metacognitive Experience and Transfer of Logical Reasoning. *The Quarterly Journal of Experimental Psychology Section A*, *35*(1), 39–49. https://doi.org/10.1080/14640748308402115

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78*(1), 246–263. https://doi.org/10.1111/j.1467-8624.2007.00995.x

Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of Training: A Meta-Analytic Review. *Journal of Management*, *36*(4), 1065–1105. https://doi.org/10.1177/0149206309352880

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, *12*(12), 752–762. https://doi.org/10.1038/nrn3122

Bowman, C. R., Iwashita, T., & Zeithamova, D. (2022). The effects of age on category learning and prototype- and exemplar-based generalization. *Psychology and Aging*, *37*(7), 800–815. https://doi.org/10.1037/pag0000714

Bowman, C. R., Valdez, M. R., & Obarski, S. A. (2023). Learning new categories in older age: A review of theoretical perspectives and empirical findings. *Psychology and Aging*, *38*(3), 174–187. https://doi.org/10.1037/pag0000715

Burnette, J. L., Billingsley, J., Banks, G. C., Knouse, L. E., Hoyt, C. L., Pollack, J. M., & Simon, S. (2023). A systematic review and meta-analysis of growth mindset interventions: For whom, how, and why might such interventions work? *Psychological Bulletin*, *149*(3–4), 174–205. https://doi.org/10.1037/bul0000368

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies* (pp. ix, 819). Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Casadevante, C., Romero, M., Fernández-Marcos, T., & Hernández, J. M. (2019). Category Learning in Schoolchildren. Its Relation to Age, Academic Marks and Resolution Patterns. *The Spanish Journal of Psychology*, *22*, E48. https://doi.org/10.1017/sjp.2019.56

Chen, L., Meier, K. M., Blair, M. R., Watson, M. R., & Wood, M. J. (2013). Temporal characteristics of overt attentional behavior during category learning. *Attention, Perception, & Psychophysics*, *75*(2), 244–256. https://doi.org/10.3758/s13414-012-0395-8

Chikha, A. B., Khacharem, A., Trabelsi, K., & Bragazzi, N. L. (2021). The Effect of Spatial Ability in Learning From Static and Dynamic Visualizations: A Moderation Analysis in 6-Year-Old Children. *Frontiers in Psychology*, *12*, 583968. https://doi.org/10.3389/fpsyg.2021.583968

Clark, R. E. (1994). Media Will Never Influence Learning. *Educational Technology Research and Development*, *42*(2), 21–29.

Corriveau Lecavalier, N., Ouellet, É., Boller, B., & Belleville, S. (2018). Use of immersive virtual reality to assess episodic memory: A validation study in older adults. *Neuropsychological Rehabilitation*, *30*(3), 462–480. https://doi.org/10.1080/09602011.2018.1477684

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International*, *34*, 819B.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837. https://doi.org/10.1016/0042-6989(95)00294-4

Di Natale, A. F., Repetto, C., Riva, G., & Villani, D. (2020). Immersive virtual reality in K-12 and higher education: A 10-year systematic review of empirical research. *British Journal of Educational Technology*, *51*(6), 2006–2033. https://doi.org/10.1111/bjet.13030

Dolguikh, K., Tracey, T., & Blair, M. R. (2021). The ubiquity of selective attention in the processing of feedback during category learning. *PLOS ONE*, *16*(12), e0259517. https://doi.org/10.1371/journal.pone.0259517

Duffy, G., Sorby, S., Reves, P. R., Delahunty, T., Perez, L., & Ravishankar, J. (2018). The Link between Spatial Skills and Engineering Problem-Solving. *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 272–278. https://doi.org/10.1109/TALE.2018.8615193

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*, 256–273. https://doi.org/10.1037/0033-295X.95.2.256

Epler-Ruths, C. M., McDonald, S., Pallant, A., & Lee, H.-S. (2020). Focus on the notice: Evidence of spatial skills' effect on middle school learning from a computer simulation. *Cognitive Research: Principles and Implications*, *5*, 61. https://doi.org/10.1186/s41235-020-00263-0

Espinet, S. D., Graziosi, G., Toplak, M. E., Hesson, J., & Minhas, P. (2022). A Review of Canadian Diagnosed ADHD Prevalence and Incidence Estimates Published in the Past Decade. *Brain Sciences*, *12*(8), 1051. https://doi.org/10.3390/brainsci12081051

Faraone, S. V., Banaschewski, T., Coghill, D., Zheng, Y., Biederman, J., Bellgrove, M. A., Newcorn, J. H., Gignac, M., Al Saud, N. M., Manor, I., Rohde, L. A., Yang, L., Cortese, S., Almagor, D., Stein, M. A., Albatti, T. H., Aljoudi, H. F., Alqahtani, M. M. J., Asherson, P., … Wang, Y. (2021). The World Federation of ADHD International Consensus Statement: 208 Evidence-based conclusions about the disorder. *Neuroscience & Biobehavioral Reviews*, *128*, 789–818. https://doi.org/10.1016/j.neubiorev.2021.01.022

Feldman, J. (2021). Mutual Information and Categorical Perception. *Psychological Science*, *32*(8), 1298–1310. https://doi.org/10.1177/0956797621996663

Filoteo, J. V., & Maddox, W. T. (2004). A quantitative model-based approach to examining aging effects on information-integration category learning. *Psychology and Aging*, *19*(1), 171–182. https://doi.org/10.1037/0882-7974.19.1.171

Fisher, S. C. (1916). The process of generalizing abstraction; and its product, the general concept. *The Psychological Monographs*, *21*(2), i–218. https://doi.org/10.1037/h0093097

Fisk University, HTC VIVE, T-Mobile and VictoryXR Launch 5G-Powered VR Human Cadaver Lab. (2021, August 3). *Fisk University*. https://www.fisk.edu/university-news-and-publications/fisk-university-htc-vive-t-mobile-and-victoryxr-launch-5g-powered-vr-human-cadaver-lab/

Gabay, Y., & Goldfarb, L. (2017). Feedback-based probabilistic category learning is selectively impaired in attention/hyperactivity deficit disorder. *Neurobiology of Learning and Memory*, *142*, 200–208. https://doi.org/10.1016/j.nlm.2017.04.012

Gauthier, I., & Tarr, M. J. (1997). Becoming a “Greeble” Expert: Exploring Mechanisms for Face Recognition. *Vision Research*, *37*(12), 1673–1682. https://doi.org/10.1016/S0042-6989(96)00286-6

Gittinger, M., & Wiesche, D. (2023). Systematic review of spatial abilities and virtual reality: The role of interaction. *Journal of Engineering Education*. https://doi.org/10.1002/jee.20568

Gouravajhala, R., Wahlheim, C. N., & McDaniel, M. A. (2020). Individual and age differences in block-by-block dynamics of category learning strategies. *Quarterly Journal of Experimental Psychology*, *73*(4), 578–593. https://doi.org/10.1177/1747021819892584

Grace, R. C., Sargisson, R. J., & White, K. G. (2012). Evidence for a magnitude effect in temporal discounting with pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*(1), 102–108. https://doi.org/10.1037/a0026345

Hammer, R., Sloutsky, V., & Grill-Spector, K. (2012). The Interplay between Feature-Saliency and Feedback Information in Visual Category Learning Tasks. *CogSci ... Annual Conference of the Cognitive Science Society. Cognitive Science Society (U.S.). Conference*, *2012*, 420–425.

Hartley, J., & Homa, D. (1981). Abstraction of stylistic concepts. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(1), 33–46. https://doi.org/10.1037/0278-7393.7.1.33

Höffler, T. N., & Leutner, D. (2011). The role of spatial ability in learning from instructional animations – Evidence for an ability-as-compensator hypothesis. *Computers in Human Behavior*, *27*(1), 209–216. https://doi.org/10.1016/j.chb.2010.07.042

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, *57*(6), 787–795. https://doi.org/10.3758/BF03206794

Homa, D., Rhoads, D. J., & Chambliss, D. F. (1979). Evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning & Memory*, *5*, 11–23.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(6), 418–439. https://doi.org/10.1037/0278-7393.7.6.418

Hou, L., Chi, H.-L., Tarng, W., Chai, J., Panuwatwanich, K., & Wang, X. (2017). A framework of innovative learning for skill development in complex operational tasks. *Automation in Construction*, *83*, 29–40. https://doi.org/10.1016/j.autcon.2017.07.001

Huang-Pollock, C. L., Maddox, W. T., & Tam, H. (2014). Rule-based and information-integration perceptual category learning in children with attention-deficit/hyperactivity disorder. *Neuropsychology*, *28*(4), 594–604. https://doi.org/10.1037/neu0000075

Hughes, G. I., & Thomas, A. K. (2021). Visual category learning: Navigating the intersection of rules and similarity. *Psychonomic Bulletin & Review*, *28*(3), 711–731. https://doi.org/10.3758/s13423-020-01838-0

Hull, C. L. (1920). Quantitative aspects of evolution of concepts: An experimental study. *Psychological Monographs*, *28*(1), i–86. https://doi.org/10.1037/h0093130

Jensen, L., & Konradsen, F. (2017). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, *23*(4), 1515–1529. https://doi.org/10.1007/s10639-017-9676-0

Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & de Haan, E. H. F. (2000). The Corsi Block-Tapping Task: Standardization and Normative Data. *Applied Neuropsychology*, *7*(4), 252–258. https://doi.org/10.1207/S15324826AN0704_8

Khalil, M. K., & Elkhider, I. A. (2016). Applying learning theories and instructional design models for effective instruction. *Advances in Physiology Education*, *40*(2), 147–156. https://doi.org/10.1152/advan.00138.2015

Klingenberg, S., JÃ¸rgensen, M. L. M., Dandanell, G., Skriver, K., Mottelson, A., & Makransky, G. (2020). Investigating the effect of teaching as a generative learning strategy when learning through desktop and immersive VR: A media and methods experiment. *British Journal of Educational Technology*, *51*(6), 2115–2138. https://doi.org/10.1111/bjet.13029

Kourtesis, P., Korre, D., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2020). Guidelines for the Development of Immersive Virtual Reality Software for Cognitive Neuroscience and Neuropsychology: The Development of Virtual Reality Everyday Assessment Lab (VR-EAL), a Neuropsychological Test Battery in Immersive Virtual Reality. *Frontiers in Computer Science*, *1*. https://doi.org/10.3389/fcomp.2019.00012

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*(13), 1897–1916. https://doi.org/10.1016/0042-6989(94)00279-U

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44. https://doi.org/10.1037/0033-295X.99.1.22

Kühl, T., Fehringer, B. C. O. F., & Münzer, S. (2022). Unifying the Ability-as-Compensator and Ability-as-Enhancer Hypotheses. *Educational Psychology Review*, *34*(2), 1063–1095. https://doi.org/10.1007/s10648-021-09650-5

Lee, E. A.-L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers & Education*, *79*, 49–58. https://doi.org/10.1016/j.compedu.2014.07.010

Li, G., Anguera, J. A., Javed, S. V., Khan, M. A., Wang, G., & Gazzaley, A. (2020). Enhanced Attention Using Head-mounted Virtual Reality. *Journal of Cognitive Neuroscience*, *32*(8), 1438–1454. https://doi.org/10.1162/jocn_a_01560

Likens, S., & Mower, A. (n.d.). *What does virtual reality and the metaverse mean for training?* PwC. Retrieved April 25, 2024, from https://www.pwc.com/us/en/tech-effect/emerging-tech/virtual-reality-study.html

Little, J. L., & McDaniel, M. A. (2015). Individual differences in category learning: Memorization versus rule abstraction. *Memory & Cognition*, *43*(2), 283–297. https://doi.org/10.3758/s13421-014-0475-1

Makransky, G., Andreasen, N., Baceviciute, S., & Mayer, R. (2021). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*, *113*(4), 719–735. https://doi.org/10.1037/edu0000473

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educational Psychology Review*, *33*(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

Massen, C., Vaterrodt-Plünnecke, B., Krings, L., & Hilbig, B. E. (2009). Effects of instruction on learners' ability to generate an effective pathway in the method of loci. *Memory*, *17*(7), Article 7. https://doi.org/10.1080/09658210903012442

Mathews, R. C., Stanley, W. B., Buss, R. R., & Chinn, R. A. (1984). Concept Learning: What Happens When Hypothesis Testing Fails? *The Journal of Experimental Education*, *53*(2), 91–96.

Mayer, R. E. (Ed.). (2014). *The Cambridge Handbook of Multimedia Learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, *86*(3), 389–401. https://doi.org/10.1037/0022-0663.86.3.389

McColeman, C. M., Barnes, J. I., Chen, L., Meier, K. M., Walshe, R. C., & Blair, M. R. (2014). Learning-Induced Changes in Attentional Allocation during Categorization: A Sizable Catalog of Attention Change as Measured by Eye Movements. *PLoS ONE*, *9*(1), e83302. https://doi.org/10.1371/journal.pone.0083302

McColeman, C. M., Thompson, J., Anvari, N., Azmand, S. J., Barnes, J., Barrett, R. C. A., Byliris, R., Chen, Y., Dolguikh, K., Fischler, K., Harrison, S., Hayre, R. S., Poe, R., Swanson, L., Tracey, T., Volkanov, A., Woodruff, C., Zhang, R., & Blair, M. (2020). Digit eyes: Learning-related changes in information access in a computer game parallel those of oculomotor attention in laboratory studies. *Attention, Perception, & Psychophysics*, *82*(5), 2434–2447. https://doi.org/10.3758/s13414-020-02019-w

McDonald, M. M., Defever, A. M., & Navarrete, C. D. (2017). Killing for the greater good: Action aversion and the emotional inhibition of harm in moral dilemmas. *Evolution and Human Behavior*, *38*(6), 770–778. https://doi.org/10.1016/j.evolhumbehav.2017.06.001

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295x.85.3.207

Meier, K. M., & Blair, M. R. (2013). Waiting and weighting: Information sampling is a balance between efficiency and error-reduction. *Cognition*, *126*(2), 319–325. https://doi.org/10.1016/j.cognition.2012.09.014

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*, *140*, 103603. https://doi.org/10.1016/j.compedu.2019.103603

Midkiff, B., Langer, M., Demetriou, C., & Panter, A. T. (2018). An IRT Analysis of the Growth Mindset Scale. In M. Wiberg, S. Culpepper, R. Janssen, J. González, & D. Molenaar (Eds.), *Quantitative Psychology* (pp. 163–174). Springer International Publishing. https://doi.org/10.1007/978-3-319-77249-3_14

Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, *56*(3), 769–780. https://doi.org/10.1016/j.compedu.2010.10.020

Morgan, P., Patrick, J., & Patrick, T. (2010). Increasing information access cost to protect against interruption effects during problem solving. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *32*, 949–954. https://escholarship.org/uc/item/6kp074gs

Muller Queiroz, A. C., Nascimento, A. M., Tori, R., Alejandro, T. B., Melo, V. V. de, Meirelles, F. de S., & Leme, M. I. da S. (2018). Immersive Virtual Environments in Corporate Education and Training. *AMCIS 2018 Proceedings*. https://aisel.aisnet.org/amcis2018/Education/Presentations/12

Nejati, V. (2021). Effect of stimulus dimension on perception and cognition. *Acta Psychologica*, *212*, 103208. https://doi.org/10.1016/j.actpsy.2020.103208

Neuburger, S., Ruthsatz, V., Jansen, P., & Quaiser-Pohl, C. (2015). Can girls think spatially? Influence of implicit gender stereotype activation and rotational axis on fourth graders' mental-rotation performance. *Learning and Individual Differences*, *37*, 169–175. https://doi.org/10.1016/j.lindif.2014.09.003

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206. https://doi.org/10.1016/j.actpsy.2013.06.003

Oser, R., & Fraser, B. J. (2015). Effectiveness of Virtual Laboratories in Terms of Learning Environment, Attitudes and Achievement among High-School Genetics Students. *Curriculum and Teaching*, *30*(2), 65–80. https://doi.org/10.7459/ct/30.2.05

Pajares, F., & Schunk, D. H. (2002). Self and self-belief in psychology and education: A historical perspective. In *Improving academic achievement: Impact of psychological factors on education* (pp. 3–21). Academic Press. https://doi.org/10.1016/B978-012064455-1/50004-X

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74–98. https://doi.org/10.1016/j.edurev.2017.08.004

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, *110*(6), 785–797. https://doi.org/10.1037/edu0000241

Parsons, T. D. (2015). Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00660

Pellas, N., Dengel, A., & Christopoulos, A. (2020). A Scoping Review of Immersive Virtual Reality in STEM Education. *IEEE Transactions on Learning Technologies*, *13*(4), 748–761. https://doi.org/10.1109/TLT.2020.3019405

Pérez-Gay Juárez, F., Thériault, C., Gregory, M., Rivas, D., Sabri, H., & Harnad, S. (2017). How and Why Does Category Learning Cause Categorical Perception? *International Journal of Comparative Psychology*, *30*(0). https://escholarship.org/uc/item/8rg6c087

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A Redrawn Vandenberg and Kuse Mental Rotations Test—Different Versions and Factors That Affect Performance. *Brain and Cognition*, *28*(1), 39–58. https://doi.org/10.1006/brcg.1995.1032

Petersen, G. B., Klingenberg, S., Mayer, R. E., & Makransky, G. (2020). The virtual field trip: Investigating how to optimize immersive virtual learning in climate change education. *British Journal of Educational Technology*, *51*(6), 2099–2115. https://doi.org/10.1111/bjet.12991

Petersen, G. B., Stenberdt, V., Mayer, R. E., & Makransky, G. (2023). Collaborative generative learning activities in immersive virtual reality increase learning. *Computers & Education*, *207*, 104931. https://doi.org/10.1016/j.compedu.2023.104931

R Core Team. (2020). *R: A language and environment for statistical computing* (4.0.0) [Computer software]. https://www.R-project.org/

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, *147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Rajsic, J., Wilson, D. E., & Pratt, J. (2018). The price of information: Increased inspection costs reduce the confirmation bias in visual search. *Quarterly Journal of Experimental Psychology*, *71*(4), 832–849. https://doi.org/10.1080/17470218.2016.1278249

Rayner, K., McConkie, G. W., & Ehrlich, S. (1978). Eye movements and integrating information across fixations. *Journal of Experimental Psychology. Human Perception and Performance*, *4*(4), 529–544. https://doi.org/10.1037//0096-1523.4.4.529

Reetzke, R., Maddox, W. T., & Chandrasekaran, B. (2016). The role of age and executive function in auditory category learning. *Journal of Experimental Child Psychology*, *142*, 48–65. https://doi.org/10.1016/j.jecp.2015.09.018

Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1–41. https://doi.org/10.1016/j.cogpsych.2004.11.001

Rehder, B., & Hoffman, A. B. (2005b). Thirty-Something Categorization Results Explained: Selective Attention, Eyetracking, and Models of Category Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 811–829. https://doi.org/10.1037/0278-7393.31.5.811

Romppel, M., Herrmann-Lingen, C., Wachter, R., Edelmann, F., Düngen, H.-D., Pieske, B., & Grande, G. (2013). A short form of the General Self-Efficacy Scale (GSE-6): Development, psychometric properties and validity in an intercultural non-clinical sample and a sample of patients at risk for heart failure. *GMS Psycho-Social-Medicine*, *10*.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Symposium on Eye Tracking Research & Applications - ETRA '00*, 71–78. https://doi.org/10.1145/355017.355028

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The Relationship between Eye Movements and Spatial Attention. *The Quarterly Journal of Experimental Psychology Section A*, *38*(3), 475–491. https://doi.org/10.1080/14640748608401609

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., & Sanchez-Vives, M. V. (2006). A Virtual Reprise of the Stanley Milgram Obedience Experiments. *PLoS ONE*, *1*(1), e39. https://doi.org/10.1371/journal.pone.0000039

Smith, D. J., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 3–27. https://doi.org/10.1037/0278-7393.26.1.3

Smith, S. A. (2019). Virtual reality in episodic memory research: A review. *Psychonomic Bulletin & Review*, *26*(4), 1213–1237. https://doi.org/10.3758/s13423-019-01605-w

Soranzo, A., Lugrin, J.-L., & Wilson, C. J. (2013). The effects of belongingness on the Simultaneous Lightness Contrast: A virtual reality study. *Vision Research*, *86*, 97–106. https://doi.org/10.1016/j.visres.2013.04.012

Soranzo, A., & Wilson, C. J. (2014). *Virtual environments in visual perception: Applications and challenges*. https://www.semanticscholar.org/paper/Virtual-environments-in-visual-perception%3A-and-Soranzo-Wilson/6362cde9fd4c62246f5b94918e830c098e483ffc

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

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. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Thompson, J., O'Camb, J. W., Barrett, R. C. A., Harrison, S., & Blair, M. R. (2023). Robustness of performance during domain change in an esport: A study of within-expertise transfer. *PLOS ONE*, *18*(12), e0295037. https://doi.org/10.1371/journal.pone.0295037

Trabasso, T., & Bower, G. (1964). Memory in concept identification. *Psychonomic Science*, *1*(1–12), 133–134. https://doi.org/10.3758/BF03342827

Unity Technologies. (2020). *Unity (2020.1.6f1)* [Computer software]. https://unity.com/

Usher, E. L., & Pajares, F. (2008). Self-Efficacy for Self-Regulated Learning: A Validation Study. *Educational and Psychological Measurement*, *68*(3), 443–463. https://doi.org/10.1177/0013164407308475

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604. https://doi.org/10.2466/pms.1978.47.2.599

Virtual Reality for Schools. (n.d.). ClassVR. Retrieved April 25, 2024, from https://www.classvr.com/

Wahlheim, C., Mcdaniel, M., & Little, J. (2016). Category Learning Strategies in Younger and Older Adults: Rule Abstraction and Memorization. *Psychology and Aging*, *31*. https://doi.org/10.1037/pag0000083

Wang, D.-Y., Lee, M.-H., & Sun, C.-T. (2013). Effects of Thinking Style and Spatial Ability on Anchoring Behavior in Geographic Information Systems. *Educational Technology & Society*, *16*(3), Article 3.

Watson, M. R., & Blair, M. (2008). *Attentional Allocation During Feedback: Eyetracking Adventures on the Other Side of the Response*. https://www.semanticscholar.org/paper/Attentional-Allocation-During-Feedback%3A-Eyetracking-Watson-Blair/dd849a01ffdad0a5c4132211100816117d750bc0

Weng, C., Rathinasabapathi, A., Weng, A., & Zagita, C. (2019). Mixed Reality in Science Education as a Learning Support: A Revitalized Science Book. *Journal of Educational Computing Research*, *57*(3), 777–807. https://doi.org/10.1177/0735633118757017

Williams, G. F. (1971). A model of memory in concept learning. *Cognitive Psychology*, *2*(2), 158–184. https://doi.org/10.1016/0010-0285(71)90007-7

Yang, X., Park, T., Wickens, C. D., Siah, K. T. H., Fong, L., & Yin, S. Q. (2013). The effect of information access cost and overconfidence bias on junior doctors' pre-handover performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 1755–1759. https://doi.org/10.1177/1541931213571391

Yang, X., Wickens, C. D., Park, T., Fong, L., & Siah, K. T. H. (2015). Effects of Information Access Cost and Accountability on Medical Residents' Information Retrieval Strategy and Performance During Prehandover Preparation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(8), 1459–1471. https://doi.org/10.1177/0018720815598889

Appendix. Survey Presented to Participants

This appendix contains all questions and instruments that were administered throughout the course of this experiment. The order of the instruments is as follows:

A. Demographics Survey

B. Spatial Ability Measurement

C. Working Memory Measurement

D. Self Efficacy Measurement

E. Achievement Mindset Preferences

A. Demographics Survey

1. What is your age? [Numeric Value: Years]
2. How would you describe your sex?
    a. Male
    b. Female
    c. Intersex
    d. Prefer not to answer
3. How often do you play video games <u>per week</u> (console, PC, handheld, mobile, etc.)?
    a. 0-3 hours
    b. 3-6 hours
    c. 7-10 hours
    d. 10-15 hours
    e. 16+ hours
4. How often do you use immersive virtual reality technology (HTC Vive, Oculus Rift, Valve Index, Samsung VR, etc.)?
    a. I have never tried it before
    b. I’ve tried it once or a few times before
    c. repeatedly, but less than once a month
    d. monthly
    e. weekly

f. daily

5. Do any of the following eye conditions apply to you [check all that apply]?

    a. I wear glasses to see

    b. I wear contact lenses

    c. I have some form of colour-blindness

    d. Other condition not listed here

6. Have you ever received a diagnosis for ADHD, or any other attention-related disorder or neurodiversity?

    a. No

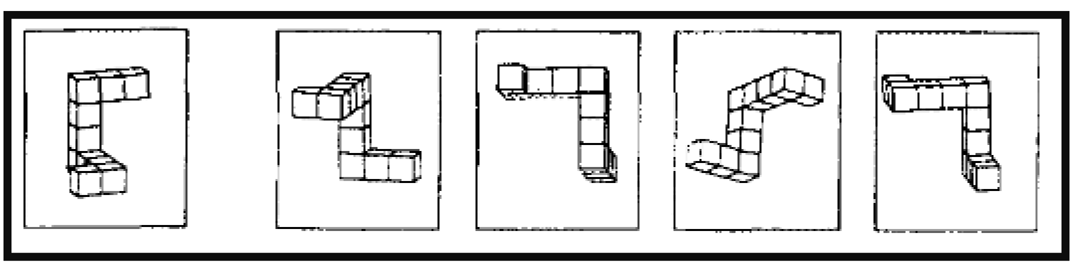
    b. Yes – ADHD

    c. Yes – Other [Please Describe]

    d. Prefer not to answer

B. Spatial Ability Measurement

To assess spatial ability, the Vandenberg & Kuse Mental Rotation Test (1978) as redrawn by Peters et al. (1995) is used to capture a participant's ability to do mental rotation. For each trial of this task, a 2D representation of 3D shape is presented to the participant to be compared against 4 other shapes on the same page. One of these shapes is a rotated version of the first shape, and it is up to the participant to correctly identify this shape. An example problem from the full set appears below, with the target shape on the far left. The participant must choose which 1 of the 4 shapes on the right is a rotated version of the first shape. All other questions on this test use similar stimuli.



C. Working Memory / Attention Measurement

Using the Corsi Block-Tapping Task (a visual spatial version of digit span task as described in Wilhelm et al. (2013)), participants are asked to observe a series of lights presented one at a time at various locations on a screen and recall the sequence in exact order immediately afterwards. Starting with just 2 lights in a sequence, each time the participant remembers the sequence correctly, an additional light is added until the participant makes a mistake. When a mistake is made, the current length is repeated, and if a second mistake is made, this triggers the end of the task. The person's working memory span is defined as the maximum length correctly recalled.

D. Self Efficacy Measurement

The following Generalized Self-Efficacy scale comes from Schwarzer & Jerusalem (1995), as adapted by Romppel et al. (2013):

Indicate for each statement below how true it is for you.

1. If someone opposes me, I can find the means and ways to get what I want.
2. I am certain that I can accomplish my goals.
3. I am confident that I could deal efficiently with unexpected events.
4. Thanks to my resourcefulness, I can handle unforeseen situations.
5. I can remain calm when facing difficulties because I can rely on my coping abilities.
6. I can handle whatever comes my way.

Responses to all prompts are made using the following scale:

not at all true

hardly true

moderately true

exactly true

E. Achievement Mindset Preferences

The following scale of the Growth Mindset was adapted and validated by Midkiff et al. (2018), based on the work of Dweck & Leggett (1988).

How much do you personally agree or disagree with the following statements?

1. You have a certain amount of intelligence, and you can't really do much to change it.
2. Your intelligence is something about you that you can't change very much.
3. No matter who you are, you can significantly change your intelligence level.
4. To be honest, you can't really change how intelligent you are.
5. You can always substantially change how intelligent you are.
6. You can learn new things, but you can't really change your basic intelligence.
7. No matter how much intelligence you have, you can always change it quite a bit.
8. You can change even your basic intelligence level considerably.

Responses to all prompts are made using the following scale:

strongly disagree

somewhat disagree

neither agree or disagree

somewhat agree

strongly agree