THE UNIVERSITY OF CHICAGO

DOES ACTIVE EXPERIENCE SUPPORT CHILDREN'S LEARNING FROM INSTRUCTION?

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY

NATALIE BREZACK

CHICAGO, ILLINOIS

JUNE 2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

In my research, I examine how children learn on their own while receiving guidance from other people. While the research that went into this dissertation included opportunities for independent learning, I could not have achieved this milestone without the support of my teachers, both formal and informal. Huge thanks are owed to my advisor, Amanda Woodward, for guiding me through my PhD and pushing me to work to my fullest potential. Through your guidance, I have become a more rigorous scientist; you have taught me to design compelling studies and engage meaningfully with theory, and you have helped me develop as a writer and a thinker. In addition, I would like to thank my committee members, Susan Levine and Susan Goldin-Meadow, for your support through this process. I have learned so much from your thoughtful questions about my research and from your inspiring course work.

I would like to thank my lab mates, current and past, for providing a community where conducting research was collaborative and fun: Marlene Meyer, Haerin Chung, Tess Fulcher, Nicole Burke, Grace Hwang, Marc Colomer, and Carol Medina. Your feedback has helped me grow as a scientist. In particular, Marlene, Haerin, and Tess: You have also supported me on a personal level, and I always know I can turn to you to talk through challenges. I would like to thank the honors thesis students (Mia Radovanovic, Sarah Pan, Daryn Ellison), undergraduate research assistants (Lauren Kingsly, Anya Samtani, Melissa Gaviria, Jess Chandler), and lab managers past and present (Annika Hendrickson, Riley Abeles, Kayleigh Watson, Marla Jean Douma, and Laura Tharsen) for their tireless efforts supporting these projects, including recruiting and testing participants and coding hours of videos. In addition, I would like to thank Dain Woods for building and rebuilding the Lock and Keys Task following many rounds of changes. Thank you as well to the families and children who participated in these studies.

Finally, I would like to thank my family and friends. To my parents, Laura and Jim Brezack, sister Hannah, and grandma Barbara Grant: Words cannot express how much your support has meant to me throughout this challenging PhD journey. Your unwavering assurances that I can finish this dissertation helped me cross the finish line. To my fiancé Adam LaMarra: You have been an incredibly supportive and patient partner during this process, and you became the best coworker during COVID-19. Your support means everything to me – I could not have done this without you. And to my friends Emily Steiner, Julia Stafford, Elise Gaffney, Ellen Ingamells, Wes and Caylee Mercado, Ed Hendel, and Nicole Wong-Homer: I feel so fortunate to have a supportive community around me that I can always turn to and learn from.

ABSTRACT

From early in life, young children learn conventional information, including artifact uses and formal systems like mathematics. Children's active experience has long been considered beneficial for learning (Piaget, 1953). However, for some types of learning, children require instruction from an expert; instruction transmits information efficiently and may provide children with all the information they need to learn (Kirschner et al., 2006). When children receive instruction, does active experience improve their learning? Children may engage actively while they are instructed, or children may actively explore new material before or after instruction. Active experience can involve several features, including physical engagement with materials as well as agency, which includes cognitive activities such as decision-making and hypothesis generation. These aspects of experience could, in turn, have implications for several aspects of learning, including mastery of taught information, generalization of learned rules or regularities to new problems, and long-term memory for the taught material.

In this dissertation, I report six preregistered experiments that examine whether and how active experience supports children's learning of conventional information in instructed contexts. In Chapter 2, I present two experiments investigating an early case of instructed learning, asking whether active engagement during instruction enhances toddlers' learning about assembling simple toys. In a series of four experiments in Chapters 3 and 4, I turn my attention to older children's, and adults', learning of a novel two-factor rule in a problem-solving game, asking whether active exploration before or after instruction improves learning. I find that active experience does not universally enhance learning. In some cases, however, active experience improved particular learning outcomes: Guided active experience supported children's long-term

memory for taught information (Chapter 2) and active experience following instruction improved children's generalization of learned rules to new contexts (Chapters 3 and 4).

Chapter 1: Introduction

In the field of developmental psychology, active experience has been described as central for young children's learning. This seminal idea was formulated by Piaget (1953, 1964) and Montessori (1918), who described the fundamental contributions of active experience to children's conceptual development and learning. Active experience's effects on learning are evident across a range of approaches in the current literature, including work that stresses the importance of hypothesis testing (Gopnik & Wellman, 2012), active explanation (Danovitch & Mills, 2018), productive failure (DeCaro & Rittle-Johnson, 2012), and curiosity-based learning (Gruber et al., 2019; Jirout & Klahr, 2012). This idea has also informed classroom curricula for children, which highlight the benefits of children's active engagement for learning (Lillard, 2016; Hmelo-Silver et al., 2007; Alfieri et al., 2011).

In contrast, in many situations, children can learn from instruction provided by knowledgeable adults, without needing to discover the answers themselves. This phenomenon is seen across studies on infants' robust ability to learn by observing adult action demonstrations (Meltzoff, 1988; Bauer, 1996; Fagard & Lockman, 2010) and research documenting the importance of pedagogical instruction for supporting learning (Csibra & Gergely, 2009). Indeed, classroom curricula that rely on instruction enhance students' learning, often more effectively than students' unguided exploration of new materials (Kirschner et al., 2006; Stockard et al., 2018; Mayer, 2004). Much of what children work on learning from a young age is conventional information, such as learning about objects in children's environments and formal systems like mathematics (Keen, 2011; Tomasello, 2001). Conventional information can be challenging or impossible for children to discover independently; instruction may be necessary for learning (Mayer, 2004) or would at least make learning more efficient (Kirschner et al., 2006).

In this dissertation, I ask whether children who receive instruction about conventional information learn more effectively if they are actively engaged with the material to be learned. Active experience involves multiple components which may support learning, including physically engaging with the materials and acting with a sense of agency, which involves higher-order cognitive processes such as hypotheses testing, decision making, and discovery. Children can be active while receiving instruction, or children may actively explore the material to be learned before or after instruction. In turn, learning can be measured in multiple ways including learning of taught information, generalizing or applying learned information to new contexts, and storing information in long-term memory.

Physical Activity

The most concrete property of active experience is its physicality. Physical activity involves motor components (reaching, grasping), perceptual components (touch, visual processing), and sensorimotor integration (integrating visual and motor cues; Rochat, 1989). These perceptual and motor systems are coordinated with cognitive systems (Thelen, 2000) including attention (Boudreau & Bushnell, 2000; Amso & Scerif, 2015), action planning (McCarty et al., 1999; Barrett et al., 2007), and enacting motor commands (Trewartha et al. 2015). A number of theoretical perspectives on "embodied cognition" have suggested that bodily engagement influences children's thinking, learning, and remembering (Wilson, 2002; Adolph & Joh, 2007; Kontra et al., 2012; Adams, 2010).

Some evidence suggests that physical activity affects children's learning while children receive instruction. Four-year-old children who took turns with an experimenter to perform instructed actions learned and remembered the actions more accurately than actions performed by either person separately (Sommerville & Hammond, 2007). However, beyond the work of

Sommerville and Hammond (2007), research has yet to examine the potential effects of physical activity on learning while children are instructed. Research with infants suggests that physical activity can support conceptual learning (e.g., addition and subtraction: Lubin et al., 2015; mental rotation: Mohring & Frick, 2013; rhythm: Phillips-Silver & Trainor, 2005; audiovisual synchrony: Gerson et al., 2015; goal understanding: Sommerville et al., 2008). More research is necessary to examine whether instructed, physically-active experience supports children's learning of taught information more effectively than comparable non-active experiences.

Agency

When children act on their environments, they often have agency over the actions they perform: Children are "in the driver's seat." Acting with agency, or exploring, involves a number of cognitive activities that might support learning. These activities include searching for information in a goal-directed (Saylor & Ganea, 2018) or curiosity-driven way (Gruber et al., 2019; Jirout & Klahr, 2012) to resolve information gaps (Sobel & Letourneau, 2018; Danovich & Mills, 2018). Children choose what to learn (Lucca & Wilbourn, 2018; Begus et al., 2014) and control the learning environment and the flow of information (Gureckis & Markant, 2012). Children may also encounter failure (DeCaro & Rittle-Johnson, 2012) and make discoveries (Dean & Kuhn, 2007). Indeed, control over the learning process is a central tenant of guided play, or child-led activities with learning goals scaffolded by adults (Yu et al., 2018; Zosh et al., 2018).

Specifically, some of the effects of agency on learning may be due to hypothesis testing (Gopnik & Wellman, 2012). Children learn about cause and effect by performing actions and observing other people perform actions. From a Bayesian perspective, children have a hypothesis space, or set of priors, when they perform actions to generate causal evidence (Gopnik &

Wellman, 2012). Children have different priors when they act as agents performing actions versus situations where children watch other people perform actions (Schulz, Kushnir, & Gopnik, 2007). This can make children's own actions more effective for learning than receiving expert instruction (Yang, 2017) because children weigh their own actions more heavily than someone else's actions, showing a "self-agency bias" (Kushnir et al., 2009). Indeed, when children learned about an ambiguous causal system, they were more likely to believe that their own actions, rather than someone else's actions, caused the effects (Kushnir et al., 2009); children rely on their own actions when learning.

Children's active exploration during instruction has been found to support learning. Four- to five-year-old children learned and remembered shape concepts more accurately after a guided play activity compared to free play or direct instruction (Fisher et al., 2013). Adding a guided play activity to a vocabulary lesson enhanced children's vocabulary growth (Han et al., 2010). Similarly, children learned more about how gears worked after their caregivers guided them to explore a museum exhibit (Callanan et al., 2020). In classroom settings, guided play (Weisberg et al., 2016), active exploration of materials (e.g., the Montessori method; Lillard, 2016), guided discovery learning (Alfieri et al., 2011), and scaffolded, self-directed learning (Hmelo-Silver et al., 2007) have been theorized to improve children's learning.

However, relatively little research has examined whether children's active exploration before or after instruction similarly supports learning. One such study examined whether four-year-old children learned the causal rules governing a light box when they explored before or after seeing an experimenter's demonstration (Sobel & Sommerville, 2010). Children who explored before the demonstration learned the causal rules more effectively than those who explored after instruction (Sobel & Sommerville, 2010). Exploring prior to receiving instructions allowed children to independently discover the information to be learned; indeed, preschoolers who discovered causal information learned it effectively (Schulz, Gopnik, & Glymour, 2007).

In addition, exploring after instruction may also support young children's learning. Three- to-five-year-old children who explored a novel causal system (a box that could be activated by certain blocks) after seeing a demonstration applied the causal rules to new blocks more accurately than those who were instructed without an opportunity to explore (Sim et al., 2017). Instruction can target children's behavior towards specific actions (Bonawitz et al., 2011), reducing the demands of unguided exploration (Sweller et al., 1998). However, more research is needed in addition to that of Sobel and Sommerville (2010) and Sim and colleagues (2017) to examine whether active exploration before or after instruction supports young children's learning.

Exploring before or after instruction has been found to improve older children's learning of academic topics in school settings. Second- to fourth-grade students who explored mathematics equivalence problems prior to receiving instruction had better conceptual understanding of mathematics equivalence than students who had been instructed before exploring (DeCaro & Rittle-Johnson, 2012). Similar results were found when students learned about variance (Kapur, 2012). Exploration before instruction can allow students to engage in "productive failure," where learners compare the errors they made while exploring to subsequent instruction (Loibl & Rummel, 2013). In addition, while third- and fourth-grade students learned the control-of-variables strategy to design science experiments from instruction better than unguided exploration (Klahr & Nigam, 2004), when students had more time to explore, they learned more effectively from exploration after instruction or exploration alone compared to instruction alone (Dean & Kuhn, 2007). Additional research is needed to examine whether active exploration before or after instruction supports learning in young children.

Learning Outcomes

Features of active experience, including physicality and agency, have been linked to children's learning. However, there are several ways to define learning; in addition to measuring children's learning of taught information, learning can also be measured as children's ability to generalize taught information to new contexts and store information in long-term memory. Much of the research on young children's active learning has measured how well children learned the taught information (e.g., Sobel & Sommerville, 2010) without testing whether children could apply what they had learned to new situations or remember learned information after a delay.

Generalization is considered a hallmark of effective learning; "generative learning" involves transforming learned information into knowledge that children can apply to new situations (Fiorella & Mayer, 2016). Rather than merely representing learned information, generalization includes mental models, schemas (Fiorella & Mayer, 2016), and conceptual or symbolic information (Bruner, 1966) that can be applied more broadly to new contexts. Active experience has been theorized to support children's generalization (Hmelo-Silver et al., 2007; Alfieri et al., 2011).

In addition, studies rarely test children's long-term memory for taught information. Physically-active experience has been found to promote memory: The "motor-induced encoding effect" proposes that the degree to which learners engage in motor planning and execution affects the strength of encoded memories (Kinder & Buss, 2021). Indeed, when infants performed actions, they remembered the actions for months (Bauer et al., 1994). Similarly, children had better action recall after a 4-month delay for actions they had taken turns performing with a partner compared to actions performed separately by each person (Sommerville & Hammond, 2007). In addition to testing children's mastery of learned information, testing children's generalization and long-term memory can potentially reveal the strength and lasting impacts of active experience on learning.

Dissertation Overview

In six preregistered studies in this dissertation, I address gaps in the prior literature regarding the effects of active experience on learning in young children. In Chapter 2 (Studies 1 and 2), I ask whether the physical component of active experience supports children's learning in contexts where children receive instruction while they act. My prior research (Brezack et al., 2021) began to answer this question. In Brezack et al. (2021), I examined how toddlers learned to assemble novel toys through everyday instruction provided by their caregivers. Caregivers first taught their toddlers to perform target actions to construct the toys. An experimenter who was not present during the teaching phase then tested children on their action learning. Detailed behavioral coding captured children's active experience (child actions), observational experience (caregiver actions), and caregiver instructions (speech, including directions and praise). I found that while toddlers were instructed, those who were given more active experience performing the actions demonstrated better learning. In contrast, viewing more caregiver action demonstrations negatively related to children's learning. Children were therefore central agents of their own action learning as they received instruction.

However, the instructed experiences in Brezack et al. (2021) were a mix of opportunities for children to be active and to observe their caregivers' actions. In Chapter 2 (Studies 1 and 2) of this dissertation, I follow up on the findings from Brezack et al. (2021) by testing whether instruction that allowed children to be active in performing target actions supported children's action learning more than instruction in which children observed a teacher perform the actions. Children's active experience was physical in nature: Children performed the actions themselves, providing an embodied experience with the information to be learned. However, children's actions were heavily guided by an experimenter; children had little control over the order or timing of their actions. I therefore tested whether instructed, physically-active experience supported learning compared to a matched observational experience. In addition, I measured learning in three ways: children's mastery of the taught actions, generalization of taught actions to new toys, and long-term memory for taught information.

In Chapters 3 and 4 (Studies 3, 4, 5, and 6), I ask whether children's active exploration before or after instruction affects learning. My colleagues and I began to answer this question in another recent study (Radovanovic et al., in preparation) in which we examined whether 6-year-old children who explored before receiving instruction learned more than children who received instruction without opportunities to explore, or only explored without instruction. In a novel problem-solving task (the Lock and Keys Task), children were tasked with determining the rule that governed which keys would unlock a set of locks. After exploration and instruction, instruction alone, or exploration alone, children were tested on their ability to unlock the locks as a test of immediate learning and were also tested on their ability to generalize the rule to a novel set of locks and keys. The results indicated that children learned best when they had been instructed, regardless of whether or not they had also explored prior to the instructions. Crucially, children who had explored prior to instruction generalized the rule more effectively than children who had not explored before instruction. This suggested that active exploration before instruction supported children's generalization of the rule.

In Study 3 (Chapter 3) of this dissertation, I examined whether children who participated in an online version of the problem-solving task would learn the rule under similar conditions as children in the prior in-person study. As in Radovanovic et al. (in preparation), children either explored before instruction, received instruction without exploring, or explored without instruction. Those who actively explored did so separately from the instruction they received, which allowed children to have agency over their actions. Children were then tested on their immediate learning and generalization of the unlocking rule. In addition, in Study 4, I tested a comparison sample of adult participants in the same task. I then compared the results from Study 3 to Radovanovic et al. (in preparation) to test whether the task context (in-person or online) affected learning.

Several questions remained unanswered from Radovanovic et al. (in preparation) and Studies 3 and 4, which I addressed in Chapter 4 (Studies 5 and 6). In particular, children in Study 3 explored prior to instruction; instruction followed by exploration might also promote learning. In addition, to understand whether active experience improved learning compared to a matched non-active experience, I implemented an observational learning comparison condition. There were also methodological challenges that informed the design of the experiments in Chapter 4. I carefully controlled the time during which children were exposed to the material to be learned across conditions, and I implemented a stringent test phase to assess children's learning. Study 5 tested these questions in children, and Study 6 tested a comparison sample of adults.

Across the six preregistered experiments presented in this dissertation (Table 1), children learned new information from physically-active experience during instruction (Chapter 2) and active exploration before or after instruction (Chapters 3 and 4). The effects of active experience on learning were compared with observational experience (Chapter 2, Chapter 4). Children's learning was measured as mastery of taught information (Chapters 2, 3, and 4), generalization to new contexts (Chapters 2, 3, and 4), and long-term memory (Chapter 2). Given the learning benefits of active experience that includes physical activity and agency, I hypothesized that children would learn better from opportunities to be active compared to matched observational experiences.

Type of activity → Timing of activity ↓	Physical activity	Exploration (physical activity & agency)	Matched observational experience
Concurrent with instruction	Chapter 2 *Learning, generalization, memory*		Chapter 2 *Learning, generalization, memory*
Before instruction		Chapter 3, Chapter 4 *Learning, generalization*	
After instruction		Chapter 4 *Learning, generalization*	Chapter 4 *Learning, generalization*
Alone (no instruction)		Chapter 3, Chapter 4 *Learning, generalization*	

Table 1: Dissertation Chapters. Types of active experience (top), timing of active experience relative to instruction (left), and comparison observational experience (right column) tested in each chapter of this dissertation with learning outcomes italicized (cells).

This dissertation begins to tackle some of the questions at the heart of understanding active experience's effects on learning. In particular, I experimentally address different types of active experience in different instructional contexts with multiple learning outcomes. The results of this dissertation shed light on the types of active engagement that matter for learning, while leaving open avenues for future research to continue examining how and why active experience may support children's learning in instructed contexts.

Chapter 2: Toddlers' Action Learning and Memory from Active and Observed Instruction

Children face the challenging task of learning the conventional actions of their culture. Objects designed by people (artifacts) have associated conventional actions meant to be performed on them (Legare, 2019; Tomasello, 1999). For example, keys are inserted into locks and twisted to unlock doors, Velcro straps are pressed together keep children's shoes on their feet, and Mr. Potato Head's arms are inserted into holes on his body to assemble him. Learning to perform conventional actions appropriately is an important skill for development that is unique to humans (Tomasello, 2001) and begins early in life (Keen, 2011). Children cannot learn conventional actions without receiving some form of input from experts; without support, children would not know which actions were culturally acceptable to perform with keys, Velcro straps, or Mr. Potato Head's arms. Children observe adults perform actions (Bandura, 1977; Rogoff, 1990) or receive direct instruction (Vygotsky, 1980; Wood et al., 1976) to learn conventional actions. Yet, children's active experience has been theorized as central for learning (Piaget, 1964). In the current studies, we evaluated whether, and how, active engagement may support children's learning about conventional actions from instruction.

Adults can support children's conventional action learning by providing demonstrations of appropriate actions. Children are facile imitators of adult action demonstrations from infancy (Meltzoff, 1988). Nine-month-old infants imitate simple actions on objects (Meltzoff, 1988), and twelve-month-old infants imitate two-step actions (Bauer, 1996). When learning conventional actions, children cannot discover target actions without guidance; in these cases, viewing action demonstrations supports learning better than independent, unguided child activity (Meltzoff, 1985; Fagard & Lockman, 2010). Indeed, sixteen-month-olds learned to use a new tool more effectively by observing someone else than engaging in active training (Somogyi et al., 2015).

Adult demonstrations can be pedagogical, where teachers intentionally teach children; pedagogy been theorized as central for learning (Csibra & Gergely, 2009). However, children learn through observing non-pedagogical demonstrations as well (Shneidman & Woodward, 2015; Gaskins & Paradise, 2010). Whether models are provided intentionally or incidentally, children could learn all the necessary information through observation alone. Is there an added learning benefit for children to act during instruction?

Active experience has been proposed to be critical for young children's learning (Piaget, 1964), and wide-ranging research has lent support to this idea. Active experience benefits learning in contexts that involve instruction about abstract causal systems and mathematical concepts: Actively engaging with new material prior to instruction supported 4-year-olds' learning of causal systems (Sobel & Sommerville, 2010), second-fourth-graders' math learning (DeCaro & Rittle-Johnson, 2012), and fourth-graders' science learning (Dean & Kuhn, 2007). Similarly, caregivers who guided their children to actively explore a museum exhibit about gears had children who learned more about the causal structure of gears (Callanan et al., 2020). Consistent with these findings, curricula for preschoolers (Lillard, 2016) and elementary school children (Hmelo-Silver et al., 2007) often feature child activity in a guided context.

Does active engagement benefit toddler's learning of conventional actions? On the one hand, infants' and toddlers' robust ability to imitate others' actions, even after a considerable delay (Bauer, 1996), indicates that active engagement may not be needed in this context. On the other hand, the findings from older children's causal and academic learning suggest that even when instruction provides all the needed information, active engagement may boost learning. Moreover, active engagement has been linked to cognitive development in uninstructed contexts from early in infancy. Active experience with objects benefits infants' object exploration

(Needham et al., 2002), understanding of others' actions (Gerson & Woodward, 2014; Sommerville et al., 2008; Sommerville et al., 2005), and knowledge of object properties (Soska et al., 2010). Theories of embodied cognition propose that when children are actively engaged with objects, their perceptual, motor, and cognitive systems are integrated: Children's actions on the environment support their cognition, and cognition supports future actions (Wilson, 2002).

A recent study tested how toddlers' active and observational experience supported conventional action learning during everyday teaching interactions with caregivers (Brezack et al., 2021). The conventional actions to be learned were novel assembly actions on toys, for example, fitting pieces together to create an object. Having been told how the pieces should go together, caregivers taught their two-year-old children to perform the target assembly actions in whatever way caregivers wanted to. Then, toddlers were tested on their ability to perform the target actions by an experimenter who was unaware of which of several toy sets children had been taught. Results indicated that caregiver instructions benefitted learning; children did not spontaneously discover the target actions without instruction. Even so, toddlers who performed more actions during caregiver instruction demonstrated better action learning at test. In contrast, when caregivers demonstrated more actions for their children, children did less well in the test phase. Thus, children's tendency to actively engage rather than simply view caregivers' demonstrations was a positive predictor of learning.

These findings suggest that active experience may be key to toddlers' learning about conventional actions. However, the prior correlational study leaves open the possibility that the relation between toddlers' active engagement and their learning was due to another factor, rather than reflecting the direct effect of active engagement on learning. It is possible that individual differences in toddlers' cognitive maturity drove both their engagement during instruction and their ability to learn and remember the action sequences. Brezack et al. (2021) attempted to rule out this explanation by controlling for children's skills with objects, measured by their propensity to assemble multi-piece toys for which no instruction had been provided. Children who were better at spontaneously constructing toys learned more from their caregivers; even so, controlling for children's skills, those who were more active during instruction demonstrated better learning. Still, an independent measure of children's cognitive maturity would better elucidate the roles of children's developmental level and active experience in learning.

In the current study, we tested whether active experience *per se* supported action learning, or whether children's developmental level was more relevant. We used an experimental teaching manipulation that controlled whether children were able to be active during instruction to test whether active experience benefitted learning. In a within-subjects design, children were introduced to target actions for which instruction was either active or observational. In the active condition, children were coached to perform the target actions themselves; the teacher guided children to perform the correct actions, but children learned via actively engaging in the actions rather than by observing the adult's actions. In the observational context, children saw the teacher produce the actions, but they did not engage in the actions themselves. Both types of instruction therefore allowed children access to the target actions to be learned but differed by whether children performed or observed the actions. Following instruction, children's learning was assessed by an assistant who was unaware of the training configuration to which children had been assigned. Children were also tested on their ability to generalize taught actions to novel toys. In a follow-up study, we evaluated whether the instruction conditions led to differences in children's long-term memory for the taught material.

We used an independent measure of children's developmental level to test whether children with greater cognitive maturity learned more effectively from the instructions. We assessed variation in children's readiness to learn via the Bayley Scales of Infant and Toddler Development, Cognition subscale (Bayley, 2006), which measured children's cognitive and motor performance on items similar to those learned from active and observational instructions (e.g., assembling puzzles). The methods and analyses were preregistered (Study 1: https://osf.io/y35dp/?view_only=e7261a7c1460448ba8b1309fa1bf855a; Study 2: https://osf.io/k2sx9/?view_only=b6dac5e6e0374712b5fc7af48f8cb992). Videos and coding manuals can be found in the Databrary video repository (http://doi.org/10.17910/b7.1328).

Study 1

Methods

Participants. 46 full-term toddlers exposed primarily to English participated (average age = 23.4 months, range = 22.2 – 26.0 months; 21 male). Approximately half of the sample was White (European or White American: 20, African American: 11, Asian: 2, Hispanic: 1, multiple races: 12). The sample had generally high levels of maternal education (post-graduate degree: 22, bachelor's degree: 12, some college: 8, associate's degree: 2, did not report: 2). Six additional children were excluded following preregistered criteria (hearing more than 25% of another language at home: 3, prematurity: 1, refusal to perform actions during Active instruction: 2). The preregistered sample size was 48, but two children were excluded after data collection was complete because they did not watch Observed instruction long enough (as preregistered) for a final sample of 46.

Procedure. Children participated in a laboratory of a research university in a large Midwestern city with a primary caregiver between May and October of 2019. Children were recruited from a database of families who had agreed to participate in research. The study had three phases that were administered in a fixed order: Teaching, Test, and Cognition (Figure 1). An additional phase in which caregivers taught their toddlers to assemble a toy occurred prior to Teaching, but was not relevant to the research questions tested here. Children participated individually in a quiet testing room with an experimenter and one of two trained assistants. Caregivers sat behind toddlers, were instructed not to interfere. Caregivers also completed a measure of children's productive vocabulary (MacArthur Bates Communicative Development Inventory, Level II; MCDI; Fenson et al., 2000). The session lasted approximately one hour. Video was recorded simultaneously from four webcams and audio was recorded from one webcam. Families were given $20 and children were given a t-shirt or book and a certificate for participating.

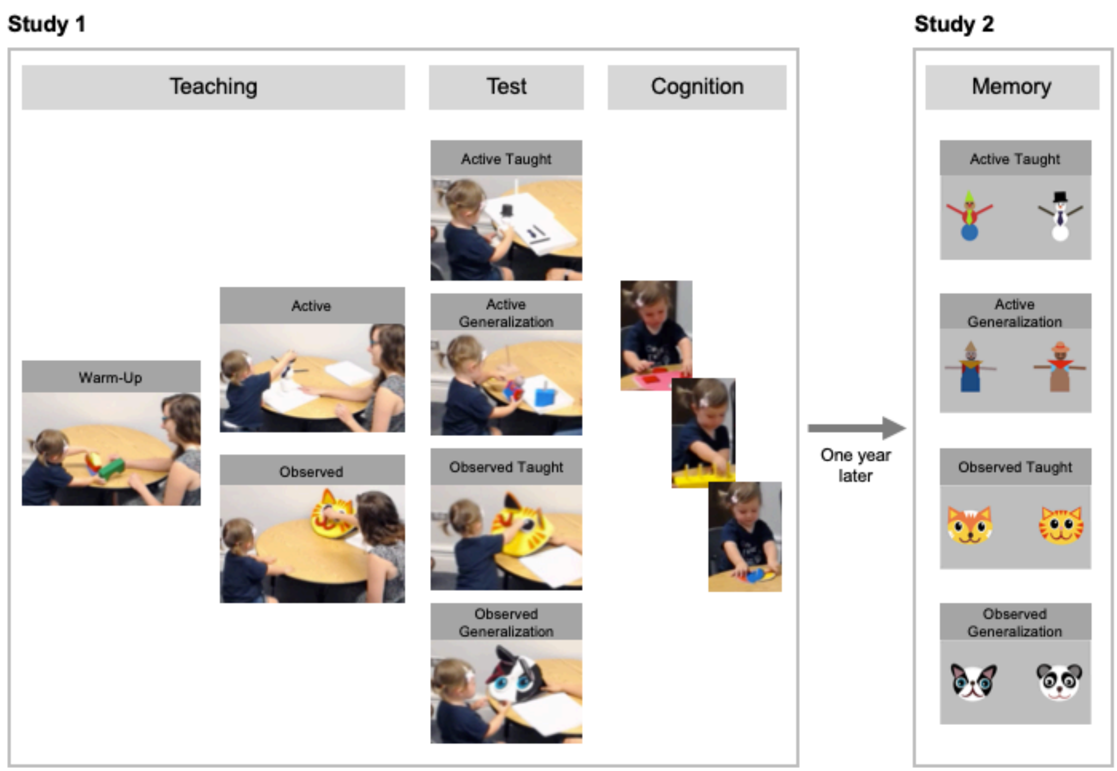


Figure 1: Study 1 and Study 2 Design. Toddlers were taught by the experimenter (Teaching): After a Warm-Up, children were taught actions on two toys in different instruction styles: Active and Observed (here: Active – Snowman, Observed – Cat). Children were then tested on their

Figure 1: Study 1 and Study 2 Design (continued): learning of the toys taught in Teaching (here: Active Taught – Snowman, Active Generalization – Scarecrow, Observed Taught – Cat, Observed Generalization – Snowman). Toddlers then completed the Cognition Subscale of the Bayley Scales of Infant and Toddler Development (Cognition; Bayley, 2006; e.g., shape puzzle, peg board, ball puzzle). Study 2: Children participated in a memory task where they saw each target toy from Study 1 with a matched foil (here: Active Taught – Snowman, foil – Clown; Active Generalization – Scarecrow, foil – Cowboy; Observed Taught – Cat, foil – Fox; Observed Generalization – Dog, foil – Panda).

Materials. Four novel toys were designed for the study, modeled after toys in Brezack et al. (2021), each with associated target actions to be learned (Figure 2). Each toy had a base and six pieces. Target actions (Action Types, 6 per toy) consisted of placing pieces onto the base to assemble a final end state. For example, the Action Types for the Cat toy were ear 1, ear 2, eye 1, eye 2, nose, and mouth. During Teaching, the experimenter taught children to perform target actions to assemble one animal face toy (Cat or Dog) and one stacking figure toy (Snowman or Scarecrow). Children were taught one pair of toys, either the Cat and Snowman (Pair 1) or the Dog and Scarecrow (Pair 2). One toy in the pair was taught in an active context, and the other was taught in an observational context (within-subjects).

Children were tested on all four toys at Test: the two toys taught by the experimenter (Taught Toys; e.g., Pair 1: Cat and Snowman) and two toys matched to Taught Toys that were not taught (Generalization Toys; e.g., Pair 2: Dog and Scarecrow). Actions were designed to transfer between the two animal face toys (Cat and Dog) and between the two stacking figure toys (Snowman and Scarecrow) such that at Test, action performance on the untaught toys reflected children's ability to generalize taught actions to parallel toys. For example, the Cat and Dog had ears designed to be inserted into slots on the base, but the ears differed in color and shape between toys. A set of four colorful foam blocks was used in the Warm-Up phase that began Experimenter Teaching.

Teaching. Teaching consisted of three parts: Warm-Up, Active, and Observed. The order of Active and Observed instructions were counterbalanced across children. The *Warm-Up* familiarized toddlers with the procedure later used during Active and Observed instructions: Sometimes, the child performed actions (Active), and sometimes the experimenter performed actions (Observed). During the Warm-Up, children built a block tower, then the experimenter built the tower. This procedure was repeated (four sequences total: child, experimenter, child, experimenter) to provide toddlers familiarity with performing actions on their turn and observing actions on the experimenter's turn.

During *Active* instruction, children were coached by the experimenter to perform a series of target actions to assemble a toy. Prior to performing each action, the experimenter provided instructions about how to perform the action with language and pointing. She then handed toddlers the associated piece so they could perform the action (e.g., Cat: "First you do the ear. The ear goes right here [*point*]. Can you do that?" [*hand piece to child*]). The experimenter assisted children in performing the correct actions if necessary by repeating instructions or pointing gestures, finishing children's actions, adjusting placements, hovering pieces over their correct locations, or briefly demonstrating actions. This process was repeated for each action in the sequence in the same fixed order (e.g., Cat: ear 1, ear 2, eye 1, eye 2, nose, mouth).

During *Observed* instruction, toddlers learned the target actions on a different toy by watching the experimenter perform the actions. As in the Active sessions, the experimenter provided instructions for each action with language and pointing. However, the experimenter then performed each action in the sequence (e.g., Cat: "First I'll do the ear. The ear goes right here [*point*]. I'll do that." [*place piece in correct location*]). The procedure was matched between Active and Observed sessions; the only difference was the person performing the actions:

Children performed target actions during Active sessions and the experimenter performed actions during Observed sessions. Instruction order (Active or Observed first), taught toy pair (Pair 1: Cat and Snowman, Pair 2: Dog and Scarecrow), and active toy (Cat or Dog active, Snowman or Scarecrow active) were counterbalanced across children.

Test. Following Teaching, one of two trained assistants who was not present for Teaching tested children on their target action performance. Toddlers were tested on the caregiver-taught toy first (which was not relevant to the research questions presented here), then the two Generalization Toys, and finally the two Taught Toys. Children's action performance on the toys they were taught during Teaching reflected their learning from instruction. Performance on the matched, untaught Generalization toys reflected children's action transfer skills. Children were tested on each toy individually for up to two minutes or until children refused to continue performing actions. The assistant remained neutral during Test and assisted only when necessary to help children complete actions.

Cognitive Maturity. To test children's cognitive maturity, the experimenter administered the Bayley Scales of Infant and Toddler Development, Cognition Scale (Bayley, 2006) following standardized testing procedures. This subscale requires similar skills as those needed to perform target actions in the study (e.g., simple puzzles, peg board).

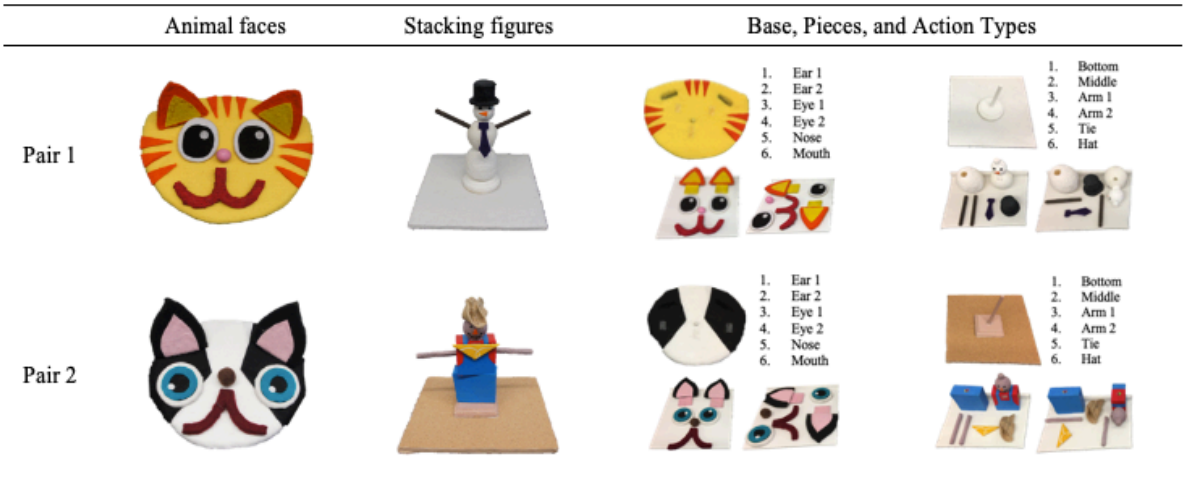


Figure 2: Study 1 Materials. Four toys with 6 novel target assembly actions (Action Types) used in Teaching are shown. Two animal faces (Cat and Dog) and two stacking figures (Snowman and Scarecrow). The experimenter taught one pair (Pair 1: Cat and Snowman or Pair 2: Dog and Scarecrow). Each taught toy had a matched toy parallel in structure but with pieces that differed in color, shape, and texture (Generalization: Cat - Dog, Snowman - Scarecrow). Toddlers were tested on all toys at Test: two Taught Toys (from Active and Observed instruction) and two matched, untaught Generalization Toys. Bases, Pieces, and Action Types (within each cell): Each toy had a base (top left) onto which six pieces were placed (bottom). Six Action Types (top right) were taught in a fixed order. Children were presented with the base and pieces in an organized array during Teaching (bottom left) and in a shuffled array during Test (bottom right).

Coding

Coding was performed to answer two research questions. (1) Did Active or Observed instruction differentially support action learning or generalization? To assess action learning differences, Test sessions were coded for children's ability to perform taught actions, generating Test Scores for each toy. (2) Did individual differences in children's cognitive maturity relate to learning? Specifically, we examined children's cognitive maturity as measured by the Bayley.

In addition, control measures were included to test whether session- or child-level individual differences related to learning. Session Factors included the amount of time children spent in each teaching session (Active and Observed). Child Factors included children's age and MCDI score. We also coded children's engagement during Active and Observed instruction to ensure that children performed actions accurately in Active instruction and were visually attentive to Observed instruction. In addition, we coded the assistance that the experimenter provided to children during Active instruction. See Table 2 for all measures. Coders were trained with manuals and practiced prior to performing final coding (Mangold Interact; Mangold, 2017). One coder coded Caregiver Teaching and Test sessions and was blind to the toys taught in Teaching. A second coder coded Teaching and was blind to children's performance at Test.

Action Learning: Test. Children's target action performance was coded during Test to examine whether toddlers learned and generalized differently from Active and Observed instructions. Children's action attempts on each toy were coded by Action Type (e.g., Cat: eye 1, eye 2, ear 1, ear 2, nose, mouth) and assigned a numerical score reflecting accuracy (maximum score per action: 1). Because children often attempted actions more than once, children's highest scoring action attempt per Action Type was used to calculate Test Scores. For example, if a child placed a Cat ear in the upper right corner of the base, that attempt would receive one full point. However, if the child then placed the ear near the bottom of the base, the attempt would receive a lower score (.2). The highest scoring attempt of each Action Type (e.g., the score of 1 rather than .2 for the Cat ear) was averaged to generate a **Test Score** for each toy (maximum score: 1). Children's **Active Taught Test Score** and **Observed Taught Test Score** reflected learning from Experimenter Teaching. Children's scores on the generalization toys reflected their ability to transfer actions learned in experimenter instruction to matched toys: **Active Generalization Test Score** and **Observed Generalization Test Score**.

Cognitive Maturity: Bayley. Children's cognitive maturity was measuring by scoring children's performance on the Cognition subscale of the Bayley. Children's performance on the Bayley Cognition Scale was scored following standardized scoring guidelines to generate an age-normed **Bayley** score for each child.

Engagement During Teaching. We coded children's visual attention to ensure children watched the Observed sessions for at least 80% of the demonstration. To measure visual attention, we first coded the total time Active and Observed instructions lasted. **Active Time** and **Observed Time** (included as Session Factors) were coded from the onset of experimenter speech about the toy until the toy was taken away. Then, durations of time toddlers were looking away from the demonstration (experimenter and toy) were coded. **Observed Attention** was calculated as the proportion of time children were watching the instruction out of 1 (Observed Attention = (Observed Time - look away time)/Observed Time).

Children's actions were coded analogously to their actions during Test, but actions performed during Active instruction reflected children's practice with taught information, while Test performance reflected learning. Analogous to Test Scores, each action was coded by Action Type and accuracy with a maximum of one point per action. The highest scoring attempt of each Action Type was averaged to yield a measure of **Child Best Performance Score** out of 1. Actions were coded to ensure children reached at least .8, as preregistered. Each instance of experimenter assistance necessary for children to perform taught actions was coded, including repeating instructions or pointing gestures, assistance finishing children's actions, adjusting placements during or after children performed an action, hovering pieces over their correct locations, or briefly demonstrating actions. The total number of times the experimenter assisted children was counted per child for a measure of **Experimenter Assistance**.

Session and Child Factors. In addition to Active Time and Observed Time, Child Factors were calculated to control for child-level individual differences that might relate to learning, including children's **Age** in months and the number of words caregivers reported children said on the **MCDI** (a measure of productive vocabulary).

Reliability Coding. Videos of 10 children (21.7%) were coded by a second coder. Each measure was calculated for each child separately per coder and compared between coders to calculate reliability (e.g., Taught Test Scores on each toy for each child were compared). Across all measures, reliability was high: Test Scores: average ICC = 0.90, all *p*'s < .001, Teaching: average ICC = 0.87, all *p*'s < .013, Session Factors: average ICC = 0.99, all *p*'s < .001. Reliability could not be calculated for Child Best Performance Score because children were coded as perfectly accurate by one coder; coding was 90% identical between coders.

Study Phase	Measure	Description of Measure	Mean (SD); range
Teaching	Observed Attention	Visual attention (proportional): (Observed Time - look away time)/Observed Time	.99 (.02); .90-1
	Child Best Performance Score	Accuracy in attempting target actions	37 of 46 children scored 1
	Experimenter Assistance	Number of instances of experimenter assistance during Active sessions: repeating instructions or pointing, finishing an action, adjusting a placement, briefly demonstrating an action	10.1 (5.5); 2-27
Test	Active Taught Test Score	Active learning: Accuracy in target action performance on toy taught in Active session	.52 (.26); 0-1
	Observed Taught Test Score	Observational learning: Accuracy in target action performance on toy taught in Observed session	.57 (.27); 0-1
	Active Generalization Test Score	Active generalization: Accuracy in target action performance on toy matched to toy taught in Active session	.55 (.25); 0-.97
	Observed Generalization Test Score	Observational generalization: Accuracy in target action performance on toy matched to toy taught in Observed session	.55 (.28); .05-1
Cognition	Bayley	Cognitive maturity; age-normed score on the Bayley Cognition subscale	61.5 (5.9); 48-74
Session Factors	Active Time	Amount of time Active instruction lasted (seconds)	95.7 (17.0); 66.1-156.9
	Observed Time	Amount of time Observed instruction lasted (seconds)	74.3 (6.1); 61.8-86.7
Child Factors	MCDI	Productive vocabulary	40 (17.9); 5-77
	Age	Child's age at date of test (months)	23.4 (1.1); 22.2-26.0

Table 2: Study 1 Measures. Measures and descriptions of measures from Study 1 coded during Teaching, Test, and Cognition. Additional measures included as Session and Child Factors. Descriptive statistics are provided (mean, standard deviation, and range).

Analysis

Inclusion Criteria. As preregistered, children included in analyses performed actions with at least .8 accuracy during Active instruction (Child Best Performance Score) and visually attended for at least 80% of the Observed Time (Observed Attention). All children were included based on Child Best Performance Score. Data from two children were excluded for insufficient Observed Attention (.58; .75), leaving 46 children for analysis. Three toddlers did not complete the Bayley and were not included in analyses of the Bayley.

Analysis Strategy. The analyses were run to test whether active experience during instruction or cognitive maturity related to children's action learning from Active and Observed instruction. The analyses were run as linear mixed-effects models with subjects as random effects and the outcome as Test Score on the four experimenter-taught items (Active Taught Test Score, Active Generalization Test Score, Observed Taught Test Score, Observed Generalization Test Score) using the lme4 package in R (Bates et al., 2015). Across measures, values more than three standard deviations from the mean were excluded.

Results

Action Learning. Prior to testing whether learning differed by Active and Observed instruction, we checked for effects of the items within each type of toy (animal faces and stacking figures; scores did not differ: all p's > 0.109). We next tested whether Session Factors (Active Time: average = 95.7 seconds; Observed Time: average = 74.3 seconds; instruction order: Active or Observed first; taught toy pair: Cat and Snowman or Dog and Scarecrow; and active toy: Cat or Dog active, Snowman or Scarecrow active) or Child Factors (Age and MCDI: average = 40) were related to Test Scores, combining two preregistered models. As preregistered, we included children's cognitive maturity (Bayley score; average = 61.5) in this model.

Bayley score was significantly related to Test Score ($\beta = 0.02$, SE = .01, $p = .045$); no other Session or Child factors were related to learning (all p's > .18; Experimenter Assistance was also not significantly related to Active Taught Test Scores, $p > .67$). Individual differences in children's cognitive maturity related to learning: More advanced children performed better at Test. Our main preregistered analysis examined whether children learned differently from Active or Observed instruction, beyond variations in children's cognitive maturity. A model was run with Test Scores as the outcome, predicted by Instruction experience (Active and Observed), Test Type (Taught and Generalization), and the interaction between Instruction experience and Test Type, controlling for Bayley. Surprisingly, the main effects and interaction did not reach significance (all β's < 0.06, all p's > .156; Figure 3A); children learned similarly from Active and Observed instruction, and on Taught and Generalization items. Only Bayley scores significantly related to learning ($\beta = 0.02$, SE = .01, $p < .001$).

We exploratorily tested whether learning from Active and Observed instruction differed by performance on the Bayley. A model with Test Score predicted by Instruction experience (Active and Observed), Bayley score, and their interaction revealed a marginal interaction between Instruction experience and Bayley ($\beta = 0.01$, SE = .01, $p = 0.083$; Figure 3B) such that toddlers with higher Bayley scores learned marginally more from Observed than Active instruction. In sum, children's cognitive maturity, not the opportunity to act during instruction, supported children's immediate action learning and generalization.

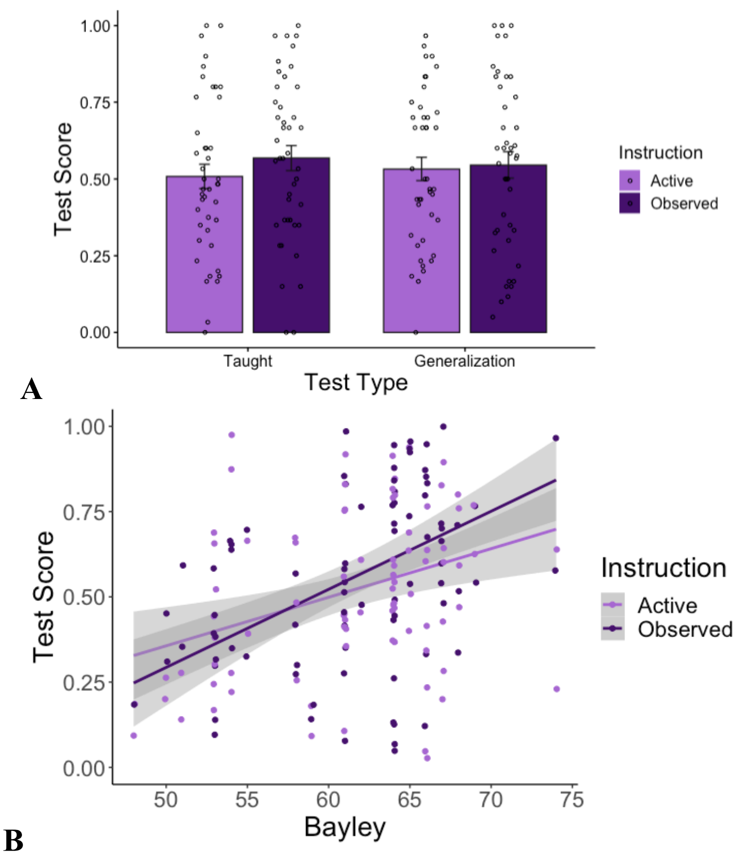


Figure 3: Study 1 Test Scores from Active and Observed Instruction. (A) Bar graph showing average Test Scores from Active and Observed instruction by Taught and Generalization test types. Error bars: +/- 1 SE. (B) Scatterplot showing relation between Bayley scores and Test Scores separated by Active and Observed instruction. Shaded areas represent 95% confidence intervals.

Discussion

We tested whether active experience *per se* supported children's action learning, or children's cognitive maturity was more relevant. Brezack et al. (2021) found that beyond children's skill levels, their propensity to actively engage with caregivers benefitted action learning. Here, we used an independent measure of children's cognitive maturity (the Bayley), and found that developmental level, not active experience, supported learning. Children with greater cognitive maturity demonstrated better learning regardless of instruction condition.

While children learn effectively through viewing action demonstrations (Bauer, 1996), active experience has been argued to be central for learning (Piaget, 1964) because it integrates perceptual, motor, and cognitive systems ("embodied cognition"; Wilson, 2002). The neural motor system (Marshall & Meltzoff, 2014) is also recruited to a greater extent for actively-learned than observationally-learned information (James & Swain, 2011). Surprisingly, children learned and generalized actions similarly from instruction that did and did not offer opportunities to act. Children may have learned similarly from both contexts because they were efficient (e.g., Kirschner et al., 2006), targeted only the material to be learned, and were free from distractions. It is possible that when information is presented in supportive contexts during short, information-rich teaching sessions, instruction may support immediate learning regardless of the presence of child activity. Both types of instruction also provided all the relevant information for learning, potentially rendering active experience unnecessary.

While active experience did not improve immediate action learning or generalization, it is possible that acting during instruction could affect children's ability to retain taught information over a longer period of time. Here, children were tested immediately after instruction. Testing children's memory after a delay could reveal underlying benefits of active experience. As with immediate learning, children's cognitive maturity could again support memory regardless of the way information was originally taught. Alternatively, instruction that included opportunities for children to be active could support children's long-term memory. We therefore measured children's memory for information taught in Active and Observed contexts one year after instruction to test whether active engagement benefitted children's long-term memory.

Study 2

Toddlers have the capacity for robust long-term memories: Young children remember people for years (Lie & Newcombe, 1999). In particular, children have strong memory capacities for actions. Infants as young as 6 months old remembered actions they were taught in a lab setting (Barr et al., 1996). Young children's action memory has been shown to persist for months (Bauer et al., 1994) and up to a year after being taught new actions (McDonough & Mandler, 1994). Active experience may be particularly important for children's long-term memory: Eighteen-month-olds who practiced an action after seeing a demonstration remembered the action after a six-week delay better than children who only observed the demonstration (Hayne et al., 2003).

Active experience when learning similarly benefits adults' memory. Adults had better memory for actions they produced compared with actions they observed (Cohen, 1989; Nilsson, 2000). Adults also demonstrated better memory for object locations after performing self-propelled reaching actions compared to robot-propelled reaches (Trewartha et al., 2014). This suggests that initiating motor commands when acting may underlie active engagement's memory benefits. Physical activity also involves encoding sensorimotor associations, which can create richer memories that are more likely to be recalled (Markant et al., 2016). We tested whether children's long-term memory was similarly enhanced by active compared to observational instruction one year after children were instructed.

Methods

Participants. All 46 children were contacted to participate in the memory follow-up task, and caregivers of 32 children agreed for their children to participate. Of those 32 children, 6 demonstrated a side bias and were excluded from analyses as preregistered. The final sample consisted of 26 children (12 male, Study 1 average age = 23.3 months, range = 22.2-25.3 months; Study 2 average age = 36.7 months, range = 33.8-40.6 months; time difference average = 13.4 months, range = 10.4-17.0 months). Children's ethnicity and maternal education were similar to that of Study 1 (European or White American: 11, African American: 3, Asian American: 2, Hispanic: 1, multiple races: 9; post-graduate degree: 18, bachelor's degree: 5, some college: 2, associate's degree: 1).

Procedure. Children were tested on their memory for toys they had learned to assemble one year after instruction. After providing informed consent, caregivers and children attended a recorded Zoom session with a trained assistant. Data was collected between July and October of 2020. The study took approximately 10 minutes and families received a $5 gift card.

Materials. Computer-drawn versions of each experimenter-taught toy (target toys: Cat, Dog, Snowman, Scarecrow) were created. A foil image was drawn to match each target toy; foils were designed to be equally salient with structures similar to those of the target toys (Cat – Fox, Dog – Panda, Snowman – Clown, Scarecrow – Cowboy; Figure 1). Each target toy and its corresponding foil image were presented side-by-side in the same fixed order for all children, target toy side counterbalanced. In addition, the Warm-Up toy and caregiver-taught toy were drawn with foil images and were used to familiarize children with the procedure. Two practice trials (orange circle and blue circle, yellow star and green heart) were used to prepare children for the task. To remind children about their original lab visit, three images from children's sessions (caregiver and child, experimenter and child, assistant and child) were taken from the video recordings.

Memory Task. Caregivers set up their computer screens so children could see only the images that the assistant shared via screen sharing on Zoom. Caregivers hid the view of their own video to minimize distraction and were told not to intervene. The assistant administered two practice trials (orange circle and blue circle: "Can you point to the blue circle?"; yellow star and green heart: "What shape is this?") to encourage children to respond to questions either by naming the item or by pointing. Then, the assistant showed children the pictures from their original lab visit to remind them about the previous session ("A long time ago, you played with toys with your mom and some friends! Here you are playing with your mom!"). Children then saw the Warm-Up trial and the caregiver-taught toy trial as additional practice (e.g., "Did you make a Tower or did you make a Castle?").

Next, children saw the four Test Trials: Fox and Cat, Scarecrow and Cowboy, Dog and Panda, and Clown and Snowman. On each trial, children were asked which toy they remembered playing with (e.g., "Did you make a Fox or did you make a Cat?"). When each toy was named, it expanded slightly on the screen for emphasis, then returned to its original size. Children responded by pointing to or labeling one of the two toys. Test Trials corresponded with toys children had been previously taught: Active Taught Toy, Active Generalization Toy, Observed Taught Toy, Observed Generalization Toy (toy assignment was counterbalanced across participants in Study 1; Figure 1). Four additional trials were presented to make the game more engaging: Clown and Scarecrow, Fox and Dog, Snowman and Cowboy, and Cat and Panda. However, children's memory for those additional trials could have been contaminated by their responses to the four Test Trials; thus, only responses to the four Test Trials were analyzed.

Coding

Each session was coded for children's forced-choice response to each item (correct: selected the target toy, or incorrect: selected the foil image) based on the child's verbal label, point, or both. Each trial therefore received a Memory Score of 1 or 0: **Active Taught Memory**

Score, **Active Generalization Memory Score**, **Observed Taught Memory Score**, and **Observed Generalization Memory Score**.

Reliability. Videos of 6 children (23.1%) were double coded by another coder. Reliability was high (23 out of 24 judgements were identical; 95.8% agreement).

Analysis

Inclusion Criteria. As preregistered, children needed to respond to at least two of the four Test Trials to be included; all children did so. Six additional children demonstrated a side bias on the four Test Trials (i.e., selected the image on right side of the screen on all four trials) and were excluded from the analyses as preregistered for a final sample of 26 children.

Analysis Strategy. The analyses were run as logistic mixed-effects models with subjects as random effects and the outcome of binary Memory Score for the four toys using the lme4 package in R (GLMM; Bates et al., 2015).

Results

Session and Child Factors. First, preregistered control analyses were run to test whether Memory Scores differed by session or child factors. Age during Study 1, the time difference between studies, and gender were tested as predictors of Memory Score; no predictors reached significance (all p's > .219). Toy type (Cat, Dog, Snowman, Scarecrow) also did not significantly relate to memory (all p's > .316), and a chi-square test did not show a significant difference in the distribution of toys previously taught in Active or Observed conditions ($\chi^2(3) = 1.54$, $p = .673$). Analyses were therefore collapsed across session and child factors.

Memory. To test whether children's memory for toys differed by way the associated actions had been originally instructed, a preregistered model was run with Memory Score per toy as the outcome predicted by Instruction experience (Active and Observed), Test Type (Taught and Generalization), and their interaction, controlling for Bayley score (measured during Study 1). Children demonstrated significantly better memory for items learned through Active than Observed instruction (Instruction experience: β = 1.65, SE = 0.70, $p = .017$; exponentiated coefficients: odds ratio of remembering a toy taught in Active compared to Observed condition = 5.23). Test Type, the interaction between Instruction experience and Test Type, and Bayley score did not significantly relate to memory (all p's > .131). Children therefore showed better memory for actively-learned than observationally-learned toys. Children's cognitive maturity previously supported their immediate learning; in contrast, children's maturity was not related to their memory. Instead, the ability to be active during instruction supported children's memory.

Since no difference between memory for Taught and Generalization measures emerged, we exploratorily computed average Active Memory Scores and Observed Memory Scores across Taught and Generalization items for each child and compared average memory scores to chance (.5). Children's average Active Memory Score was greater than chance ($t(25) = 2.85$, $p = .009$, $d = 0.56$), while average Observed Memory Score was not significantly different from chance ($t(25) = 0$, $p = 1$; Figure 4). Children therefore remembered the toys they had learned through Active instruction, but did not recall those learned through Observed instruction. This was found despite the fact that all children had experienced both instruction conditions originally – yet, only instruction including active experience related to children's memory. Of note, children's learning from Active and Observed instruction did not relate to memory: Test Scores from Study 1 were not related to Memory Score (controlling for Bayley; β = -0.80, SE = 1.02, $p = .435$). In sum, children recalled toys learned actively, but not observationally, one year after receiving instruction in both conditions. Children's cognitive maturity was not relevant for recall. Instead, active experience appeared pivotal for long-term memory.

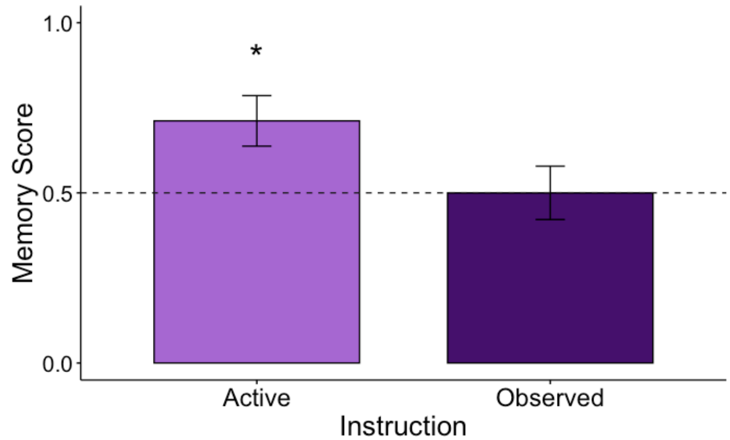


Figure 4: Study 2 Memory for Active and Observed toys. Bar graph showing average Memory Score for Active toys (across Taught and Generalization items) and average Memory Score for Observed toys (across Taught and Generalization items). Chance is 0.5. Error bars: +/- 1 Standard Error. *$p < .01$.

Results Summary. Despite demonstrating equivalent immediate learning from instruction where children acted and observed, children remembered actively- but not observationally-taught toys one year after instruction. Children's cognitive maturity, not the ability to be active, were relevant for immediate learning. However, children's maturity was unrelated to their long-term memory. Instead, active experience during instruction supported memory. Children remembered the toys they learned through instruction where they had been active, but not through instruction where they had instead observed a teacher's actions. Instruction including opportunities for children to act benefitted children's memory after a considerable delay of one year.

Discussion

Brezack et al. (2021) found that during everyday interactions with caregivers, toddlers who were active during instruction demonstrated better action learning. However, the prior correlational study could not discern whether active experience *per se* benefitted learning, or whether another factor (such as children's cognitive maturity) was responsible for this relation.

Results from the present study suggest that children's developmental level certainly plays a role in learning – cognitive maturity was the most influential predictor of immediate action learning and generalization. However, beyond maturity, active experience itself was crucial for memory: Children's long-term memory for taught information was enhanced for information they had learned through active compared to observational instruction. In sum, children's developmental level supported their ability to learn from instruction, but active experience played a critical role in supporting children's memory.

Active experience is thought to support learning and memory because physical activity integrates perceptual and motor systems, which in turn support cognition ("embodied cognition"; Wilson, 2002). Active experience also includes planning and executing actions (Barrett et al., 2007), focusing attention on actions (Boudreau & Bushnell, 2000), and problem-solving, remembering, and learning (Adolph & Joh, 2007). When children were physically active, as in the active instruction condition, they integrated their perceptual, motor, and cognitive systems; when children instead observed actions, their motor systems were not active in the same way as when they were acting. We expected that this difference in physical activity would enhance learning and memory from active compared to observational instruction.

Unexpectedly, despite the wealth of research on the benefits of active experience for learning (e.g., Piaget, 1964), we found that children demonstrated equivalent immediate learning from active and observational instruction. This could be because both teaching conditions were highly supportive of learning. In this within-subjects design, children experienced active and observational instruction on separate toys (with the instruction order counterbalanced). The teaching conditions were carefully controlled and varied only by whether the child or the experimenter performed the actions to be learned. The instructions were efficient (Kirschner et al., 2006), free from distractions, and contained all the information children needed to learn the target actions. Children also generalized taught actions similarly following active and observational instruction (similar to results from Wakefield et al., 2018). While there was variability in children's action learning, with some children learning the actions better than others, this was due to children's cognitive maturity rather than the instruction style in which the actions were taught.

The instructions were also presented in a highly collaborative context. Children first assembled a novel puzzle toy with their caregiver, and all children were active with their caregivers. Children then engaged in a warm-up where the experimenter and child took turns building a block tower. Children were then taught in active and observational styles. This collaborative environment may have caused children to take an active stance when learning (i.e., mentally simulating actions; Knoblich & Sebanz, 2006), even when children were not physically active. Indeed, prior research has shown that when children are in collaborative environments, they may become confused about who performed actions and may overclaim another person's actions as their own (Sommerville & Hammond, 2007). Even implied active engagement, induced by the presence of a person, supported children's action recall (Howard et al., 2020). Further, taking turns with a social partner enhanced infants' (Meyer et al., 2022) and toddlers' (Meyer et al., 2011) neural mirroring of observed actions. Children may have learned more than expected from observational instruction because the collaborative environment caused children to take an active stance when observing.

It is also possible that children's active experience, while physically active, lacked key aspects of active engagement that could have improved their learning. In the active condition, the experimenter explained where each piece should be placed, handed children the piece, and guided children to perform each action correctly. Thus, children were physically active, but their actions were constrained by the experimenter. Research with older children and adults has shown that when learners can make decisions during active learning, such as by controlling the flow information (Gureckis & Markant, 2012), making discoveries (Dean & Kuhn, 2007), and encountering failure (DeCaro & Rittle-Johnson, 2012), learning is enhanced. Here, we manipulated only physical activity to ensure that children had equal access to the actions to be learned, providing a direct comparison of instruction contexts. Young children may learn more effectively when active experience includes opportunities to make decisions.

However, despite learning equivalently from both conditions when tested immediately, children's long-term memory for taught toys was enhanced for toys they had learned actively rather than observationally. This was found even though children had initially been taught in both conditions with tightly matched instructions. In addition, toddlers had been tested on all toys during the immediate test phase following instruction; children had therefore acted on every toy, even those taught observationally, at minimum during test. Even so, children had better memory for the toys that had initially been *taught* actively compared to observationally.

Why did physical activity during instruction specifically support children's memory? Sensorimotor encoding could have improved children's memory for information they had learned actively (Markant et al., 2016). Physical activity allows for the integration of visual and manual information, as well as proprioceptive cues about body positions in space (Knoblich et al., 2006), which contribute to rich multimodal representations of information (Rochat, 1989). Encoding multiple cues during learning could provide stronger episodic representations of performed actions compared to observed actions, which would be more likely to be stored and recalled than information learned without sensorimotor integration (Markant et al., 2016). In addition, enacting motor commands could have similarly deepened mental representations of learned information (Trewartha et al., 2014). The physical act of performing actions could enhance memory due to deeper encoding while acting compared to observing.

In addition, physical activity could have guided children's attention to relevant information. Though children were visually attentive to both active and observed instructions, during active instruction, children could have targeted their attention towards the specific toy pieces they were manipulating and the actions they were performing. From early in life, infants direct their attention towards stimuli selectively via their gaze and actions (Baek et al., 2020). Focusing attention on manipulated objects supports children's learning of object names (Periera et al., 2009). Attention subsequently affects both short- and long-term memory (Amso & Scerif, 2015). Physical activity may have therefore targeted children's attention towards self-performed actions, supporting their memory for taught information (Markant et al., 2016); when actions were viewed, children may not have focused their attention in the same targeted way.

Children learned to perform actions physically, but were tested on their ability to visually recognize the structures they had built. Prior research suggests that information initially learned via physical activity can contribute to visual recognition memory. When sensorimotor information and coordinated muscle movements (Pouw et al., 2014) contribute to information encoding, learners can mentally simulate actions and activate visual representations (Johnson et al., 1989). Indeed, objects that adults manually explored were recalled one week later in a visual recognition test, demonstrating "cross-modal object recognition" (Hutmacher & Kuhbandner, 2018). Further, adults who explored objects while looking at them (integrating sensory, motor, and visual information) recalled the objects more accurately three weeks later than objects they had only seen or only touched (Novak & Schwan, 2021).

Why was there a disconnect between active experience's effect on immediate learning versus memory? This disconnect could be due to the type of information children were asked to recall. One year after instruction, we asked children which *items* they remembered. In contrast, other studies have tested children's memory for *actions* (e.g., Sommerville & Hammond, 2007). It is possible that different components of active engagement support learning and memory for different types of information. Here, children engaged their motor systems by acting, which supported their memory for items. Sommerville and Hammond (2007) found that children who took an active stance while learning (overclaiming another's actions as their own) had better memory for actions. It is possible that item memory is specifically enhanced by physical activity, while action memory is enhanced when children take an active stance during learning. Alternatively, if we had tested children's immediate learning of items or long-term memory for actions, we may have found that physically-active experience supported item learning and action memory; this is an area for continued research.

Of note, when toddlers were active during learning, they were active in the context of instruction. A teacher guided children to the correct actions by providing prompts, corrections, and demonstrations when children struggled. With the teacher's guidance, almost all children performed the taught actions with perfect accuracy. Children could not have learned through completely unguided activity because they would not have known which actions were the conventional, target actions to be learned. Instruction was therefore necessary for action learning. All the relevant action information to be learned could be gained from observing the teacher's actions. Even so, in a situation where instruction was necessary for learning, guided active experience supported children's memory.

Here, we used novel toys to induce familiar, playful learning environments, but active experience may also be important for children's memory when learning conventional actions on real-world artifacts like utensils and tools. When caregivers or teachers teach children to use artifacts in everyday contexts, providing children with opportunities to actively engage during learning could similarly benefit their memory. Alternatively, in real-world learning, children's active experience may be discouraged to avoid costly mistakes (Gaskins & Paradise, 2010). Future studies should address whether active experience during instruction benefits children's memory across more diverse contexts and cultures.

In sum, when children were instructed to perform novel conventional actions by acting rather than observing actions, they had better long-term memory for the taught material. Children's cognitive maturity, not the instruction condition, affected their immediate learning: Children who were more developmentally advanced learned more regardless of whether or not they had been active during the learning process. Despite the role that children's cognitive development plays in immediate learning, active experience during instruction benefitted children's long-term memory. This may be due to sensorimotor encoding and focused attention that occurs when children are actively engaged in learning. Instructions featuring opportunities for children to act seem particularly important for supporting children's memory.

Chapter 3: Exploration and Instruction Support Children's Rule-Learning in an Online Problem-Solving Game

Children often encounter new learning challenges: for example, children learn to tie their shoes and navigate tablet applications. These learning situations are examples of conventional problems, which have been designed by adults in children's culture. How do children learn these conventional systems? Receiving instructions is an efficient way to learn (Kirschner et al., 2006). For example, a caregiver may explicitly teach their child to tie their shoes. In contrast, children may learn some conventional information by exploring independently (Piaget, 1953). For example, a child may explore a tablet and learn to navigate to new applications without instruction. But, it could be faster and more efficient for children to learn through instruction because instruction provides children with all the relevant information they need to learn. In situations where instruction can make learning more efficient, does independent exploration benefit children's learning? In addition, children often apply what they have learned to new contexts: Shoe tying skills can be applied to tying knots with yarn, and tablet navigation skills can be transferred to using computer programs. What roles do exploration and instruction play in supporting children's generalization of novel conventional information?

Instruction efficiently transfers information to children (Kirschner et al., 2006; Stockard et al., 2018), and can support learning better than exploration, particularly when information is difficult or impossible for children to figure out independently. Indeed, when students learned logic puzzles and computer coding, instructions were more effective for learning than children's unguided exploration (Mayer, 2004). Similarly, when students in third and fourth grade learned the control of variables strategy to design science experiments, children learned and generalized the strategy better after receiving instruction than exploring (Klahr & Nigam, 2004). In particular, pedagogy, or child-directed instructions featuring cues to focus children's attention on relevant information, has been considered a key way children learn new information (Csibra & Gergely, 2006; 2009). When children can learn all they need to know from instruction, is it beneficial for children to explore?

Children's active, exploratory experience has been theorized as central for learning (Piaget, 1964). Active exploration allows children to make decisions (Gureckis & Markant, 2012), test hypotheses (Gopnik & Wellman, 2012), discover new information (Sobel & Sommerville, 2010), control the learning environment (Begus et al., 2014), and focus children's attention on their actions (Markant et al., 2016). Exploration benefits young children's understanding of objects and novel causal systems. Indeed, preschoolers discovered more functions of a novel object when their exploration was not constrained by instructions (Bonawitz et al., 2011). Similarly, toddlers discovered which blocks activated a machine through exploration (Sim & Xu, 2017) and 3-6-year-olds who systematically explored a gear exhibit had better causal understanding of gear functionality (Callanan et al., 2020). Exploration may be more effective than instruction in supporting learning if children are provided with sufficient time to explore. Though children learned the control of variables strategy better from instruction than exploration (Klahr & Nigam, 2004), when children were given more time to explore (6 months instead of one week), exploration supported learning more effectively than instruction (Dean & Kuhn, 2007). Instruction can therefore support learning when information is challenging to learn independently, but when children have sufficient time to explore and learn from their own actions, exploration may be more effective for learning than instruction.

While instruction alone and exploration alone support learning in certain contexts, combining exploration and instruction may be particularly beneficial. When children explore before receiving instruction, they gain information about the problem space and can focus their attention on relevant features (Markant et al., 2016). Exploring prior to instruction has been found to support children's causal and academic learning. For example, preschoolers learned the causal rules governing which buttons activated lights on a light box more effectively when they explored prior to instruction rather than exploring after instruction (Sobel & Sommerville, 2010). Exploring before instruction might also improve learning because children may encounter errors that they can compare to later instruction. This "productive failure" supported school-aged children's learning of math equivalence (DeCaro & Rittle-Johnson, 2012) and variance (Kapur, 2014).

However, the contributions of exploration, instruction, and their combination for children's learning are relatively unstudied beyond causal reasoning and formal school subjects like math and science. In addition, though generalization is an indicator of effective learning (Fiorella & Mayer, 2016), it is rarely studied in young children (e.g., Sobel & Sommerville, 2010). Radovanovic and colleagues (in preparation) tested the contributions of exploration, instruction, and exploration followed by instruction for 6-year-olds' learning and generalization in a novel problem-solving game: the Lock and Keys Task. Children's goal in the task was to figure out which keys unlocked a set of locks. There was a two-factor rule governing unlocking, which children learned by exploring, receiving instruction, or both. Children were then tested on their immediate learning and ability to generalize the rule to novel locks and keys. The results showed that all instructed children (regardless of prior exploration) learned the rule. However, exploring before instruction improved children's generalization of the rule to new examples. This study provided compelling evidence that exploring before receiving instruction can support generalization more effectively than instruction alone or exploration alone.

Study 3 replicated the design of Radovanovic et al. (in preparation) in an online environment. The in-person Lock and Keys Task used in Radovanovic et al. (in preparation) was motorically challenging: Some children struggled to unlock the locks with the keys. Shifting to an online environment allowed children greater ease in unlocking locks by clicking and dragging keys to locks. In addition, the locks and keys were complex stimuli with multiple sides and physical properties to explore. The online version provided easier access to information relevant for learning the unlocking rule. In addition, Radovanovic et al. (in preparation) measured children's exploration behaviors, which involved coding videos from the sessions. The online format allowed children's behaviors to be readily accessible from gameplay data generated after each child participated, without the need for intensive video coding.

Study 3 tested whether 6-year-olds learned and generalize the unlocking rule from the same three learning contexts as in Radovanovic et al. (in preparation): Exploration Then Instruction, Instruction Alone, and Exploration Alone. Study 4 tested this same question with a comparison sample of adult participants. We then compared the results of the prior in-person study to the online study (In-Person versus Online). The methods and analyses of both studies were preregistered (Study 3: https://osf.io/n86k3/?view_only=784ddf8a19a54c88a382903edd1e4b55; Study 4: https://osf.io/68q5v/?view_only=2a979518d52a433bab836ecfae59181b)

Study 3

Methods

Children participated in an online version of the Lock and Keys Task from Radovanovic et al. (in preparation) to test whether exploration, instruction, or exploration then instruction supported children's rule learning and generalization. In the online problem-solving task, as in the in-person task, children learned to unlock locks using keys that followed a two-factor rule. Children were not told the rule; instead, they figured out the rule by exploring, receiving instruction, or both. Children's immediate learning and generalization were then evaluated: Children were tested on their ability to unlock the locks in a test phase. Then, children participated in a generalization phase where they applied the rule to novel locks and keys.

Participants. 120 children (mean age = 6.4 years, range = 6.0-7.0 years; 59 male, 61 female) participated. Children were randomly assigned to one of three conditions with 40 children per condition: Exploration Then Instruction (mean age = 6.4 years; 19 male, 21 female), Instruction Alone (mean age = 6.4 years; 20 male, 20 female), or Exploration Alone (mean age = 6.4 years; 20 male, 20 female). 11 additional children were tested by excluded due to colorblindness (1; the task involved color matching), developmental delay (2), parental interference (1), failure to complete the Lock and Keys Task (1), quitting and restarting the Lock and Keys Task (1), or technological problems with the task (5). Participants were recruited from a database of families who had volunteered to participate in research at a university in a large Midwestern city. The majority of caregivers reported that their children were of European or White-American descent (67; Asian or Asian-American: 18, Hispanic or Latino-American: 7, African or African-American: 3, multiple races or ethnicities: 21, other: 1, did not report: 3). Children's primary caregivers were highly educated (post-graduate degree: 55, bachelor's degree: 46, some college: 11, associate's degree: 3, high school: 1, other: 1, did not report: 3). Data was collected between March 2021 and February 2022.

Procedure. Children were tested individually in a single synchronous Zoom session with an experimenter that lasted approximately thirty minutes. Children participated in the online Lock and Keys Task, then additional tasks were administered which were not central to the main research questions and are not reported here. Caregivers could choose to stay with their children during the session, but were asked not to intervene. All participants were audio and video recorded, and children's computer screens were recorded as they participated in the Lock and Keys Task. Prior to participating, caregivers provided informed consent online and completed a background questionnaire about their children's age, race or ethnicity, language exposure, caregiver education, and whether their child had been diagnosed with colorblindness. At the beginning of the study, caregivers and children provided verbal assent. Families received a $5 Amazon gift card and a virtual certificate to thank them for participating.

Lock and Keys Task. An online problem-solving task based on Radovanovic et al. (in preparation) was programmed for the study in Unity. The game was hosted on a university-maintained website and could be played on a laptop or desktop computer. To access the game, participants navigated to the website and entered a Game ID, which loaded the game in their assigned condition (Exploration Then Instruction, Instruction Alone, or Exploration Alone). Gameplay data (all of children's click, drag, and drop actions with associated timing information) were tracked and automatically stored in text files on a university-maintained server for analyses.

In the Lock and Keys Task, children saw four locks and 20 keys. The locks and keys were digitally-edited photographs of the locks and keys in Radovanovic et al. (in preparation). Each key had a colored drawing on the top (e.g., an orange lion) and keys had different shaped tops. Locks were solid colors (blue, purple, green, orange) and had white stars on them, which were irrelevant to the task. Each lock could be opened by a single key that followed a two-factor rule: 1) the colored drawing on the key matched the color of the lock (color rule), and 2) the key had a circular top (shape rule). Children were not told the rule and learned them through the game.

The Lock and Keys Task had five phases that varied between condition (Practice, Exploration, Instruction, Test, and Generalization). All conditions began with the Practice phase. The Exploration Then Instruction condition included Exploration, Instruction, Test, and Generalization. The Instruction Alone condition included Instruction, Test, and Generalization. The Exploration Alone condition included Test and Generalization (Figure 5).

Exploration Then Instruction	**Practice**	**Exploration** (2 min)	**Instruction** (1 min)	**Test** (4 min)	**Generalization**
Instruction Alone	**Practice**		**Instruction** (1 min)	**Test** (4 min)	**Generalization**
Exploration Alone	**Practice**			**Test** (6 min)	**Generalization**

Figure 5: Study 3 Design. Children were assigned to participate in one of three conditions: Exploration Then Instruction (phases: Exploration, Instruction, Test, Generalization), Instruction Alone (phases: Instruction, Test, Generalization), or Exploration Alone (phases: Test, Generalization). Time spent in each phase is listed with the phase. The image below each phase shows how the locks and keys were arranged on the screen at the start of each phase. The image for Generalization is an example of one of the six Generalization trials.

Practice. The Practice phase gave children experience with the game mechanics for unlocking locks. Children saw one black lock with a red stripe at the top of the screen and a gray key without an image on it at the bottom of the screen. Voiceover instructed children to use the key to unlock the lock, while showing an animation of the key moving to and unlocking the lock: "It's your job to unlock the lock! Click the key [*key highlighted*]. Drag the key to the lock [*key moved to the lock*]. Drop the key on the lock [*lock highlighted as key stopped over the lock*]. Look, the lock unlocked! [*Key unlocked the lock with a clicking sound.*] Now it's your turn." The practice key then moved back to its original position and children had two minutes to unlock the lock. If children struggled to click and drag, the experimenter asked caregivers to teach their children to click and drag. After the Practice phase, caregivers were told not to intervene. All children unlocked the practice lock independently or with help from a caregiver.

When children moved their mouse over the key, yellow highlighting appeared around the key. Then, the key could be clicked, dragged, and dropped anywhere on the screen. When the key was dragged over the lock, the lock highlighted. When children dropped the key on the lock, an animation showed the key upside down and moving under the lock. Then, the top latch of the lock opened and a clicking sound played, followed by voiceover that said, “You unlocked the lock!” The key then appeared in its original orientation under the lock (Figure 6).

Exploration. Children in the Exploration phase of the Exploration Then Instruction condition had two minutes to explore the locks and keys. Children saw four locks at the top of the screen and 20 keys at the bottom of the screen. Children heard voiceover that said, “Use the keys [*keys highlighted*] to unlock the locks [*locks highlighted*].” Children could click, drag, and drop the keys anywhere on the screen, though the locks remained stationary. Exploration ended after all four locks were unlocked or after two minutes had elapsed, whichever occurred first.

As in Practice, keys highlighted when children moved their mouse over them and children could click, drag, and drop the keys anywhere on the screen. If a key was dragged on top of a lock, as in Practice, the lock’s outline highlighted in yellow. When a correct key (i.e., color-matched to the lock and round) was dropped on a lock, an animation showed the key unlocking the lock with the same sound effect and voiceover as in Practice. When an incorrect key (i.e., not color-matched to the lock or not round) was dropped on a lock, an animation showed the key failing to unlock the lock: The key turned upside down and bounced under the lock three times while making a tapping sound. The key then sat under the lock in its original orientation (identical to the end state of successful unlocking; see Figure 6).

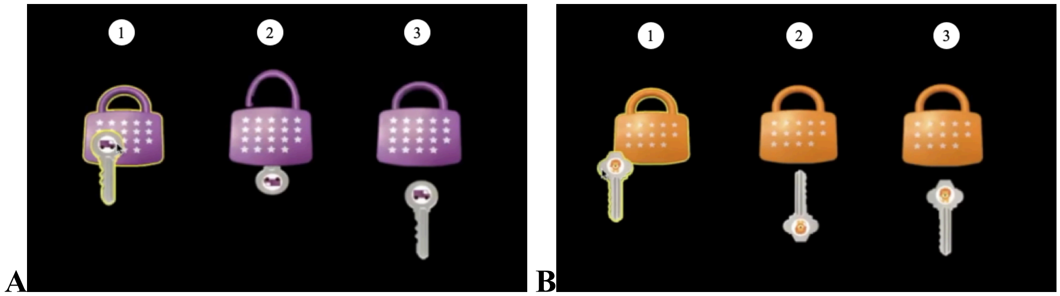


Figure 6: Study 3 Unlocking Examples. (A) Images showing an example sequence of successful unlocking (the round purple key is dropped on the purple lock): The key is selected (key is highlighted) and dropped on a lock (lock is highlighted; 1), animation showed the lock unlocking with unlocking sound (2), the key ends under the lock (3). This sequence also occurred during Practice. (B) Images showing an example sequence of unsuccessful unlocking (the orange clover key is dropped on the orange lock – it is color-matched but not round): The key is selected (key is highlighted) and dragged over a lock (lock is highlighted; 1), animation shows the key bouncing under the lock three times with a tapping sound (2), the key ends under the lock (3).

Instruction. The Instruction phase in the two conditions including Instruction (Exploration Then Instruction and Instruction Alone) showed all four locks being unlocked by the correct keys, giving children evidence about which keys would unlock the locks without explicitly teaching children the color and shape rule. Children saw 20 keys at the bottom of the screen with the four locks at the top. An animation showed each of the correct keys unlocking each lock. Voiceover said, "Now I'll show you how to unlock the locks!" From left to right, each lock was unlocked one at a time: The correct key slid on top of the corresponding lock and unlocked the lock with the same animation and clicking sound as in Exploration. After the lock unlocked, voiceover said, "Look, the lock unlocked!" Then, the key returned to its initial position at the bottom of the screen. This process was repeated for each lock. The entire animation lasted one minute. Clicking and dragging was disabled during this phase.

Test. Following Instruction, children in the Exploration Then Instruction and Instruction Alone conditions were tested on how well they had learned which keys unlocked the locks. Children saw the four locks at the top of the screen the 20 keys at the bottom of the screen (with key positions shuffled relative to Instruction). Children were given four minutes to independently unlock the locks with the keys. The same gameplay mechanics in Exploration were used during Test. Test ended after all four locks were unlocked or after four minutes elapsed, whichever occurred first.

To assess learning from Exploration Alone, children in this condition proceeded directly to the Test phase after Practice, during which they explored and attempted to unlock the locks. Children in the Exploration Alone condition were given six minutes for Test rather than four to act as a time control relative to the other conditions: Children in the Exploration Then Instruction condition had 6 minutes of exposure to the locks and keys (two minutes of Exploration and four minutes of Test), while children in Instruction Alone had only four minutes with the locks and keys during Test.

Generalization. To measure how well children could apply the unlocking rule they had learned to new examples, children in all conditions participated in the Generalization phase after Test. On each of six trials, children saw an image of a novel lock at the top of the screen and four novel keys at the bottom of the screen. One key was correct (color-matched to the lock and round), a second key was color-matched but not round, a third key was round but not color-matched, and a fourth key was neither round nor color-matched (locations randomized across trials but fixed between participants). On each trial, children heard, "Which key [*all keys highlighted*] would unlock the lock [*lock highlighted*]?" Children then dragged a key and dropped it on the lock (as in Exploration). However, no feedback was provided: The dropped key remained on the lock for one second before moving on to the text trial.

Analysis

Measures were extracted from the gameplay data generated after each child participated. The gameplay data included time-stamped information about the start and end time of each phase and total game play. Within each phase, key movements and unlocking attempts (when keys were dropped on locks) were tracked. Data were combined for all participants and organized in R using the dplyr package (Wickham et al., 2018) to extract measures reflecting learning, generalization, and behaviors of interest. See Table 3 for measures, definitions, and descriptive statistics.

Test Performance. The number of locks unlocked out of four possible was operationalized as children's Test Performance. This measure reflected how well children learned which keys unlocked the locks. For children in Exploration Alone, Test Performance reflected the what children learned from exploring without prior exploration or instruction.

Generalization Performance. The number of correctly chosen keys (keys that were color-matched and round) across 6 trials was counted for a measure of Generalization Performance. In addition, Generalization Color Rule was calculated as the number of keys children selected that matched the locks in color despite their shapes (i.e., the number of keys that were correct plus the number of color-matched keys that were not round).

Test Success. Measures were extracted from children's game play to examine whether prior experience exploring or receiving instruction affected children's ability to successfully unlock the locks at Test. The total number of times children successfully unlocked locks was counted (note that successes reflected the *total* number of successful unlocking instances, whereas Test Score was the number of locks out of four possible unlocked). The number of times children failed to unlock locks was also counted (i.e., instances where children dropped incorrect keys onto locks). The number of successes and failures were summed as a measure of Test Total Attempts. Test Success Proportion was calculated as the number of total successes divided by the number of Test Total Attempts, which reflected children's success rate.

Test Failures. Children's failed unlocking attempts at Test were categorized by the evidence that the attempts could provide about the unlocking rule. Keys that were a different color than the lock they were tested on were categorized as Incorrect Color Keys; failures with these keys provides counterfactual evidence that could reveal the color rule. Keys that matched the lock in color but were not round were counted as Correct Color Incorrect Shape Keys. Failures with Correct Color Incorrect Shape Keys provided counterfactual evidence that could reveal the shape rule. The proportion of attempts made with each type of key was calculated: Incorrect Color Keys Proportion and Correct Color Incorrect Shape Keys Proportion.

Exploration Success. Analogous to behaviors measured at Test, exploration behaviors were extracted from the Exploration phase of Exploration Then Instruction to compare behaviors in the online game to the in-person findings from Radovanovic et al. (in preparation; In-Person versus Online). Exploration Total Attempts was extracted and Exploration Success Proportion was calculated.

Phase	Measure	Definition	Mean (SD); range
Exploration*	Exploration Total Attempts	Total number of successful and failed unlocking attempts	10.50 (4.77); 2-23
	Exploration Success Proportion	Number of successful unlocking attempts divided by Total Exploration Attempts	.28 (.26); 0-.80
Test	Test Performance	Number of locks unlocked (maximum 4)	3.45 (1.17); 0-4
	Test Total Attempts	Total number of successful and failed unlocking attempts	11.86 (8.80); 2-45
	Test Success Proportion	Number of successful unlocking attempts divided by Total Test Attempts	.48 (.29); 0-1
	Incorrect Color Keys Proportion	Number of incorrect unlocking attempts made with keys that did not match the lock in color divided by Test Total Attempts	.25 (.31); 0-1
	Correct Color Incorrect Shape Keys Proportion	Number of incorrect unlocking attempts made with keys that did match the lock in color but were not round divided by Test Total Attempts	.27 (.23); 0-1
Generalization	Generalization Performance	Number of correct keys (maximum: 6)	4.02 (1.61); 0-6
	Generalization Color Rule	Number of color-matched keys (maximum: 6)	5.76 (.50); 4-6

Table 3: Study 3 Measures. Measure calculated from gameplay data during each phase of the study with associated definition. Descriptive statistics are provided across conditions (mean, standard deviation, and range). *Measures from Exploration were used in the analyses comparing the in-person study to Study 1.

Analysis Strategy. As preregistered, the analyses tested whether children who explored or were instructed differed in their learning or generalization. The analyses were conducted with linear regressions separately predicting Test Performance and Generalization Performance. Predictors were planned contrasts that tested for an effect of **Exploration** (Exploration Then Instruction: 1, Instruction Alone: -1, Exploration Alone: 0) and **Instruction** (Exploration Then Instruction: 1, Instruction Alone: 1, Exploration Alone: -2). All proportion measures were arcsine square-root transformed prior to the analyses.

Results

Age ($F(2, 117) = 0.52$, $p = .597$) and gender (male vs. female: $\chi^2(2) = 0.07$, $p = 0.970$) did not differ by condition. The analyses collapsed across these factors.

Test Performance. On average, children unlocked 3.5 out of 4 locks at Test. There was a significant effect of Exploration on Test Performance such that children who explored prior to instruction unlocked more locks at Test than those who did not explore ($\beta = .263$, SE = .129, $p = .044$; no significant effect of Instruction: $\beta = -0.01$, SE = .075, $p = .867$; Figure 7A). Exploring before receiving instruction increased the number of locks children unlocked at Test.

Generalization Performance. Children scored on average 4.0 out of 6 at Generalization. Children in all conditions selected color-matched keys at above-chance levels (Generalization Color Rule performance was significantly above chance (50%): all p's < .001). Compared to the stringent 50% chance level, children's Generalization Performance was significantly above chance in all conditions (all p's < .01). There was no evidence of an effect of Exploration ($\beta = 0.03$, SE = 0.18, $p = .890$) or Instruction ($\beta = 0.16$, SE = .10, $p = .129$) on Generalization Performance (Figure 7B). Children in all conditions generalized the rule at above-chance levels.

Generalization Performance: Perfect Test. We then exploratorily analyzed Generalization Performance for children who unlocked all four locks at Test (N = 92, or 77% of participants; no difference in number of participants by condition: Exploration Then Instruction: 33, Instruction Alone: 29; Exploration Alone: 30; $\chi^2(2) = 0.28$, $p = 0.868$). Children's Generalization Performance was above chance in all conditions (50%: all p's < .01). Among children who had successfully demonstrated rule-learning at Test, there was a significant effect of Instruction on Generalization Performance ($\beta = 0.23$, SE = 0.11, $p = .043$; no effect of

Exploration: β = -0.02, SE = 0.13, p = .917; Figure 7C). Children who had received instruction generalized the rule more accurately than those who had only explored.

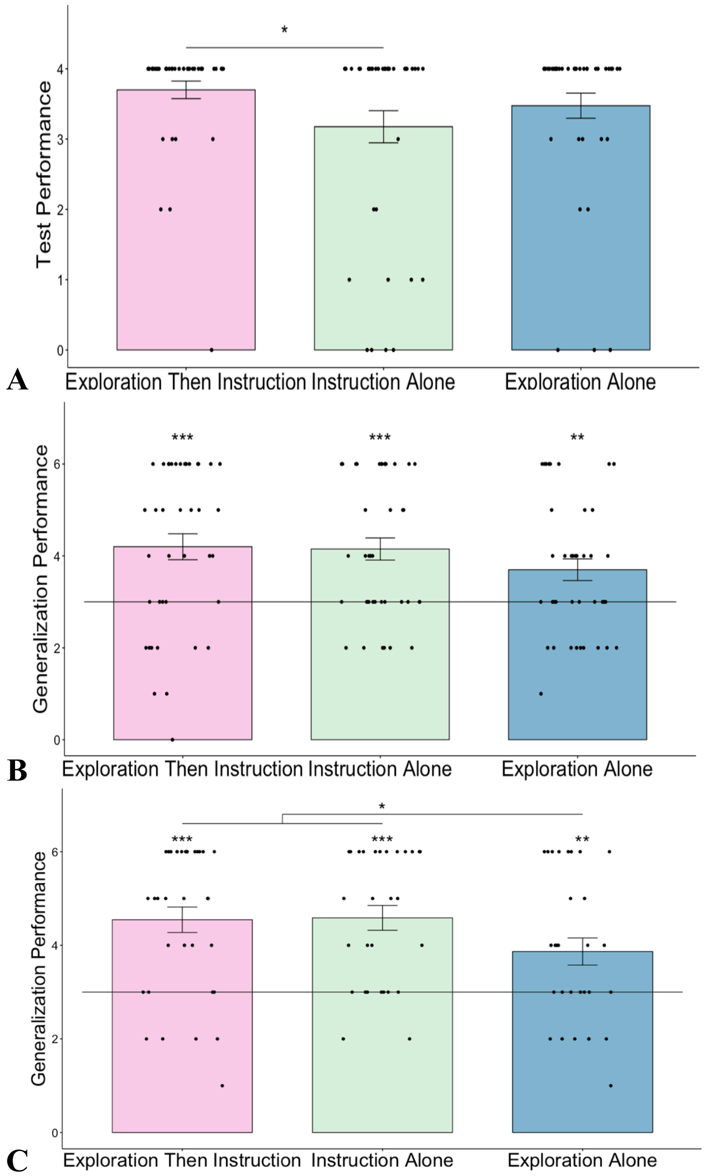
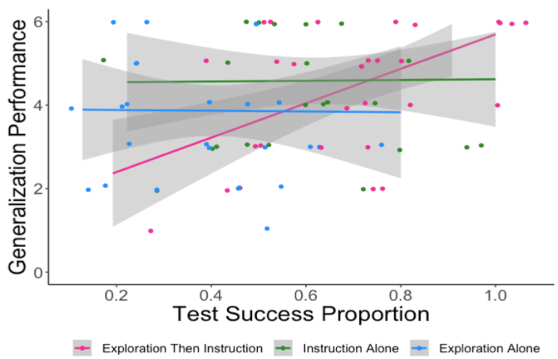


Figure 7: Study 3 Test and Generalization Performance. (A) Test Performance, (B) Generalization Performance for the full sample, and (C) Generalization Performance only for those who unlocked all locks at Test (perfect Test Performance). Generalization chance is 50% (3). Error bars: +/- 1 Standard Error. *p < .05, **p < .01, ***p < .001.

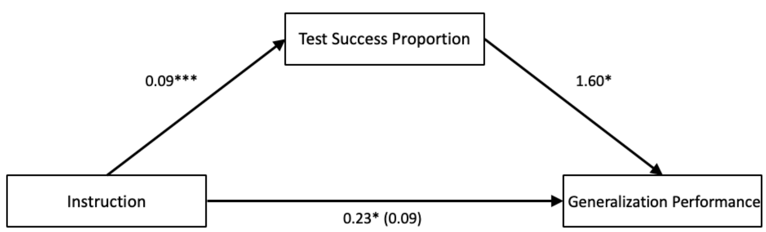
We asked whether children who were more successful in their unlocking attempts during exploration subsequently unlocked more locks during the Test phase. Indeed, children who had a higher proportion of successful unlocking attempts during exploration unlocked more of the locks during Test ($\beta = 1.01$, SE = 0.31, $p = .003$). However, it should be noted that most children in this condition unlocked all four locks at Test (33 of 40 children).

We next asked whether instruction supported generalization by making children more successful in their unlocking attempts during Test. Among children who unlocked all four locks at Test, children who only explored made significantly more unlocking attempts compared to children who were instructed (Total Attempts: $\beta = -2.67$, SE = 0.43, $p < .001$), but children who were instructed were significantly more successful than children who only explored ($\beta = 0.11$, SE = 0.02, $p < .001$), and those who explored prior to instruction were marginally more successful than those who were only instructed ($\beta = 0.07$, SE = 0.04, $p = 0.09$; Figure 8C). In turn, children who were more successful scored higher at Generalization ($\beta = 1.46$, SE = 0.49, $p = .004$; with added interactions: the relation was stronger for children who explored before instruction; Exploration: $\beta = 1.35$, SE = 0.62, $p = .031$; no interaction with Instruction: $\beta = 0.46$, SE = 0.56, $p = .408$; Figure 8A). Further, the effect of Instruction on Generalization was fully mediated by children's success at Test (average causal mediation effect: $\beta = 0.14$, 95% CI [0.01, 0.29], $p = .034$; Figure 8B). Children who were instructed were more successful in their unlocking attempts at Test, which supported generalization.

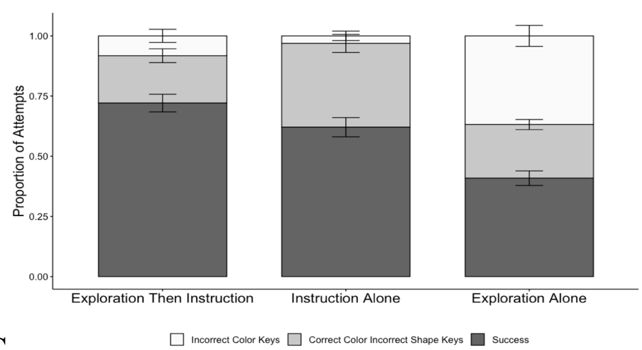
We then investigated the nature of children's failed unlocking attempts to examine whether children continued gathering evidence about the color and shape components of the rule during Test. Among children who unlocked all four locks, children who only explored made proportionally more attempts with Incorrect Color Keys than those who were instructed ($\beta = -0.17$, SE = 0.02, $p < .001$; no effect of Exploration: $\beta = 0.05$, SE = 0.03, $p = .141$), suggesting that when children only explored, they were gathering more evidence that would inform them about the color part of the rule than instructed children. In contrast, children in Instruction Alone made proportionally more attempts with Correct Color Incorrect Shape Keys than those in Exploration Then Instruction (β = -0.05, SE = 0.03, *p* = .001; no effect of Instruction: β < 0.01, SE = 0.02, *p* = .912; Figure 8C), suggesting that children who were instructed without prior exploration tested more keys to disambiguate the shape part of the rule than those who explored before instruction. In other words, exploring and receiving instruction led children to be the least likely to gather evidence about color or shape; exploration and instruction led children to intuit both parts of the unlocking rule better than those in other conditions.



A



B



C

Figure 8: Study 3 Test Success, Generalization, and Attempts. For participants who scored perfectly at Test: (A) relation between Test Success Proportion and Generalization Performance

Figure 8: Study 3 Test Success, Generalization, and Attempts (continued): separated by condition, (B) causal mediation model showing the effect of Instruction on Generalization Performance via Test Success Proportion, and (C) Proportion of unlocking attempts that were successful and failed (by failure type: Incorrect Color Keys and Correct Color Incorrect Shape Keys). Shaded areas represent 95% confidence intervals. Error bars: +/- 1 Standard Error. $*p < .05$, $**p < .01$, $***p < .001$.

Results Summary. In sum, exploration supported learning, and instruction supported generalization. Children who explored before instruction were better able to unlock all four locks at Test. In addition, for children who learned which keys unlocked the locks, instruction supported generalized more than exploring alone. Instruction improved generalization by increasing children's success rate during the Test phase. In turn, more success was related to better generalization performance. This relation was particularly strong for children who explored before instruction. Thus, instruction improved the success of children's subsequent behaviors, which boosted their ability to generalize the unlocking rule.

Interestingly, children appeared to continue learning the unlocking rule during the Test phase. Children had the best understanding of both parts of the rule after exploring and receiving instruction; they rarely tested keys that would provide evidence to disambiguate the color or shape parts of the rule. Children who were only instructed tested more color-matched keys that were the wrong shape than those who had explored before instruction, suggesting that they had intuited that color mattered for unlocking to some extent, but were still gathering evidence that would disambiguate the shape part of the rule. In contrast, children who only explored tested all the keys to provide evidence about both parts of the rule. In sum, instruction allowed children to infer that color mattered for unlocking, and exploration before instruction additionally supported children's understanding about the shape part of the rule. Children engaged in targeted testing behaviors during the exploratory Test phase to disambiguate the parts of the rule they had not yet learned, suggesting that children continued learning during the Test phase.

Study 4

In Study 4, a comparison sample of adults was tested on the same Lock and Keys Task as children in Study 3 to examine whether adults learned and generalize the rule more effectively after exploring, receiving instruction, or both.

Methods

Participants. Adults in the United States with an approval rating of at least 75% on Amazon Mechanical Turk (MTurk) were eligible to participate in the study. 305 participants were included in the analyses (mean age = 37.4 years, range = 20-77 years; 188 male, 115 female, 2 NA or other) with approximately 100 participants randomly assigned to the same three conditions as in Study 1 (Exploration Then Instruction = 102, Instruction Alone = 102, Exploration Alone = 101). The majority of participants were of European American descent (196; 34 African or African-American, 27 Asian or Asian-American, 18 Native American, 17 Hispanic or Latino-American, 12 multiple races or ethnicities, 1 other) and held bachelor's degrees (163; 49 post-graduate degrees, 42 attended some college, 22 associate's degrees, and 29 attended high school). All participants were fluent English speakers. Additional participants were excluded because they were colorblind (25), were not fluent in English (4), or provided data that were outliers (19; see Analysis). Data was collected in March and April of 2021.

Procedure. Eligible participants were directed to a Qualtrics survey with instructions that asked them to use a desktop or laptop computer and turn their audio on. They were then randomly assigned a Game ID that they entered into the game website, which loaded the Lock and Keys Task in the correct condition. Participants then participated in the Lock and Keys Task identically to children in Study 3. At the end of the task, participants were given a code that they entered into MTurk and the Qualtrics survey, which matched participants' survey responses with their gameplay data. Participants then answered questions about their age, gender, race or ethnicity, education level, spoken languages, and whether they had been diagnosed with colorblindness. Participants were provided $1.50 as compensation. The study took less than 10 minutes to complete.

Analysis

The same measures extracted from children's data in Study 3 were extracted from adults' gameplay data in Study 4: Test Performance, Generalization Performance, Generalization Color Rule, Test Total Attempts, Test Success Proportion, and the proportion of failures performed with each type of key at Test: Incorrect Color Keys Proportion and Incorrect Shape Keys Proportion (Table 4).

Phase	Measure	Definition	Mean (SD); range
Test	Test Performance	Number of locks unlocked (maximum: 4)	3.90 (.50); 0-4
	Total Test Attempts	Total number of successful and failed unlocking attempts	13.10 (15.75); 1-86
	Test Success Proportion	Number of successful unlocking attempts divided by Total Test Attempts	.61 (.31); 0-1
	Incorrect Color Keys Proportion	Number of incorrect unlocking attempts made with keys that did not match the lock in color divided by Test Total Attempts	.20 (.31), 0-1
	Correct Color Incorrect Shape Keys Proportion	Number of incorrect unlocking attempts made with keys that did match the lock in color but were not round divided by Test Total Attempts	.19 (.17), 0-.60
Generalization	Generalization Performance	Number of correct keys (maximum: 6)	4.05 (1.79); 0-6
	Generalization Color Rule	Number of color-matched keys (maximum: 6)	5.71 (.67); 0-6

Table 4: Study 4 Measures. Measure calculated from gameplay data during each phase of the study with associated definition. Descriptive statistics are provided across conditions (mean, standard deviation, and range).

Outlier Detection. In addition to the final sample of 305, 19 participants were excluded for contributing data that were outliers. Piloting revealed that adults' game play was more variable than children's; some participants spent more time than expected on the task or clicked and dragged many keys in each phase. We therefore extracted additional measures from adults' gameplay to determine outliers, as preregistered. Specifically, we extracted time measures (Total Time, Pause Time, and time in each phase of the game). During Exploration and Test, we also extracted the number of keys participants clicked and dragged (Total Keys). Outliers (more than 3 standard deviations from the mean) were found in the following measures: Total Time (3), Pause Time (8), and Generalization Time (3). Participants who clicked and dragged more than 100 Total Keys at Test (5) were excluded based on visual inspection of box plots.

Analysis Strategy. The preregistered analyses in Study 3 were conducted for Study 4. The analyses tested whether exploration or instruction were related to adults' learning or generalization with linear regressions separately predicting Test Performance and Generalization Performance from the planned contrasts of **Exploration** and **Instruction**. The proportion measures were arcsine square-root transformed prior to the analyses.

Results

Age ($F(2, 302) = 0.27$, $p = .761$) and gender (male versus female: $\chi^2(2) = 0.73$, $p = 0.690$) did not differ significantly by condition. The analyses collapsed across these factors.

Test Performance. On average, adults unlocked 3.9 out of 4 locks at Test. All participants except 16 unlocked all four locks at Test; with this limited variability in Test Performance, it was unsurprising that no significant effects of Exploration ($\beta = 0.02$, SE = 0.03, $p = .483$) or Instruction ($\beta = 0.04$, SE = 0.02, $p = .060$) emerged (Figure 9A).

Generalization Performance. On average, adults scored 4.1 out of 6 during Generalization. Adults in all conditions generalized at above-chance levels (Generalization Color Rule: chance = 50%: all p's < .001; with this chance level applied to Generalization Performance, adults in all conditions performed above chance; all p's < .01). Interestingly, there was a significant effect of Instruction on Generalization Performance ($\beta = 0.28$, SE = 0.07, $p < .001$; no significant effect of Exploration: $\beta = 0.07$, SE = 0.12, $p = .548$; Figure 9B). Like children, adults who were instructed generalized more accurately than those who only explored.

Generalization Performance: Perfect Test only. To parallel the analysis in Study 1, adults' Generalization Performance was analyzed within the subgroup of participants who unlocked all four locks at Test. 289 participants unlocked all four locks at Test (95% of participants), and performance in all conditions was above chance (50%: all p's < .05). Similar results emerged from this analysis as the analysis of the full dataset: Adults who were instructed scored higher than those who only explored ($\beta = 0.30$, SE = 0.07, $p < .0001$; no significant effect of Exploration: $\beta = 0.06$, SE = 0.13, $p = .617$; Figure 9C). Adults, like children, better transferred the unlocking rule to new contexts when they had received instruction.

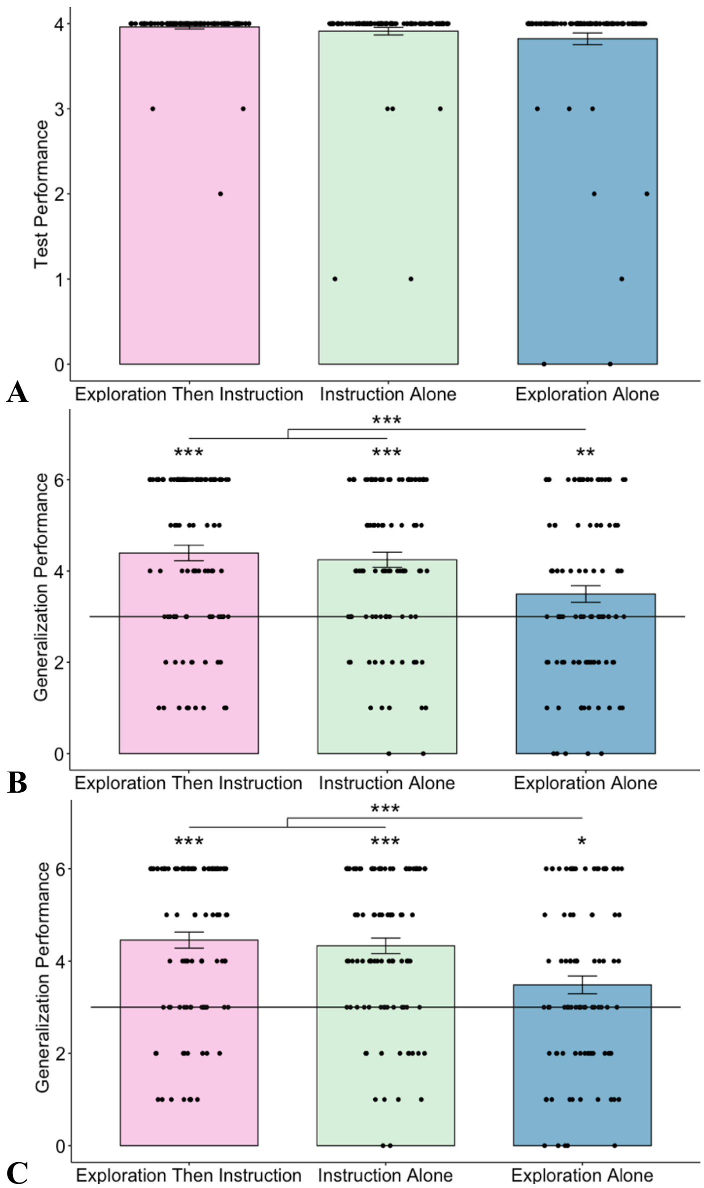


Figure 9: Study 4 Test and Generalization Performance. (A) Test Performance, (B) Generalization Performance for the full sample, and (C) Generalization Performance for participants who unlocked all locks at Test (perfect Test Performance). Generalization chance is 50% (3). Error bars: +/- 1 Standard Error. **p* < .05, ***p* < .01, ****p* < .001.

To understand why instruction supported Generalization, as in Study 3, we tested whether participants who had received instruction were more successful in unlocking locks at Test than those who were not instructed. Participants who only explored made more attempts to unlock the locks than participants who received instruction (Instruction: β = -5.79, SE = 0.55, $p < .001$; no effect of Exploration: $\beta = -0.73$, SE = 0.95, $p = .440$). Participants who explored prior to instruction were the most successful in their attempts (Exploration: $\beta = 0.13$, SE = 0.03, $p < .001$), and participants who only explored were the least successful (Instruction: $\beta = 0.16$, SE = 0.01, $p < .001$; Figure 10C). In turn, greater unlocking success was related to better Generalization Performance ($\beta = 1.44$, SE = 0.23, $p < .001$), particularly for participants who were instructed ($\beta = 0.37$, SE = 0.19, $p = .050$; no interaction with Exploration: $\beta = -0.17$, SE = 0.35, $p = .627$; Figure 10A). Further, more successful attempts at Test fully mediated the relation between Instruction and Generalization Performance ($\beta = 0.23$, 95% CI [0.14, 0.33], $p < .001$; Figure 10B). As with children, instruction improved adults' Test successes, which supported Generalization Performance.

The types of failed unlocking attempts adults made at Test were similar to children's failures. Adults who only explored tested proportionally more Incorrect Color Keys than adults who received instruction ($\beta = -0.20$, SE = 0.01, $p < .001$). Without instruction, adults were more likely to gather evidence about the color and shape rule. In contrast, those who were instructed without prior exploration tried proportionally more Correct Color Incorrect Shape Keys than those who explored before instruction ($\beta = -0.12$, SE = 0.02, $p < .001$; Figure 10C). Adults who were only instructed gathered evidence that would disambiguate the shape rule. When adults explored before instruction, they were the most successful and tested the fewest incorrect keys, indicating that they inferred both the color and shape parts of the rule more effectively than those in the other conditions.

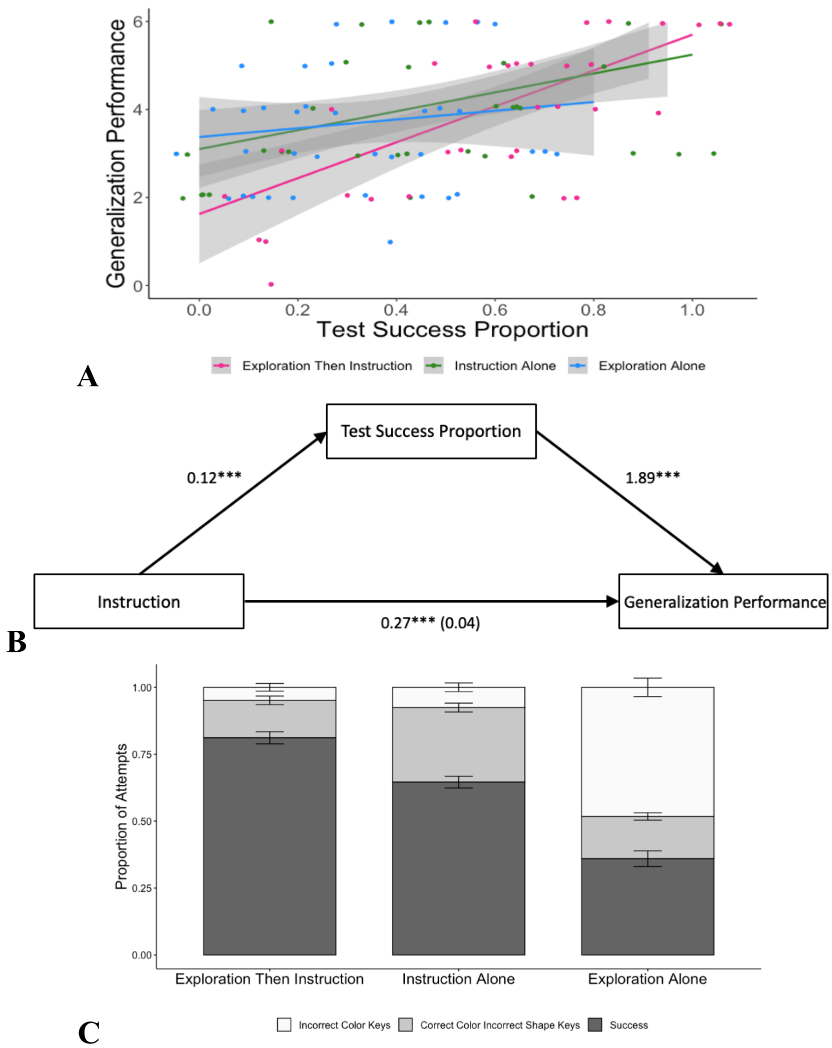


Figure 10: Study 4 Test Success, Generalization, and Attempts. (A) Relation between Test Success Proportion and Generalization Performance by condition, (B) causal mediation model showing the effect of Instruction on Generalization Performance via Test Success Proportion, and (C) proportion of unlocking attempts that were successful and failed (by failure type: Incorrect Color Keys and Correct Color Incorrect Shape Keys). Shaded areas represent 95% confidence intervals. Error bars: +/- 1 Standard Error.*$p < .05$, **$p < .01$, ***$p < .001$.

Results Summary. The results of Study 4 with adults parallels the findings from Study 3 with children. Like children, adults who were instructed generalized the rule more effectively than those who only explored. The mechanism supporting the relation between instruction and generalization was adults' success at Test: Instructed adults were more successful in unlocking the locks at Test, which in turn supported generalization. Further, adults' unlocking failures at

Test revealed that after exploring and receiving instruction, adults had the best sense of both parts of the unlocking rule; they engaged in the least testing to disambiguate either part of the rule. Instruction without exploration allowed adults to infer that color mattered for unlocking, though they engaged in targeted testing that would provide evidence to disambiguate the shape part of the rule. In contrast, those who only explored gathered evidence about both parts of the rule.

Nearly all adults successfully unlocked all four locks during the Test phase; this ceiling-level performance made it impossible to examine whether exploration or instruction affected adults' learning. Yet, adults who explored prior to instruction were more successful at Test, failing least often with the wrong keys, than those who were only instructed; exploring prior to receiving instruction improved adults' performance. In sum, instruction supported adults' successes, which in turn improved generalization, and we have some evidence that exploring before instruction further improved adults' understanding of the rule.

In-person versus Online

The results from Study 3 differ from the prior study conducted using an in-person version of the Lock and Keys Task (Radovanovic et al., in preparation). We therefore compared the findings of the in-person and online studies.

Methods

99 children (mean age = 6.5 years) participated in Radovanovic et al. (in preparation) in a laboratory of a research university with an experimenter in one of three conditions: Exploration Then Instruction (N = 39), Instruction Alone (N = 40), or Exploration Alone (N = 20). In the in-person Lock and Keys Task, children saw four locks and 18 keys (Figure 11). All children unlocked a practice lock first. Children in the Exploration Then Instruction condition then had

two minutes to explore the locks and keys prior to receiving instruction. Children in Instruction Alone proceeded directly to Instruction. During Instruction, the experimenter demonstrated successfully unlocking the four locks. After Instruction, children in both conditions with instruction participated in a four-minute Test phase during which they attempted to unlock the locks. Children in the Exploration Alone condition participated in the Test phase only for six minutes. Children in all conditions then participated in a six-trial Generalization phase where children selected which of four of novel keys would unlock each novel lock.

Figure 11: In-Person Study Materials. (A) Image taken from video-recorded in-person session: Children were presented with a pile of keys and the four locks at the beginning of the Exploration and Test phases. (B) Locks and keys from the live task with the correct key (round and color-matched) on top of the lock it unlocked. (C) Example of one of six Generalization trials.

Coding

As in Study 3, children's Test Performance was the number of locks (out of four) that children successfully unlocked during Test. Generalization Performance was the number of correct keys children chose at Generalization across six trials, and Generalization Color Rule was the number of color-matched keys children chose out of six. Failures to unlock locks were difficult to discern in children's behavior, so opportunities where unlocking attempts could occur were identified. During Exploration and Test, the time children spent holding locks and keys was coded. Overlapping time windows when children had a lock in one hand and key in the other hand were extracted. The total number of overlapping lock and key times were counted and operationalized as Total Unlocking Attempts (for Exploration and Test). In addition, the total number of times children unlocked locks was coded during Exploration and Test. Exploration and Test Success Proportions were calculated by dividing the total number of locks unlocked by the number of unlocking attempts during that phase (Table 5).

Phase	Measure	Definition	Mean (SD); range
Exploration	Exploration Total Attempts	Number of overlapping instances when a lock and a key were held	16.05 (8.39); 3-33
	Exploration Success Proportion	Number of successful unlocking attempts divided by Total Exploration Attempts	.12 (.17); 0-.67
Test	Test Performance	Number of locks unlocked (maximum: 4)	3.25 (1.29); 0-4
	Total Test Attempts	Number of overlapping instances when a lock and a key were held	23.05 (21.81); 1-111
	Test Success Proportion	Number of successful unlocking attempts divided by Total Test Attempts	.31 (.27); 0-1
Generalization	Generalization Performance	Number of correct keys (maximum: 6)	3.16 (1.86); 0-6
	Generalization Color Rule	Number of color-matched keys (maximum: 6)	5.80 (.45); 4-6

Table 5: In-Person Study Measures. Measure calculated from the in-person study during each phase with associated definition. Descriptive statistics are provided across conditions (mean, standard deviation, and range).

Analysis

Linear regressions were run separately predicting Test Performance and Generalization Performance by study context (in-person or online), planned contrasts (Exploration and Instruction), and the interaction between context and contrast. Proportions were arcsine square-root transformed prior to the analyses.

Results

Test Performance. Radovanovic et al. (in preparation) found that children unlocked more locks at Test when they received *instruction*. Online, those who *explored* unlocked more locks at Test. Between study contexts, children unlocked more locks at Test online (mean = 3.5) than in-person (mean = 3.3; β = 0.38, SE = 0.16, p = .020). This effect was driven by the number of locks children unlocked in the Exploration Alone condition: There was a significant interaction between study context (in-person or online) and Instruction such that online, children in the Exploration Alone condition unlocked more locks than children in the same condition in-person (β = -0.47, SE = 0.12, p < .001; Figure 12). In other words, without instruction or prior exploratory experience, children were more successful in unlocking locks online than in-person.

Generalization Performance. Radovanovic et al. (in preparation) found that children who *explored* generalized more accurately, whereas online, children who received *instruction* generalized more accurately. Children generalizing the rule at higher rates online (mean = 4.02) than in-person (mean = 3.2; β = 0.96, SE = 0.24, p < .001; Figure 12), though there was no evidence of interactions between the study context and planned contrasts (p's > .072).

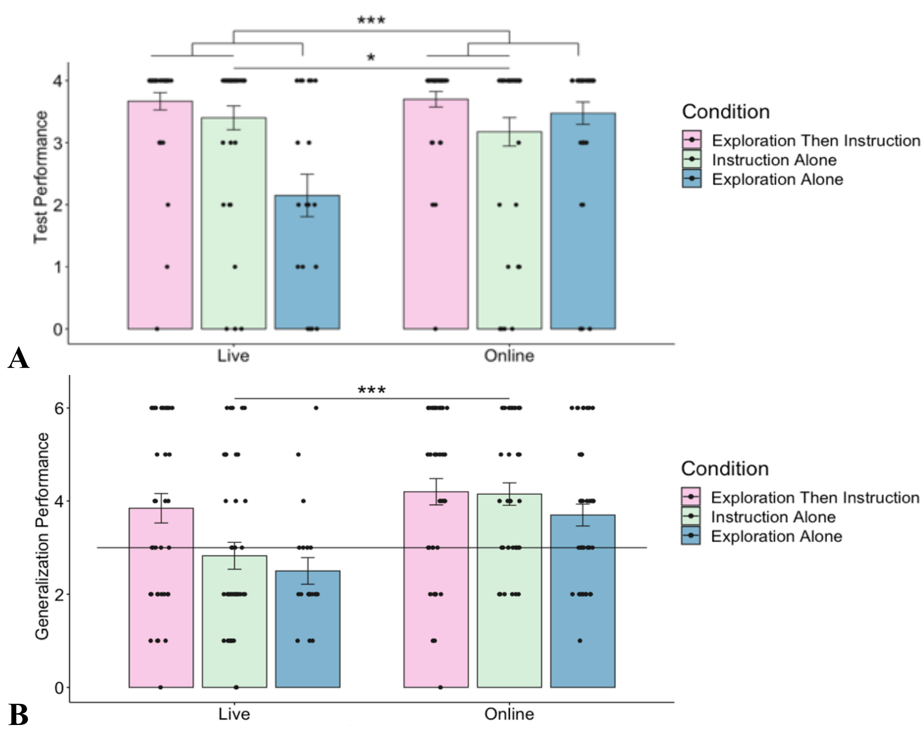


Figure 12: In-Person and Study 3 Test and Generalization Performance. (A) Test Performance by study context and condition. (B) Generalization Performance by study context and condition. Generalization chance is 50% (3). Error bars: +/- 1 Standard Error.*p < .05, **p < .01, ***p < .001.

Unlocking Attempts. To understand why children unlocked more locks online than in-person, we compared children's unlocking successes across contexts. During the Exploration phase for children in the Exploration Then Instruction condition and the Test phase for those in Exploration Alone, children made more attempts in-person than online (Exploration Then Instruction: $\beta = -5.55$, SE = 1.53, $p < .001$; Exploration Alone: $\beta = -33.6$, SE = 4.83, $p < .001$). But, children were more successful in their attempts online than in-person (Exploration: online: average = .28; live: average = .12; $\beta = 0.16$, SE = 0.05, $p = .002$; Exploration Alone: online: average = .33; live: average = .09; $\beta = 0.24$, SE = 0.05, $p < .001$; Figure 13). Children were more successful in unlocking locks independently online than in-person; this suggests that the information value of exploring was higher online than in-person.

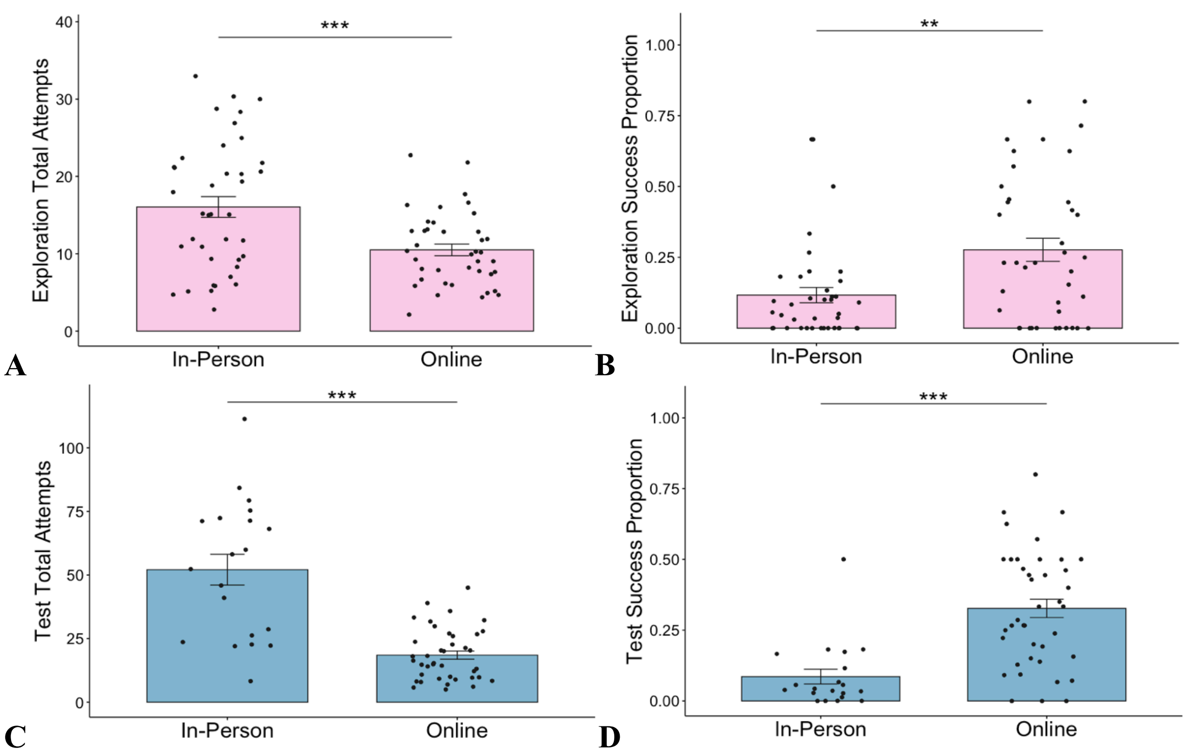


Figure 13: In-Person and Study 3 Attempts and Success Proportion. (A) Exploration Total Attempts by study context, (B) Exploration Success Proportion by study context, (C) Test Total Attempts in the Exploration Alone condition by study context, and (D) Test Success Proportion in the Exploration Alone condition by study context. Error bars: +/- 1 Standard Error.**p* < .05, ***p* < .01, ****p* < .001.

Results Summary. Online, children learned more when they explored and generalized better after instruction. In person, children learned more following instruction and generalized more accurately after exploration. This was likely because the information value of exploring differed between contexts: Online, children unlocked more locks through independent exploration and were more successful compared with children who participated in-person. This suggests that children could gain more information from their actions online than in-person.

Discussion

Exploration and instruction together supported children's and adults' learning and generalization of rule in an online problem-solving game. In Study 3, exploration improved children's learning, and in Studies 3 and 4, instruction improved children's and adults' generalization. Exploration improved children's learning by allowing children to discover how to unlock the locks independently. Instruction improved children's and adults' ability to successfully unlock locks during the Test phase, and unlocking success in turn improved their ability to apply what they had learned to new situations. For children and adults, independent success guided by instruction was key for supporting generalization.

Exploration before instruction supported children's learning. Children who were more successful in unlocking locks as they explored demonstrated better test performance. This suggests that children who figured out how to unlock more locks learned which keys unlocked the locks. This finding aligns with work on discovery-based learning. Discovery has been described as central for learning from exploration (Dean et al., 2007); self-discovered information is recalled better than information told to learners (the "generation effect"; Slamecka & Graf, 1978). In contrast, prior studies have found that when children explored before instruction, they made mistakes that they could then compare to the instruction they received (e.g., DeCaro & Rittle-Johnson, 2012). However, “productive failure” did not drive the results of the present study; instead, children who were more successful while exploring demonstrated better understanding of the material to be learned.

During the test phase, learners who had explored prior to receiving instruction were the most targeted in their unlocking attempts. Children and adults rarely tried incorrect keys, suggesting that they did not need to continue gathering evidence to learn the unlocking rule. Indeed, research suggests that exploring could give children an opportunity to examine the problem space and engage in goal-directed searches for information (Saylor & Ganea, 2018). Exploration likely allowed learners to attend to relevant information, which prepared them to receive instruction (Markant et al., 2016) and subsequently achieve success at Test.

In Studies 3 and 4, children and adults who were instructed better generalized the unlocking rule to novel locks and keys. After instruction, learners were more successful in their unlocking attempts compared to those who only explored; this unlocking success supported generalization, particularly for children who explored and were instructed, and for adults who were instructed. Instruction therefore guided learners’ subsequent actions towards success. Instruction has been shown to limit subsequent behavior; for example, children who were instructed tended to perform instructed rather than novel actions (Bonawitz et al., 2011). Instruction also reduces the demands of unguided exploration (Sweller et al., 1998).

Specifically, learners who were instructed (compared to those who only explored) were less likely to try keys that did not match the locks in color. This suggests that after learners saw the instructions, they inferred that the color of the keys was an important feature of the unlocking rule. Children and adults then engaged in focused testing of color-matched keys. Testing color-matched keys led learners to see the necessary contrastive information: Some color-matched keys unlocked the locks, but other color-matched keys did not unlock the locks. Contrasting color-matched incorrect shape keys and color-matched round keys likely led to a full understand of the unlocking rule: color-matched, round keys unlocked the locks.

Though the test phase was designed to measure how well children and adults had learned which keys unlocked the locks, learners continued learning the unlocking rule during this phase. Exploration and instruction appeared to support learning of both parts of the rule, but learners who were only instructed continued testing keys that were the wrong shape. In contrast, those who only explored tested incorrect-color and incorrect-shape keys, suggested that they used trial-and-error to learn both parts of the rule during the Test phase. Since learners continued to engage in hypothesis-drive testing at Test, this suggests that they were continuing to learn the rule.

In-person versus Online Results

Interestingly, exploration and instruction facilitated children's learning and generalization in different ways in-person (Radnovanovic et al., in preparation) than online (Study 3). In-person, instruction supported children's learning and exploration supported generalization. In contrast, the opposite results were found online: Exploration supported learning, while instruction enhanced generalization. The differences are likely due to how informative children's actions were in each context. Children's actions yielded success more often online than in-person, suggesting that children could gain more information from each action online than in-person. When children's actions were informative, children learned effectively when they had more time to explore the material to be learned, likely because they could make discoveries. In the in-person environment where children's actions were less informative for discovering new information, instruction improved their learning.

The difference in information value of actions was likely due to high motor demands and greater degrees of freedom in-person than online. Motorically, children often struggled to fit keys into locks in-person. Though some children also struggled to click and drag online, each unlocking attempt was more straightforward: Once children mastered clicking and dragging, they could make many attempts using the same motor pattern. Such a pattern would have been more challenging to master in-person when keys needed to be turned and positioned in one hand while locks were simultaneously positioned at the correct angle in the other hand. In person, children could examine the feel of the key shapes and teeth and examine both sides of the locks and keys. Online, children saw a 2D representation of the materials: The colored images on the keys were always facing the child, the locks could not be moved, and none of the materials could be rotated, which reduced the degrees of freedom online.

However, these findings do suggest that in-person, children could potentially figure out how to unlock the locks if they were given enough time or if the task complexity had been reduced. In parallel, if the exploration time had been restricted online, fewer children would have been able to unlock all the locks independently. Given enough time, few degrees of freedom, and minimal task complexity, children could learn through independent exploration. If children instead have limited time, higher task demands, and more degrees of freedom, they may fail to learn through their own actions.

Limitations and Future Directions

This study has theoretical and practical limitations that are addressed in Chapter 4 (Study 5: children, Study 6: adults). Studies 3 and 4 in this chapter tested the effects of exploration prior to instruction, but not instruction followed by exploration. Further, guided exploration could support generalization because it is an active process, but it is unknown whether a comparable observational experience would be equally effective. Practically, learners in all conditions in this chapter who were instructed also explored, minimally during the exploratory Test phase. Studies 5 and 6 in Chapter 4 tested exploration before and after instruction, and isolated active and observational exploration experiences. Studies 5 and 6 also carefully controlled the amount of time children were exposed to the locks and keys, and implemented a true test phase that did not allow additional exploration.

Conclusion

Overall, the results of Studies 3 and 4 suggest that learners' independent successes improved learning and generalization. Children who successfully explored prior to instruction learned more, and children and adults whose exploration was guided to success by instruction generalized more accurately. Together, these studies suggest that situations that allow learners to be independently successful can promote their learning of taught information and transfer of learned material to new situations.

Chapter 4: Revisiting Online Problem-Solving: Timing of Exploration and Instruction, Active versus Observational Exploration

Chapter 3 provided evidence that exploration and instruction together supported learning and generalization by increasing learners' independent successes. In a problem space where children's actions were informative for discovery, children who had more time to explore the problem learned the rule more effectively. Instruction also guided learners' behaviors towards success, which in turn supported information transfer. Yet, unanswered theoretical questions remain from these findings regarding (1) the timing of exploration relative to instruction and (2) active versus observational exploration. Additional practical issues also remained, including (3) isolating exploration from instruction, (4) controlling exposure time to the material to be learned, and (5) a true test phase. These theoretical and practical follow-up questions were addressed in Study 5 with children and Study 6 with a comparison sample of adults.

Studies 3 and 4 tested whether exploration prior to instruction supported learning. Exploration prior to instruction could give children an opportunity to examine the problem space (Saylor & Ganea, 2018), develop and test hypotheses (Gopnik & Wellman, 2012), and attend to relevant information (Markant et al., 2016) prior to receiving instruction. In contrast, exploration *after* instruction could also support children's learning. To maximize efficient learning in classroom settings, some have argued that instruction should precede exploration (Kirschner et al., 2006) so children do not have to generate problem-solving strategies themselves (Clark, 2009). Others similarly argued that children should see examples before exploring to focus children's attention, connect learning to prior knowledge, and enhance motivation (Savery & Duffy, 2001). Instruction prior to exploration benefitted children's learning of math equivalence by providing conceptual information and activating common misconceptions (Fyfe et al., 2014).

Instruction also narrowed children's behavior during exploration (Bonawitz et al., 2011) and reduced the demands on children's exploration (Sweller et al., 1998). Study 5 therefore included conditions where children explored before (Exploration Then Instruction) and after (Instruction Then Exploration) instruction.

In addition, Studies 5 and 6 tested whether active or observational experience supported learning and generalization. Active exploration could support learning outcomes because it includes motor engagement (here, clicking and dragging), which provide the learner with an embodied experience with the information to be learned (Wilson, 2002). Exploration also includes cognitive components including decision-making (Markant et al., 2016), discovery (Schulz, Gopnik, & Glymour, 2007), hypothesis testing (Gopnik & Wellman, 2012), choice about what children learn (Begus et al., 2014; Lucca & Wilbourn, 2018), and control over the flow of information (Markant et al., 2016). Active exploration is also rewarding (Gruber et al., 2014). When learners explore actively, they are "in the driver's seat," which may enhance their learning and generalization more than a comparable observational experience.

Still, some observational experiences engage children mentally in ways that might allow children to *feel* as though they are in the driver's seat even when they are not acting. When observed actions were more relevant for children's actions, the mirror neuron system was more active (Meyer et al., 2022), which supported children's action performance (Meyer et al., 2011). Further, children who overclaimed someone else's actions as their own (Sommerville & Hammond, 2007) or saw a person present while learning (Howard et al., 2020) demonstrated better action learning and memory. Observing another person's exploration may allow children to be "mentally active" if children feel that they are acting or if children recognize that the observed actions are relevant for their own actions. To ensure children who only observed had access to the same information as children who explored, children in the observational condition watched a video of another child's exploration. This allowed children to receive the same information as children who explored while varying only whether children actively generated or observed that evidence.

In addition, Studies 3 and 4 did not disentangle the possible benefits of exploration from instruction. All children who received instruction also explored, either prior to instruction or during the exploratory Test phase. Results from the Exploration Alone condition in Study 3 suggested that children learned from exploration without prior instruction. However, when task demands were higher and less information could be gained by exploring alone (in-person, Radovanovic et al., in preparation), children did not learn or generalize at high rates. Thus, sufficient exploration experience could support learning and generalization (e.g., Dean & Kuhn, 2007). Study 5 therefore included a condition where children received two opportunities to actively explore (Double Explore). This condition was compared to the condition that lacked exploration, and instead provided children with an opportunity to observe (Observation Then Instruction). This allowed for a more controlled examination of the effects of exploration versus instruction, and also tested whether active or observational exploration supported learning.

The learning conditions in Studies 3 and 4 also differed in the amount of time children were provided with information about the locks and keys: Children in Exploration Then Instruction had more time with the locks and keys than children in Instruction Alone. It is possible that children in Study 3 were more successful at Test after exploring because they had been given more time with the locks and keys. To equate exposure time, all children in Study 5 had three minutes with the locks and keys. In addition, the Test phase in Studies 3 and 4 was not a stringent test of learning. The Test phase provided learners with feedback about correct and incorrect lock-key matches, allowing children to continue learning. Study 5 therefore implemented a true test phase: Children had one chance to unlock each lock and were not given feedback. After the true Test phase, children participated in the same Generalization phase as in Study 3 to test how accurately they could apply the rule they had learned to new locks and keys.

Study 5 was conducted with 6-year-olds and the methods and analyses were preregistered (https://osf.io/tvjh2/?view_only=e5f28c406a4048aba60ca658458e5a33). Study 6 was identical in design to Study 5 and was conducted with a comparison sample of adults. The methods and analyses of Study 6 were also preregistered (https://osf.io/2aw6c/?view_only=0f96b8c4e4364b7b82dba62985efbaba).

Study 5

This study tested whether timing of exploration relative to instruction affected learning or generalization with two conditions: Exploration Then Instruction and Instruction Then Exploration. We also tested whether active exploration or observational instruction supported learning and generalization with two additional conditions: Double Explore and Observation Then Instruction. All conditions equated the amount of time that children were exposed to the locks and keys and tested children's learning with a true Test phase, followed by a Generalization phase.

Methods

Participants. 160 children (mean age = 6.3 years; 74 male, 86 female) participated. Children were recruited from a database of families who volunteered to participate in research at a university in a large Midwestern city (144) and from a local school affiliated with the university (16). The age range was wider than in Study 3 (range = 5.4 - 7.5 years) so children in kindergarteners and first grade at the school could participate. Children were primarily of

European or White-American descent (87; Asian or Asian-American: 22, African or African-American: 14, Hispanic or Latino-American: 13, multiple race or ethnicities: 17, other: 5, did not report: 2) and primary caregivers were highly educated (post-graduate degree: 81, bachelor's degree: 60, some college: 9, associate's degree: 6, high school: 2, did not report: 2). Children were randomly assigned to one of four conditions with 40 children per condition: Exploration Then Instruction (mean age = 6.2 years, 18 male, 22 female), Instruction Then Exploration (mean age = 6.3 years, 19 male, 21 female), Double Explore (mean age = 6.4 years, 19 male, 21 female), or Observation Then Instruction (mean age = 6.5 years, 18 male, 22 female). Data was collected in December of 2021 and January of 2022.

Additional children (54) were tested but excluded from analyses (39 database families, 15 school families; no difference by condition: Exploration Then Instruction: 12; Instruction Then Exploration: 17; Double Explore: 10; Observation Then Instruction: 15; $\chi^2(3) = 2.15$, $p = 0.540$) following preregistered exclusion criteria. Children were excluded due to caregiver interference (18; see Post-game Survey); failure to click and drag the practice key (8), quitting or restarting the game (14), failure to see the video in the Observation Then Instruction due to internet connection issues (6), caregiver failure to complete the Post-game Survey (1), colorblindness (4), or exposure to less than 50% English (3).

Procedure. This study was conducted asynchronously without an experimenter present. Families were recruited to participate via email. The email included a link to a Qualtrics survey that guided them through the study. After providing consent online, caregivers read instructions that asked them to have their child play the Lock and Keys Task independently on a laptop or desktop computer with their computer audio turned on in a distraction-free area. Caregivers also viewed instructions about optionally recording a video of the session using Zoom. Caregivers then followed a link to the Lock and Keys Task website and entered their child's randomly-assigned Game ID, which loaded the task in the correct condition. Children then participated in the Lock and Keys Task. Finally, caregivers completed a survey about their child's game play (Post-game Survey) and provided background information (children's age, race or ethnicity, language exposure, caregiver education, and whether the child had been diagnosed with colorblindness). Families received a $5 Amazon gift card to thank them for participating, and received an additional $5 gift card if they submitted a video of the session.

Lock and Keys Task. The Lock and Keys Task was similar to that of Study 3. Children first participated in the same Practice phase. Following Practice, children received three minutes of exposure to the locks and keys that included exploring, seeing instruction, or both. Then, children participated in a Test phase (which differed from that of Study 3) and a Generalization phase (which was identical to that of Study 3; Figure 14).

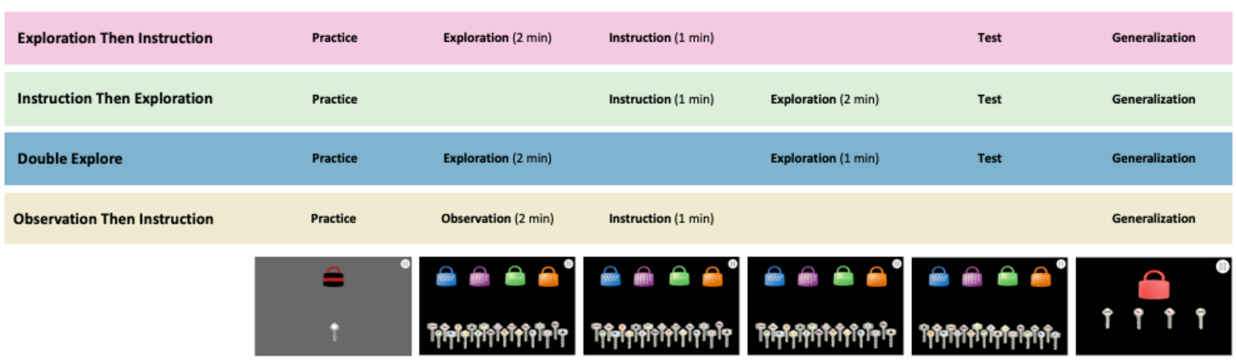


Figure 14: Study 5 Design. Children were assigned to participate in one of four conditions. Time spent in each phase is listed with the phase. An image of the screen shows how the locks and keys were arranged at the start of each phase. The image for Generalization is an example of one of the six Generalization trials.

Exploration Then Instruction and Instruction Then Exploration. In the two conditions including both exploration and instruction, children experienced two minutes of Exploration and one minute of Instruction, identical to Study 3. These conditions differed only by the order in which the Exploration and Instruction phases were presented. Exploration Then Instruction included Exploration followed by Instruction, while Instruction Then Exploration showed Instruction followed by Exploration.

Double Explore. In the condition including only exploration, children were given two minutes of Exploration identical to Study 3. Then, the keys reset: The keys slid back to their original positions at the bottom of the screen. Children were given the same prompt that began Exploration ("Use the keys [*keys highlighted*] to unlock the locks [*locks highlighted*]"). Children then had one additional minute of Exploration. Resetting the keys allowed children a second chance to unlock the locks from scratch, providing a double dose of exploration.

Observation Then Instruction. In this condition, children were randomly assigned to watch one of four two-minute video of another child's Exploration phase (Observation), then watched the same one-minute Instruction demonstration as in Study 3. The Observation videos were screen-recorded videos of children's exploration from Study 3. In all videos, children saw evidence that some keys unlocked the locks while others did not. Two of the videos provided "better than average" evidence and two provided "excellent" evidence about which keys unlocked the locks.

Data from the Exploration phase of Exploration Then Instruction and the first two minutes of the Test phase in Exploration Alone in Study 3 were analyzed to identify "better than average" and "excellent" videos (N = 50; data collection for Study 3 was partially complete at the time of this analysis). On average, when exploring, children unlocked 2 different locks. "Better than average" videos showed successful unlocking of 3 different locks, and "excellent" videos showed all 4 locks unlocking (Table 6). Four videos with these features were selected from the Exploration phase of the Exploration Then Instruction condition from children in Study

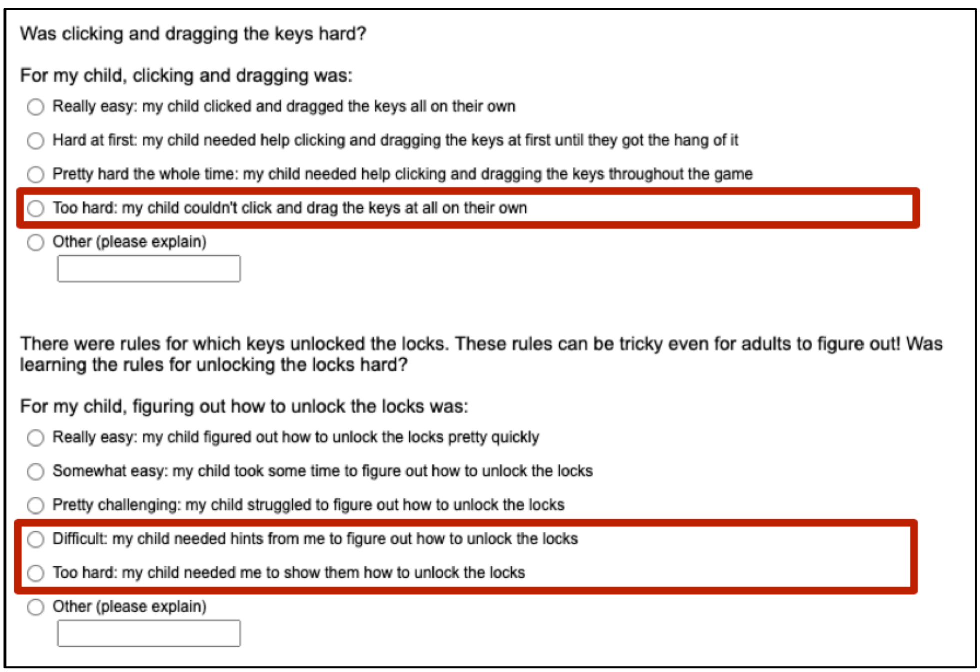
3. The audio and timing of the videos was edited. At the beginning of the video, voiceover said, "Watch the keys [*keys highlighted*] unlock the locks [*locks highlighted*]." When a key unlocked a lock, the correct unlocking sound played and voiceover said, "Look, the lock unlocked!" When an incorrect key was tested on a lock, the incorrect tapping sound played. The two "excellent" videos were also edited for timing: Freeze-frames were added between key movements to make the videos two minutes long.

		Total unlocking attempts (number of locks)	Unlocking Failures (number of locks)	Unlocking Successes (number of locks)
Average Performance		10.1 (3.8)	7.8 (3.0)	2.3 (2.1)
Better than average	Video 1	13 (4)	10 (4)	3 (3)
	Video 2	13 (4)	10 (4)	3 (3)
Excellent	Video 3	7 (4)	3 (3)	4 (4)
	Video 4	9 (4)	5 (3)	4 (4)

Table 6: Study 5 Observation Videos. Table of descriptive statistics on the average exploration performance (during the Exploration phase of Exploration Then Explore or the first two minutes of Exploration Alone) for 50 participants in Study 1. "Better than average" and "excellent" quality videos were identified from this average performance.

Test Phase. The Test phase differed from that of Study 3 to eliminate feedback and opportunities for exploration. At the start of the Test phase, children saw four keys at the top of the screen and 20 keys at the bottom of the screen, with the key positions shuffled relative to the prior phase. There were four trials, one per lock (blue, purple, green, and orange). For the first trial, voiceover asked, "Which key [*all keys highlighted*] would unlock the lock [*blue lock highlighted*]?" Children then could click and drag keys, as in Exploration. However, only the target lock (blue) highlighted when keys were hovered over it. When a key was dropped on the blue lock, the key remained on top of the lock for one second (providing no feedback), then returned to its previous position on the screen. Then, the Test phase proceeded to the next trial, during which the procedure was repeated (for the purple lock, green lock, and orange lock). All children participated in the same Generalization phase as in Study 3 after completing Test.

Post-game Survey. After children participated in the Lock and Keys Task, a link directed caregivers to complete a Post-game Survey. This survey asked about children's experience with the game to capture any help caregivers provided to children that would invalidate children's gameplay. Specifically, if caregivers performed all the clicking and dragging for their child (N = 6; data had already been excluded for other reasons) or offered hints about the unlocking rule (N = 18), their child's data was excluded (see Figure 15). Caregivers then provided information about children's backgrounds (age, race or ethnicity, language exposure, and colorblindness).


Was clicking and dragging the keys hard?

For my child, clicking and dragging was:

- Really easy: my child clicked and dragged the keys all on their own
- Hard at first: my child needed help clicking and dragging the keys at first until they got the hang of it
- Pretty hard the whole time: my child needed help clicking and dragging the keys throughout the game
- Too hard: my child couldn't click and drag the keys at all on their own
- Other (please explain)

There were rules for which keys unlocked the locks. These rules can be tricky even for adults to figure out! Was learning the rules for unlocking the locks hard?

For my child, figuring out how to unlock the locks was:

- Really easy: my child figured out how to unlock the locks pretty quickly
- Somewhat easy: my child took some time to figure out how to unlock the locks
- Pretty challenging: my child struggled to figure out how to unlock the locks
- Difficult: my child needed hints from me to figure out how to unlock the locks
- Too hard: my child needed me to show them how to unlock the locks
- Other (please explain)

Figure 15: Study 5 Post-game Survey. For the first question about clicking and dragging, children's data were excluded if their parent answered, "Too hard: my child couldn't click and drag the keys at all on their own." For the second question about caregiver hints, children's data were excluded if their parent answered "Difficult" or "Too hard" since both options indicated that parents had provided their children with information about the unlocking rule.

Analysis

Similar measures as in Study 3 were extracted from the gameplay files. Generalization Performance and Generalization Color Rule were identical to Study 3. Exploration Total Attempts, Exploration Success Proportion, and the proportion of exploration failures performed with each key type (Incorrect Color Keys Proportion, Correct Color Incorrect Shape Keys Proportion) were extracted from the following Exploration phases: two minutes of exploration in the Exploration Then Instruction and Instruction Then Exploration conditions, and one minute of exploration following exploration in Double Explore (i.e., the final minute of exploration after the keys reset). The test score differed from that of Study 3: Test Performance was the number of correctly chosen keys on the four test trials (maximum: 4). Test Color Rule was the total number of color-matched keys children selected across the four trials (the correct keys plus color-matched, incorrect shape keys; Table 7).

Phase	Measure	Definition	Mean (SD); range
Exploration*	Exploration Total Attempts	Total number of successful and failed unlocking attempts	7.81 (3.75); 2-30
	Exploration Success Proportion	Number of successful unlocking attempts divided by Total Exploration Attempts	.42 (.31); 0-1
	Incorrect Color Keys Proportion	Number of incorrect unlocking attempts made with keys that did not match the lock in color divided by Exploration Total Attempts	.27 (.35); 0-1
	Correct Color Incorrect Shape Keys Proportion	Number of incorrect unlocking attempts made with keys that did match the lock in color but were not round divided by Exploration Total Attempts	.31 (.25); 0-1
Test	Test Performance	Number of correct keys (maximum: 4)	2.66 (1.19); 0-4
	Test Color Rule	Number of color-matched keys (maximum: 4)	3.66 (.73); 0-4
Generalization	Generalization Performance	Number of correctly chosen keys across 6 trials	3.82 (1.76); 0-6
	Generalization Color Rule	Number of color-matched keys chosen across 6 trials	5.77 (.64); 2-6

Table 7: Study 5 Measures. Measure calculated from gameplay data during each phase of the study with associated definition. Descriptive statistics are provided across conditions (mean,

Table 7: Study 5 Measures (continued): standard deviation, and range). *Exploration Attempts were extracted from the 2-minute Exploration phases in the Exploration Then Instruction and Instruction Then Exploration conditions, and from the 1-minute Exploration phase following 2 minutes of exploration in the Double Explore condition.

Analysis Strategy. Four planned contrasts were preregistered to compare Test and Generalization Performance by condition. **Exploration** tested whether children who only explored would perform better than those who were only instructed: Double Explore (1) vs. Observation Then Instruction (-1). **Exploration Timing** tested whether exploring before instruction would improve performance relative to exploring after instruction: Exploration Then Instruction (1) vs. Instruction Then Exploration (-1). **Combination to Exploration** tested whether the combination of exploration and instruction would improve performance compared to a double-dose of exploring: Exploration Then Instruction (1) and Instruction Then Exploration (1) vs. Double Explore (-2). **Combination to Instruction** tested whether the combination of exploration and instruction would be more effective than only observing: Exploration Then Instruction (1) and Instruction Then Exploration (1) vs. Observation Then Instruction (-2). These contrasts were run two separate models, each predicting Test and Generalization Performance: Model 1: Exploration, Exploration Timing, and Combination to Exploration; Model 2: Exploration, Exploration Timing, and Combination to Instruction. Proportion measures were arcsine square-root transformed prior to the analyses.

Results

Age ($F(3, 156) = 2.22$, $p = .089$) and gender (male versus female: $\chi^2(3) = 0.10$, $p = 0.992$) did not differ by condition. The analyses collapsed across these participant factors. The distribution of participants who saw each Observation video did not significantly differ (video 1: 8, video 2: 9, video 3: 13, video 4: 10; $\chi^2(3) = 1.40$, $p = 0.706$). Caregivers of 71 children (44%)

submitted a video of their child's session; there was no evidence that Test or Generalization Performance differed by whether or not caregivers submitted a video (Test: $\beta < 0.01$, SE = 0.19, $p = .996$; Generalization: $\beta = 0.03$, SE = 0.28, $p = .919$).

Test Performance. On average, children were correct on 2.7 out of four Test trials. No planned contrasts significantly related to Test Performance (all p's > .227; Figure 16A). Children in all conditions performed significantly above chance in selecting the color-matched keys (Test Color Rule, chance = 50%; all p's < .0001). When this stringent chance measure was applied to Test Performance, children in all conditions performed significantly above chance (all p's < .01), suggesting that children learned which keys unlocked the locks in all conditions. However, there was no evidence that exploration, instruction, or their combination affected learning.

Generalization Performance. On average, children scored 3.8 out of six during Generalization. Interestingly, Exploration Timing significantly related to Generalization performance ($\beta = -0.60$, SE = 0.19, $p = .002$; Figure 16B): Children generalized the rule more effectively if they explored *after* instruction rather than *before* instruction. No other contrasts reached significance (all p's > .818). In addition, children in all conditions performed significantly above chance in Generalization Color Rule (chance = 50%; all p's < .0001). However, when this stringent chance level was applied to Generalization Performance, children in the Exploration Then Instruction condition did not perform significantly above chance ($t(39) = 0.87$, $p = 0.387$). Children in the other three conditions performed significantly above chance (p's < .01). This suggests that children who explored before receiving instruction did not generalize the unlocking rule, while children in the other three conditions generalized the rule.

Exploratory analyses then tested whether Generalization Performance in any condition differed from that of Instruction Then Exploration. As in the main analysis, children in the

Instruction Then Exploration condition generalized more accurately than those in Exploration Then Instruction ($\beta = -1.20$, SE = 0.38, $p = .002$). In addition, children in Instruction Then Exploration generalized marginally better than those in Double Explore ($\beta = -0.65$, SE = 0.38, $p = .093$) and Observation Then Instruction ($\beta = -0.68$, SE = 0.38, $p = .081$). Thus, children who explored after instruction generalized significantly better than those who explored before instruction, and marginally better than those who consistently explored or were instructed.

Generalization Performance: Perfect Test. To parallel analyses from Study 3, Generalization Performance was analyzed for the subsample of children who performed perfectly at Test. This left a sample of 55 children (no difference by condition: Exploration Then Instruction: 11, Instruction Then Exploration: 18, Double Explore: 14, Observe Instruct: 12; $\chi^2(3) = 2.09$, $p = 0.554$). Of note, this Test phase was a true test and was therefore more challenging than the Test phase in Study 1. Given this small sample, it is unsurprising that no contrasts reached significance (all p's > .245; Figure 16C).

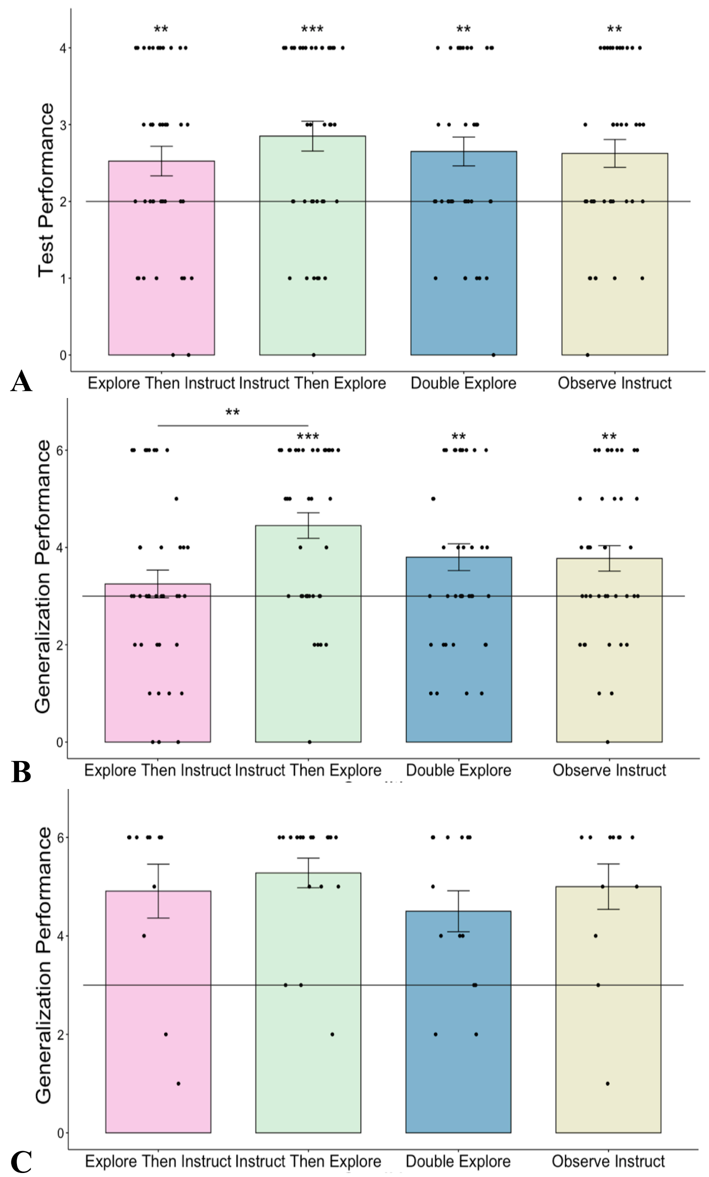


Figure 16: Study 5 Test and Generalization Performance. (A) Test Performance, (B) Generalization Performance for all participants, and (C) Generalization Performance for participants who performed perfectly at Test (perfect Test Performance). Generalization chance is 50% (3). Error bars: +/- 1 Standard Error. *$p < .05$, **$p < .01$, ***$p < .001$.

We next tested whether the difference in Generalization Performance between children who explored before or after instruction was driven by children's exploration success. Children who explored before instruction made more attempts than those who explored after instruction (Total Attempts: $\beta = 2.28$, SE = 0.75, $p = .003$); however, those who explored after instruction were more successful than those who explored before (Proportion Successful Attempts: β = -

0.14, SE = 0.04, $p < .001$; Figure 17D). In turn, higher proportions of success were related to better Generalization Performance ($\beta = 2.10$, SE = 0.49, $p < .001$), particularly for children who explored after instruction ($\beta = -1.21$, SE = 0.51, $p = .021$; Figure 17A). Further, the relation between Exploration Timing and Generalization Performance was fully mediated by Exploration Success Proportion ($\beta = -0.27$, 95% CI [-0.51, -0.08], $p < .0001$; Figure 17B). Children who explored after instruction generalized better that those who explored before instruction because they were more successful when they explored.

We also examined the nature of children's unlocking failures during exploration. After instruction, children specifically gathered information that would disambiguate the shape rule: Compared to children who explored before instruction, children who explored after instruction failed proportionally less with the Incorrect Color Keys ($\beta = -0.16$, SE = 0.04, $p < .001$) and proportionally more with the Correct Color Incorrect Shape Keys ($\beta = 0.33$, SE = 0.05, $p < .001$; Figure 17D). Children who explored without prior instruction gathered evidence to learn both parts of the rule. In contrast, instruction targeted children's exploration towards evidence that would disambiguate the shape rule, suggested that children had learned the color rule by viewing the instructions.

It is possible that instruction and exploration differentially affected children's behavior during the exploration that followed. We therefore compared attempts, successes, and failures during exploration that followed instruction (Instruction Then Exploration) to exploration that followed exploration (the second exploration phase of Double Explore). Children made more attempts after instruction than after exploration ($\beta = -1.88$, SE = 0.46, $p < .001$), but children were equally successful in their attempts ($\beta = 0.12$, SE = 0.10, $p = .215$; Figure 17D). In turn, greater success was related to better Generalization Performance ($\beta = 1.72$, SE = 0.41, $p < .001$).

Interestingly, the relation between success and Generalization was stronger for children who explored after instruction than those who explored after exploration ($\beta = -1.83$, $SE = 0.86$, $p = .038$; Figure 17C). Children therefore reached similar levels of success following instruction and exploration, but success was more relevant for generalization after instruction than after exploration.

We also examined children's failed attempts to unlock locks after instruction versus after exploration. Children who explored after instruction (rather than after exploration) were less likely to try Incorrect Color Keys ($\beta = 0.32$, $SE = 0.09$, $p < .001$) and more likely to try Correct Color Incorrect Shape Keys ($\beta = -0.43$, $SE = 0.07$, $p < .001$; Figure 17D). Instruction was therefore more effective than exploration in targeting children's behavior towards color-matched keys, suggesting that children had inferred that color mattered for unlocking more effectively from instruction than prior exploration.

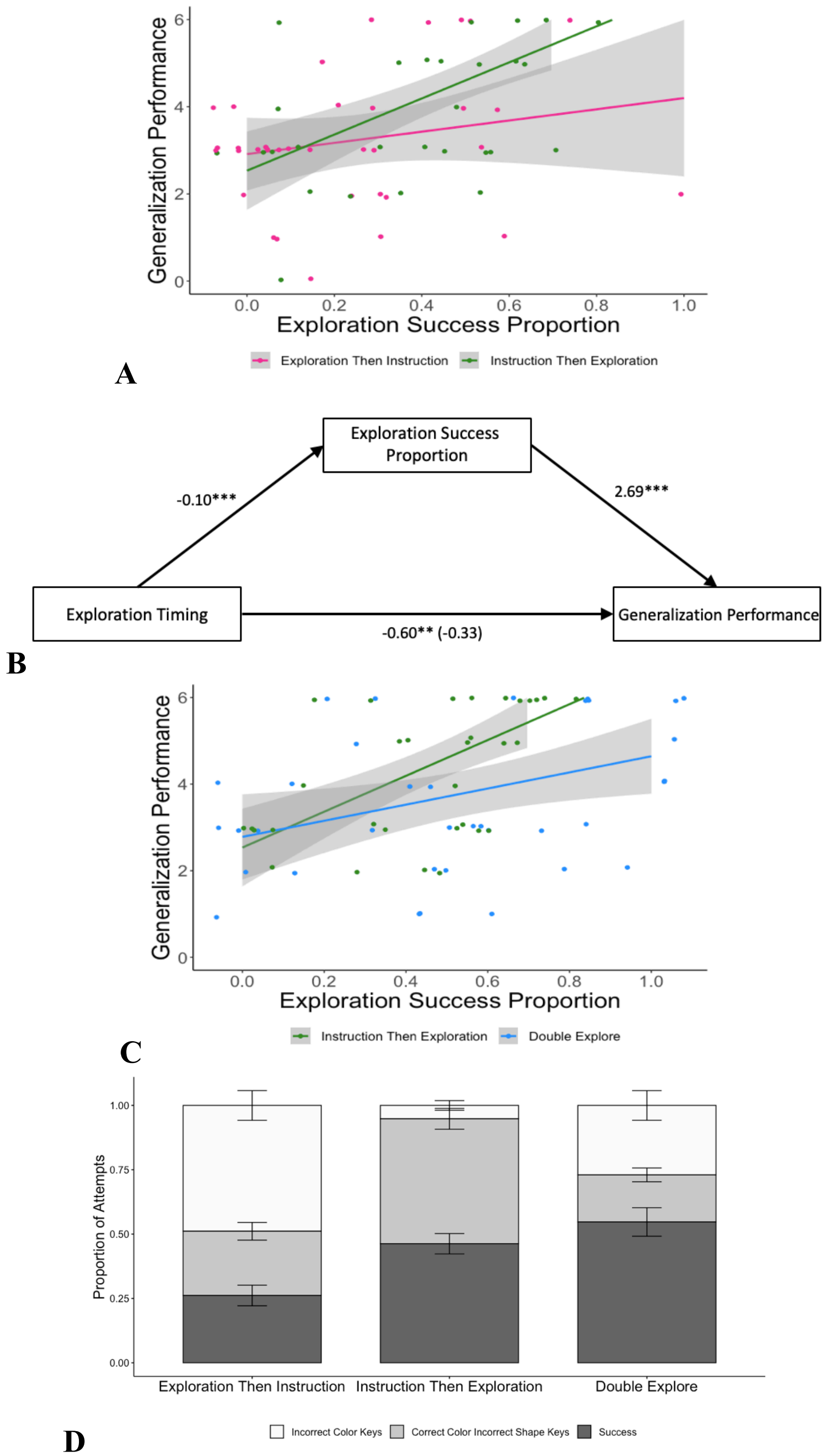


Figure 17: Study 5 Exploration Success, Generalization, and Attempts. (A) For children in the Exploration Then Instruction and Instruction Then Exploration conditions: relation between Exploration Success Proportion and Generalization Performance by condition, and (B) causal mediation model showing the effect of Exploration Timing on Generalization Performance via Exploration Success Proportion. (C) For children in the Instruction Then Exploration and Double Explore conditions: relation between Exploration Success Proportion and Generalization

Figure 17: Study 5 Exploration Success, Generalization, and Attempts (continued): Performance by condition. (D) Proportion of unlocking attempts that were successful and failed (by failure type: Incorrect Color Keys and Correct Color Incorrect Shape Keys). Shaded areas represent 95% confidence intervals. Error bars: +/- 1 Standard Error.*$p < .05$, **$p < .01$, ***$p < .001$.

Results Summary. While children in all conditions learned the taught information equally well, children who explored *after* rather than before receiving instruction generalized the rule more accurately. Instruction improved children's success in unlocking the locks, which in turn improved generalization. As in Chapter 3, guided exploration supported generalization. Children who explored twice or were instructed twice generalized at levels between instructed and uninstructed exploration. Interestingly, though prior instruction and exploration improved children's success rates similarly, instructed exploration success was more strongly related to generalization than exploration-drive success. In addition, children inferred that color was important for unlocking more effectively from instruction than from exploration: After instruction (compared to after exploration), children engaged in more targeted trial-and-error testing that would reveal evidence about the shape part of the unlocking rule.

Study 6

Study 6 was conducted with a comparison sample of adult participants on MTurk using the same methods as Study 5. Adults participated in one of four conditions: Exploration Then Instruction, Instruction Then Exploration, Double Explore, and Observation Then Instruction.

Methods

Participants. 426 participants were included in the analyses (mean age = 37.9 years, range = 20-74 years, 248 male, 117 female, 1 NA or other) with approximately 100 per condition (Exploration Then Instruction: 107; average age = 37.6 years, 57 male, 50 female; Instruction Then Exploration: 94; average age = 39.5 years, 55 male, 39 female; Double Explore: 110;

average age = 37.6 years; 69 male, 40 female, 1 NA or other; Observation Then Instruction: 115; average age = 37.2 years, 67 male, 47 female). The majority of participates were European or White-American (282; 46 Native American, 31 Hispanic or Latino-American, 29 African or African-American, 24 Asian or Asian-American, 14 multiple races or ethnicities) and held a bachelor's degree (247; 62 post-graduate degree, 62 some college, 40 high school, 14 associate's degree, 1 other). Data was collected in December of 2021.

Data from 135 additional participants were excluded (colorblindness: 119, not fluent in English: 16). 67 additional participants in the Observation Then Instruction condition were excluded because the Observation video failed to load due to poor internet connection. Extra participants were run in this condition to account for dropout. During the analysis, an additional 29 participants were excluded for contributing data that were outliers (as preregistered) and one additional participant was excluded due to a technical error.

Procedure. As in Study 4, eligible participants were recruited through MTurk and directed to a Qualtrics survey. Participants were told to use a laptop or desktop computer, turn on the computer audio, and play the Lock and Keys Task on a different website. After participating, adults entered a code into MTurk and the Qualtrics survey and were asked the same background questions as in Study 2. Participants received $1.50 as compensation. The study took approximately 10 minutes.

Analysis

The same measures as in Study 5 with children were extracted from adults' gameplay data: Test Performance, Test Color Rule, Generalization Performance, and Generalization Color Rule. During the exploration phases of Exploration Then Instruction, Instruction Then Exploration, and the second Exploration phase of Double Explore, Total Attempts were extracted and Proportion Unlocking Success, Incorrect Color Keys Proportion, and Correct Color Incorrect Shape Keys Proportion were calculated (Table 8).

Outlier Detection. As in Study 4, adults' game play was more variable that that of children in Study 5. As preregistered, behavioral data were extracted and analyzed for the presence of outliers: Total Time, Pause Time, time during each phase of the game, total keys clicked and dragged during each phase of the game, and unlocking failures. 29 additional participants were excluded for providing data that was 3 standard deviations from the mean of Total Game Play Time (8), Total Pause Time (9), Test Time (3), Generalization Time (2), and number of keys selected during Test (visual inspection of box plots: 7).

Phase	Measure	Definition	Mean (SD); range
Exploration*	Exploration Total Attempts	Total number of successful and failed unlocking attempts	9.19 (7.22); 0-40
	Exploration Success Proportion	Number of successful unlocking attempts divided by Exploration Total Attempts	.60 (.30); 0-1
	Incorrect Color Keys Proportion	Number of incorrect unlocking attempts made with keys that did not match the lock in color divided by Exploration Total Attempts	.21 (.32); 0-1
	Correct Color Incorrect Shape Keys Proportion	Number of incorrect unlocking attempts made with keys that did match the lock in color but were not round divided by Exploration Total Attempts	.21 (.18); 0-1
Test	Test Performance	Number of correct keys (maximum: 4)	2.68 (1.29); 0-4
	Test Color Rule	Number of color-matched keys (maximum: 4)	3.66 (.78); 0-4
Generalization	Generalization Performance	Number of correct keys (maximum: 6)	4.17 (1.75); 0-6
	Generalization Color Rule	Number of color-matched keys (maximum: 6)	5.76 (.64); 0-6

Table 8: Study 6 Measures. Measure calculated from gameplay data during each phase of the study with associated definition. Descriptive statistics are provided across conditions (mean, standard deviation, and range). *Exploration Attempts were extracted from the 2-minute Exploration phases in the Exploration Then Instruction and Instruction Then Exploration conditions, and from the 1-minute Exploration phase following 2 minutes of exploration in the Double Explore condition.

Analysis Strategy. As preregistered, the main analyses tested the same planned contrasts as in Study 5 predicting Test Performance and Generalization Performance: **Exploration**, **Combination to Exploration**, **Combination to Instruction**, and **Exploration Timing**. Two models were run to test these contrasts (Model 1: Exploration, Combination to Exploration, and Timing; Model 2: Exploration, Combination to Instruction, and Timing). All proportion measures were arcsine square-root transformed prior to the analyses.

Results

Age ($F(3, 421) = 0.84$, $p = .473$; 1 participant missing) and gender (male versus female: $\chi^2(3) = 2.24$, $p = 0.52$; 1 participant missing) did not differ by condition. The analyses collapsed across participant factors. To test whether there had been differential dropout from the Observation Then Instruction condition relative to the other conditions due to internet failure, participants' background characteristics were tested by condition. Participants did not significantly differ by race or ethnicity (White vs. non-White participants: $\chi^2(3) = 6.98$, $p = 0.072$) or degree attainment (bachelor's degree or higher versus less than a bachelor's degree: $\chi^2(3) = 6.37$, $p = .095$). There was also no evidence of a difference in the number of participants who saw each of the four Observation videos (video 1: 34, video 2: 24, video 3: 29, video 4: 28; $\chi^2(3) = 1.77$, $p = 0.623$).

Test Performance. On average, participants scored 2.7 out of four at Test. No contrasts reached significance in predicting Test Score (all p's > .100). In addition, participants were significantly above chance (Test Color Rule: chance = 50%; all p's < .0001; 50% chance was applied to Test Performance: all conditions were significantly above chance, all p's < .0001; Figure 18A). Like children, adults learned which keys unlocked the locks in all conditions.

Generalization Performance. Participants scored on average 4.2 out of six in Generalization. Unlike with children, no contrasts reached significance in predicting Generalization Performance (p's > .802). Participants in all conditions performed significantly above chance as well (Generalization Color Rule: chance = 50%; all p's < .0001; 50% chance was applied to Generalization Performance: all conditions were above chance, all p's < .0001; Figure 18B). Adults generalized the rule at above-chance levels in all conditions, and prior exploratory or instructional experience did not differentially affect generalization.

Generalization Performance: Perfect Test. As in prior analyses, Generalization Performance was analyzed for the group of participants who scored perfectly at Test. 162 participants (38%), were included in this analysis (no difference by condition: Exploration Then Instruction: 42, Instruction Then Exploration: 35, Double Exploration: 42, Observation Then Instruction: 43; $\chi^2(3) = 1.01$, $p = .798$). Within this subsample of participants who demonstrated perfect rule-learning, participants in all conditions performed at above-chance levels (chance = 50%; all p's < .0001). Interestingly, there was a significant effect of Combination to Explore (Model 1: $\beta = -0.17$, SE = 0.09, $p = .047$) and a significant effect of Combination to Instruction (Model 2: $\beta = -0.17$, SE = 0.09, $p = .047$; no other contrasts reached significance: p's > .145; Figure 18C). Those in the Double Explore condition demonstrated significantly better Generalization Performance compared to those in either condition that combined exploration and instruction. Similarly, those in the Observation Then Instruction condition scored higher in Generalization compared to those in the combination conditions. Among adults who demonstrated perfect test performance, those who experienced consistent exploration or instruction generalized the rule most accurately.

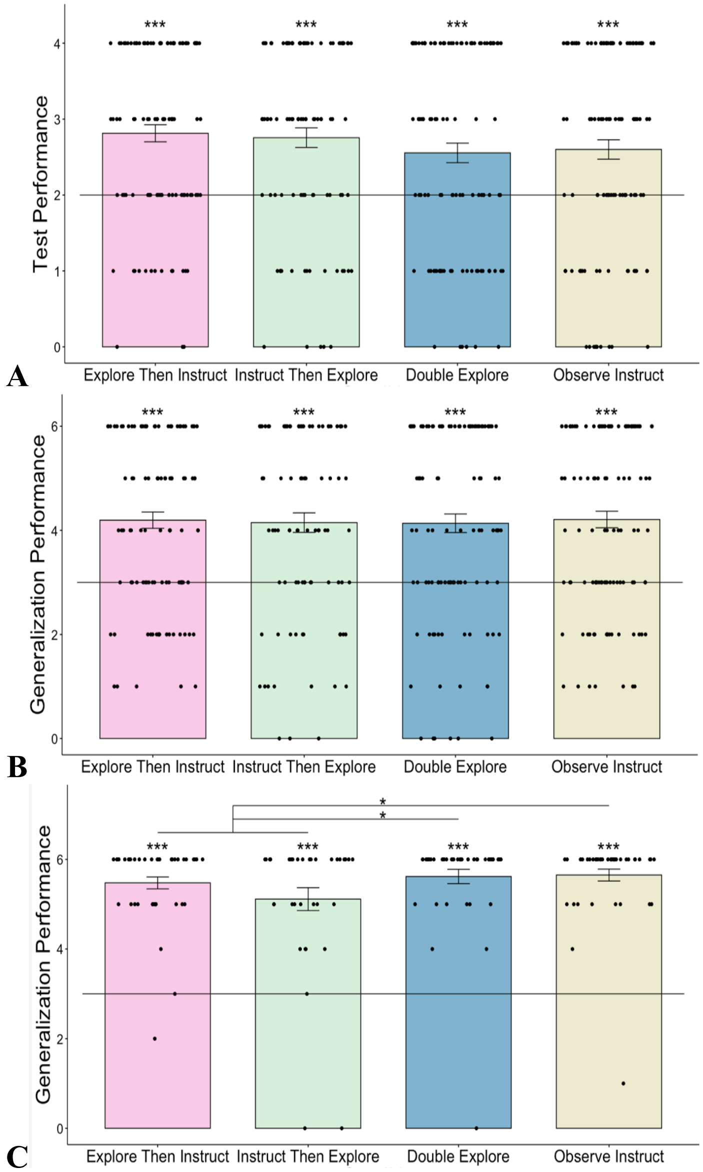


Figure 18: Study 6 Test and Generalization Performance. (A) Test Performance, (B) Generalization Performance for the full sample, and (C) Generalization Performance for those who scored perfectly at Test (perfect Test Performance). Generalization chance is 50% (3). Error bars: +/- 1 Standard Error. *$p < .05$, **$p < .01$, ***$p < .001$.

Since adults generalized better after consistent exploration than combined exploration and instruction, we compared exploration attempts, successes, and failures between conditions. Among those who scored perfectly at Test, adults made more attempts to unlock the locks when they explored before instruction compared to after exploration ($\beta = 5.02$, SE = 1.58, $p < .001$), but there was no difference in the number of attempts between participants who explored after instruction or exploration ($\beta = 1.58$, SE = 1.21, $p = .196$). Interestingly, those who explored after exploration were the most successful (Exploration Then Instruction: $\beta = -0.36$, SE = 0.08, $p < .001$; Instruction Then Exploration: $\beta = -0.20$, SE = 0.09, $p = .024$; Figure 19B). However, exploration success was not significantly related to Generalization Performance ($\beta = 0.16$, SE = 0.27, $p = .557$; Figure 19A). The lack of a relation between exploration success and generalization was likely due to limited variability in adults' generalization; in this sample of high-performing participants, most adults scored perfectly during Generalization.

We also examined the failures adults made as they explored. Exploration targted adults' behaviors most: After exploration, adults tried proportionally fewer Incorrect Color Keys (Exploration Then Instruction: $\beta = 0.29$, SE = 0.07, $p < .001$; no difference from Instruction Then Exploration: $p > .731$) and fewer Correct Color Incorrect Shape Keys (Exploration Then Instruction: $\beta = 0.15$, SE = 0.06, $p = .014$; Instruction Then Exploration: $\beta = 0.18$, SE = 0.06, $p = .005$; Figure 19B). Two exploration experiences allowed adults to intuit both parts of the unlocking rule more effectively than exploring before or after instruction.

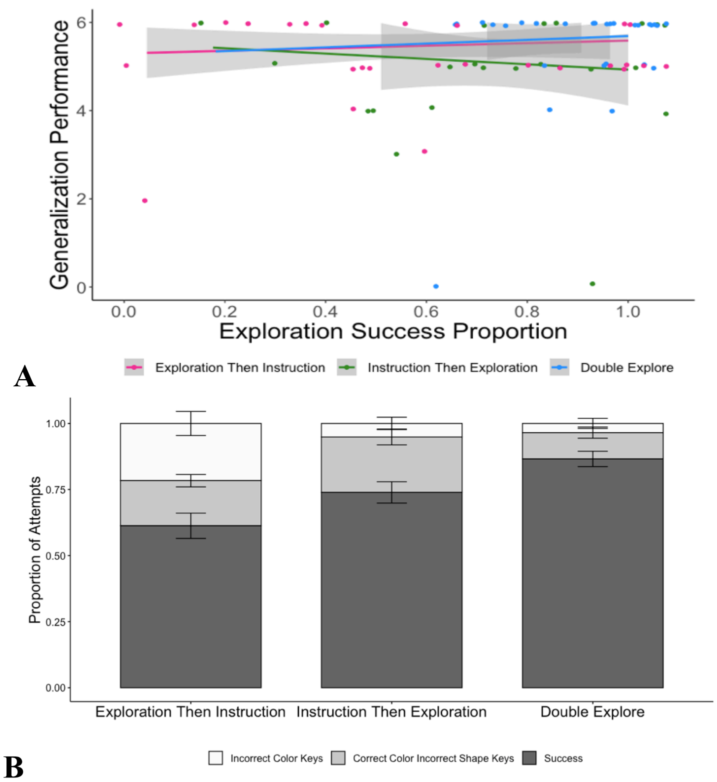


Figure 19: Study 6 Exploration Success and Attempts (Perfect Test Only). Among participants who scored perfectly at Test: (A) Relation between Exploration Success Proportion and Generalization Performance separated by condition and (B) proportion of unlocking attempts that were successful and failed (by failure type: Incorrect Color Keys and Correct Color Incorrect Shape Keys). Shaded areas represent 95% confidence intervals. Error bars: +/- 1 Standard Error. *$p < .05$, **$p < .01$, ***$p < .001$.

We additionally examined attempts, successes, and failures within the full sample of participants. Adults made more attempts when exploring before instruction than after exploration ($\beta = 6.32$, SE = 0.92, $p < .001$), but attempts were similar when participants explored after instruction or exploration ($\beta = 0.65$, SE = 0.94, $p = .489$). Adults were more successful after exploring than before instruction ($\beta = -0.28$, SE = 0.06, $p < .001$) but were similarly successful after instruction and exploration ($\beta = -0.042$, SE = 0.06, $p = .483$; Figure 20B). In addition, greater exploration success was positively related to Generalization ($\beta = 1.43$, SE = 0.23, $p < .001$; no interactions with Combination To Explore or Explore Time: p's > .684; Figure 20A).

Exploring after exploration targeted adults' behaviors compared to those who explored first (Exploration Then Instruction): After exploration, adults used proportionally fewer Incorrect Color Keys (β = 0.29, SE = 0.06, $p < .001$; no difference in Correct Color Incorrect Shape Keys: $p > .362$). Interestingly, instruction targeted adults' behavior more than prior exploration: After instruction, adults tried proportionally fewer Incorrect Color Keys (marginal: β = -0.12, SE = 0.06, $p = .059$) and more Correct Color Incorrect Shape Keys (β = 0.16, SE = 0.04, $p < .001$; Figure 20B) than those who explored after prior exploration. In the full sample of adults, like children, instruction targeted exploration more effectively than prior exploration: Adults inferred the color part of the rule more effectively from instruction than their own exploration.

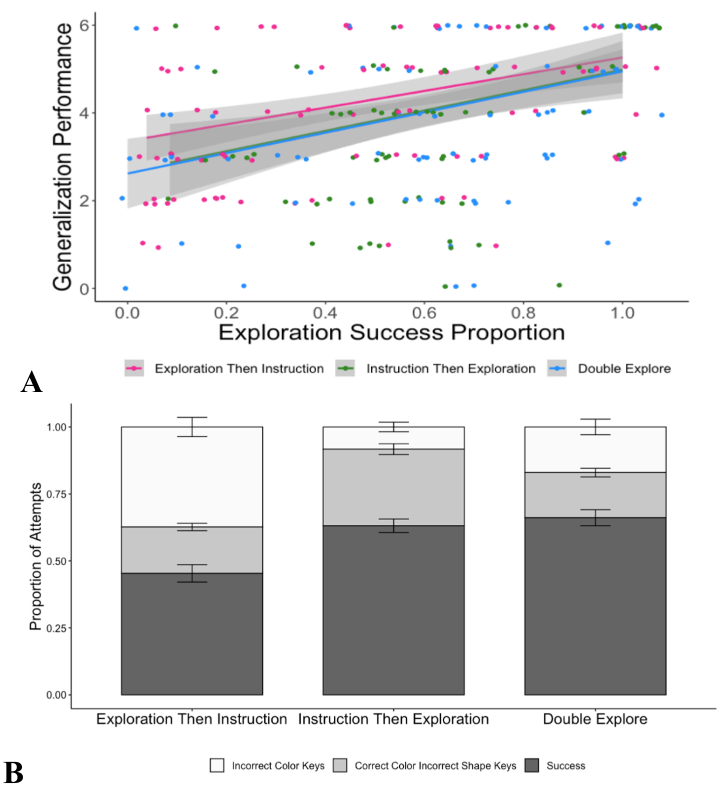


Figure 20: Study 6 Exploration Success and Attempts (Full Sample). (A) Relation between Exploration Success Proportion and Generalization Performance and (B) proportion of unlocking attempts that were successful and failed (by failure type: Incorrect Color Keys and Correct Color Incorrect Shape Keys). Shaded areas represent 95% confidence intervals. Error bars: +/- 1 Standard Error. *$p < .05$, **$p < .01$, ***$p < .001$.

Results Summary. Among the participants who scored perfectly at Test, adults generalized the unlocking rule better after consistent exploratory or instructional experience rather than combined exploratory and instructed experiences. Adults who explored after prior exploration were more successful in unlocking locks than those who explored before or after receiving instruction. Within this high-performing subgroup, exploration guided adults' learning; after exploring, adults were highly successful and rarely tried keys that were the wrong color or shape. Skilled adults learned well from exploring because their successful exploration allowed them to intuit both the color and shape parts of the rule. In addition, among adults who scored perfectly at test, consistent observational experience was as effective as consistent exploratory experience.

However, among adults with more varied learning outcomes (i.e., the full sample), generalization did not differ by condition. Adults achieved similar exploration success after instruction and after exploration. In turn, more successful exploration supported generalization across conditions. Interestingly, instruction targeted adults' exploration more effectively than prior exploration. After instruction, adults focused their attempts on keys that were color-matched, indicated that instruction allowed adults to infer the color part of the rule. However, after adults explored, they tested keys that were the wrong color, suggesting that instruction allowed adults to infer the color part of the rule more effectively than their own exploration. Among adults of different skill levels, instructed exploration was more effective than consistent exploratory experience for rule learning.

Discussion

Four preregistered studies tested whether exploration, instruction, or the combination of exploration and instruction improved children's and adult's rule-learning and generalization in an online problem-solving task. Across the studies, learners tended to generalize best when they explored after receiving instruction. Instruction targeted subsequent exploratory behaviors in ways that allowed learners to disambiguate the rule governing the problem-solving system. In turn, successful exploration supported generalization. Success achieved through guided exploration was key for promoting generalization.

The most robust evidence of this effect was seen in Study 5. Children who explored after rather than before instruction were more successful in their exploration, which in turn improved children's generalization. The results from Study 3 showed a similar effect: After instruction, whether or not children had also explored before receiving instruction, children were more successful in the exploratory Test phase. In turn, successful exploration improved generalization. and these results were replicated in Study 4 with adults.

Did exploration success reflect what learners had already learned or did children and adults learn more as they explored? Evidence that children and adults continued learning during exploration comes from three sources: (1) Generalization performance across studies, (2) the relations between exploration success and learning in conditions where learners explored first, and (3) failed unlocking attempts when learners explored after instruction.

Comparing children's generalization performance in the Exploration Then Instruction conditions in Study 3 and Study 5 indicated that children generalized better when they had more time to explore. In Study 3, children in the Exploration Then Instruction condition generalized at above-chance levels. Children in this condition explored, received instruction, and then explored again during the Test phase. In contrast, in Study 5, the Exploration Then Instruction condition did not allow children an extra opportunity to explore after instruction (the test phase was a true test without feedback); children in this condition did not generalize at above-chance levels. This suggests that when children had an extra chance to explore after receiving instruction in Study 3, they generalized more accurately than children who did not explore after instruction in Study 5. We can therefore conclude that additional learning occurred during the exploratory Test phase in Study 3. However, it is possible that any additional experience would have supported generalization. This possibility was addressed in Studies 5 and 6.

In conditions where learners explored first, there was also evidence that they learned by exploring. In the Exploration Then Instruction condition in Study 3, children who explored more successfully demonstrated greater learning at the Test phase. In addition, adults in Study 6 who were perfectly accurate at Test were highly successful when they explored after prior exploration, suggesting that they had learned both the color and shape components of the rule through their first exploration phase. In the conditions where learners explored first, exploration improved children's learning and adults' generalization.

Additionally, the types of errors learners made while exploring in the exploratory test phase (Studies 3 and 4) and the exploration phases (Studies 5 and 6) demonstrated that children and adults systematically tested specific keys while they explored. Though instruction led to the most successful exploration, learners were not completely successful; specifically, after instruction, children and adults tried correct keys and color-matched incorrect shape keys. In contrast, learners who explored without prior instruction tried more keys that did not match the locks in color. This suggested that without instruction, learners used trial and error to discover both components of the unlocking rule. However, after instruction, learners focused their efforts on testing the keys that matched the locks in color. Learners had therefore intuited that color mattered from instruction but were still learning the shape rule through systematic exploration.

These errors indicated that children and adults were engaging in hypothesis testing as they explored (Gopnik & Wellman, 2012), which likely supported generalization. Children and adults who explored first were unaware of the rule: Learners tested all types of keys, suggesting that they began the study with a simple hypothesis: "some keys unlock the locks." After receiving instruction, learners' hypothesis space was reduced. Instead of testing all the keys, children and adults focused their testing behaviors on keys that matched the locks in color (correct keys and correct color, incorrect shape keys). This suggests that instruction constrained learners' hypothesis spaces to: "color-matched keys unlock the locks." However, the instructions did not provide counterfactual evidence that some color-matched keys did not unlock the locks. Children and adults systematically tested color-matched keys that were round, which generated successes identical to the instructions, and color matched keys that were not round, which generated the counterfactual evidence necessary to learn both components of the unlocking rule. Children in Study 3 who explored and received instruction and adults who explored twice in Study 6 were the most targeted in their testing behaviors, rarely trying keys that were incorrect. This suggested that through exploration, instruction, and more exploration (children, Study 3) and successful exploration alone (among high-performing adults, Study 6), learners had developed an even more refined hypotheses: "color-matched round keys unlock the locks."

Unlike prior research on the benefits of productive failure (e.g., DeCaro & Rittle-Johnson, 2012), learners who failed *less* often more accurately applied the rule to new locks and keys. More confirming rather than disconfirming evidence supported generalization. Success was measured as a proportion of attempts; thus, the positive effect of success on generalization also necessarily reflects a negative effect of failure on generalization. As such, learners who succeeded more and failed less generalized more accurately. It is possible that enacting the material to be learned allowed learners to generalize the rule.

Instruction was crucial in guiding learners' behaviors; instruction improved children's and adults' exploration success. After instruction, learners explored during what can be considered a "practice" phase: Learners saw a demonstration of the material to be learned, then practiced enacting unlocking independently. Children and adults were successful after viewing the instruction likely because instruction reduced the degrees of freedom in the problem, helping learners narrow in on relevant information (i.e., the color rule) and ignore irrelevant information (i.e., keys that did not match the locks in color). Guided discovery learning has been found to improve students' learning in classroom settings compared to fully guided or completely unguided exploration (Alfieri et al., 2011). We found similar evidence on a smaller scale than classroom lessons; exploration guided by instruction improved learning transfer.

Instruction can also help students develop appropriate problem-solving strategies, which children might not generate independently (Klahr & Nigam, 2004; Mayer, 2004). A similar process may have occurred in the present studies. The Lock and Keys Task was highly constrained; the only possible actions were dragging and dropping keys onto locks. Even without instruction, learners engaged in a basic problem-solving strategy of testing keys. However, instruction guided learners to engage in specific forms of testing. When children explored without prior guidance, they used some correct keys, but often tried keys that did not match the locks in color. Receiving instruction allowed children to develop a more targeted problem-solving strategy focused on learning which of the color-matched keys unlocked the locks. Instruction targeted children's existing problem-solving strategies towards actions that would disambiguate the causal rule.

Although children generalized significantly (Study 3) or marginally (Study 5) more accurately from instructed exploration than exploration alone, children who only explored also generalized the rule at above-chance levels. It is possible that if children were given additional time to explore, they may have generalized at levels similar to those who received instructed exploration. While these findings do not allow definitive conclusions, they suggest that given sufficient time and multiple opportunities to explore "from scratch," children may be able to explore in ways that supported high levels of generalization. In these studies, when exploration time was limited, instructed exploration supported rule-learning.

More advanced learners learned well from unguided exploration. Specifically, for the high-performing adults in Study 6, two exploratory experiences supported generalization better than combined exploratory and instructed experiences. This was likely due to the fact that adults' exploration was highly successful; on average, adults' success rate in the second exploration phase of Double Explore was 87% (SD = 18%), while children's success rate was on average 55% (SD = 35%). Since adults generated more successes than children, this likely caused adults to generalize more accurately from two exploratory experiences. In contast, children and the whole sample of adults benefitted most from guided exploration. Children and the entire adult sample were equally successful after exploring and after instruction; the high-performing subsample of adults were more successful after exploring than after instruction. More mature or advanced learners may therefore learn from their own actions, while less advanced learners may benefit from instructed exploration.

Learning from Observing

Interestingly, children who were instructed twice also generalized at above-chance levels, though they performed marginally poorer than children who experienced guided exploration.

This was surprising given the potential benefits of being "in the driver's seat" while learning. Though children in this condition could not make decisions, control the flow of their learning, or test hypotheses (Markant et al., 2016), it is possible that children felt they were acting even when they were only watching. All children participated in a practice phase before observing, which made the actions they observed highly relevant for their own actions (Meyer et al., 2011). In addition, the practice phase may have induced a collaborative context (Sommerville & Hammond, 2007), which may have caused children to overclaim observed actions as their own. It is possible that this observational context induced mental activity that improved children's ability to learn from watching.

Even if children were aware that they were watching rather than acting, children are skilled in learning from observation. Children learn through viewing examples (Bandura, 1977) that they can imitate (Tomasello, 2001), particularly when learning conventional information (Bandura, 1993). Given the breadth of research on children's robust ability to learn from watching (e.g., Rogoff, 1990; Bauer, 1996), it is unsurprising that children could learn from viewing videos of other children's exploration. The high-performing subsample of adults also generalized best from consistent observational (and consistent exploratory) activity. Children and adults could have learned well by watching because the videos showed some of the most successful examples of children's exploration. We chose videos that were better than children's average exploration; learners saw three or four of the locks being unlocked successfully, and also saw crucial counterfactual evidence that some keys did not unlock locks. Watching highly successful exploration may be just as effective as exploring independently, particularly when learners generated less successful exploration independently than the evidence they observed. Still, informed exploration improved children's generalization marginally more than observing, which provides some evidence that guided active exploration may support generalization more than observation.

With the current study design, it is challenging to discern whether active or observational exploration was more effective for generalization. To test this question, a new condition could be run in place of Observation Then Instruction that would more closely match the Instruction the Exploration condition: Instruction Then Observation. Comparing generalization from these matched conditions would directly test whether active or observational exploration that followed the instruction demonstration was more effective. It is possible that enacting the material to be learned after instruction would support generalization more than observing another child's exploration after instruction. This direct comparison of instructed active versus observational exploration is an area for continued research.

Learning Outcomes

All children learned which keys unlocked the locks, performing at above-chance levels in all conditions in Study 5 and at high levels of success in Study 3 (though exploration prior to instruction improved learning). Combining exploration and instruction, exploring alone, and receiving instruction alone communicated the relevant information about the material to be learned. Through acting, observing, and the combination of acting and observing, all children successfully matched the keys that unlocked each lock. It is possible that a more challenging test phase could have shown benefits of guided active experience for learning.

It would have been interesting to examine whether children could maintain the unlocking rule in long-term memory months after learning. While two exploratory experiences and two observational experiences yielded similar generalization, active exploration may have improved children's memory compared to observational experience. Indeed, we tested children's memory for the color and shape rule three years after children participated in the in-person study (Radovanovic et al., in preparation). However, no differences in memory for the rule were seen between the three conditions. Three years was likely too long an interval over which to measure memory; testing children's memory a few months after participating in the online Locks and Keys Task could be an interesting avenue for future research.

Limitations & Future Directions

Open questions remained from Chapter 3, which were addressed in Chapter 4. However, additional questions remain from the findings of Studies 5 and 6. One area of future research is to examine whether exploration could be guided in ways that would support generalization. Prior research has shown that when children were provided with subtle hints about patterns in the materials, they learned more about mathematics (one-to-one correspondence) by exploring independently (Mix et al., 2011). In the Lock and Keys Task, the keys were arranged in a pseudo-randomized order on the screen, but the paired color-matched keys could have been sorted or highlighted to draw attention to the shape rule. Exploration could also be set up in ways that limited the degrees of freedom in the problem. For example, children could see one lock at a time with a limited set of keys to learn the rule in a more step-wise way. These manipulations could help children learn more quickly through exploration and potentially experience less failure and frustration. Guided exploration could even be more effective in improving generalization than instructed exploration.

The instruction demonstration could also be manipulated in different ways to enhance learning. Here, we simply showed each correct key unlocking its corresponding lock accompanied by simple narration. However, in the prior in-person study (Radovanovic et al., in preparation), the experimenter provided gestural cues indicating the color and shape rule. We ran a pilot study of adults who were tested identically to Study 4, but the Instructions included cues that highlighted the color and shape rule. Adults who saw instructions including cues generalized the rule better than those in Study 4 who did not see the cues. Future studies could manipulate instruction cues to make the relevant information to be learned more or less apparent and measure the effects of cues on learning and generalization.

The types of observational experiences children receive could also be addressed through future research. It is possible that the Observation videos tested here supported learning because the videos showed better-than-average exploration; we tested whether an *accurate* and *informative* observational experience would be as effective as active experience in supporting generalization. Ideally, a yoked design would be used where each child in the Observation condition would watch another child's exploration. This would better test whether exploration is effective only when it is an active process or whether it could also be effective as an observational experience. In addition, children may learn well from certain types of observed videos. For example, videos that show clear hypothesis testing could have been more effective for learning than watching a child explore who was less organized in their behavior.

Conclusion

In sum, exploration and instruction together promoted rule-learning in 6-year-olds and comparison samples of adults. Across four preregistered studies, learners' success during exploration promoted rule generalization. Exploration success was most commonly supported by instruction that targeted learners' exploration towards relevant information. Learners engaged in systematic hypothesis-testing to learn through instructed exploration. In turn, success during exploration facilitated the transfer of learned material to new contexts. Exploration guided to success by instruction was key for supporting children's and adults' rule generalization.

Chapter 5: General Discussion

This dissertation addressed the research question: Does active experience support children's learning in instructed contexts? Across six preregistered studies presented in this dissertation, the results demonstrated that active experience supported learning when children's activities were concurrently guided by a teacher or when active experience followed instruction. Guided active experience supported young children's learning of conventional actions and rule-based systems. However, the ways that active experience, instruction, and learning were defined affected the answers to this question.

In Chapter 2, toddlers were physically active during instruction. Children controlled their actions by engaging in motor planning and action execution, but the order and timing of their actions were guided by an experimenter, which limited children's agency. The experimenter also guided children to perform each action with high levels of accuracy, which did not allow children to make errors. Physical-only active experience was compared with observational experience, where children viewed the actions as the experimenter performed them. The results from Study 1 showed that toddlers learned and generalized the taught actions from active and observational instruction. But, in Study 2, children demonstrated better long-term memory for the toys they had assembled through active experience compared to observational experience. These results were particularly impressive since children's memory was tested after a considerable delay of one year.

In Chapters 3 and 4, 6-year-old children were not only physical agents in a problem-solving task; children also had decision-making power over their own actions. In Studies 3 and 5, children had opportunities to actively explore an online problem-solving task before or after receiving instruction; i.e., children were not guided by instruction while they explored. Within an online environment where only certain actions were possible, children could explore to perform actions in any sequence that they decided, allowing them to test hypotheses, make discoveries, encounter failure, and use trial-and-error to learn the rules governing the system. In Study 3, children learned the taught information when they had more time to explore independently. In Studies 3 and 5, children who received instruction prior to exploring were the most successful in their independent actions, which in turn supported generalization. Similar results were found in comparison samples of adult in Studies 4 and 6.

In addition to manipulating the types of active engagement that children experienced and the timing of activity relative to instruction, this dissertation tested different forms of learning. Chapter 2 measured children's immediate action learning, generalization, and long-term memory for taught information. Chapters 3 and 4 measured children's and adults' rule-learning and generalization. Interestingly, children learned and generalized taught actions equivalently from active and observational instruction in Chapter 2, and children learned the unlocking rule equivalently from nearly all active, instructed, and combined experiences in Chapters 3 and 4. Active experience specifically enhanced long-term memory in Study 2, and guided active experience improved generalization in Chapters 3 and 4.

Physical Activity

In Chapter 2, the mechanism supporting children's learning was physical activity. Children who performed the actions to be learned had better long-term memory for the structures they had built compared to the structures they saw the experimenter build. The physical component of active experience included sensorimotor integration (Rochat, 1989), action planning (McCarty et al., 1999; Barrett et al., 2007), and attention to performed actions (Amso & Scerif, 2015). These components contributed to an embodied learning experience (Wilson, 2002) that improved information encoding for storage in long-term memory (Kinder & Buss, 2020). Importantly, though children did not have agency over the actions they performed, they still planned and executed each action independently, which improved children's memory.

In Chapters 3 and 4, children were active agents of their own learning. In that way, children experienced embodied learning (Wilson, 2002), but they did not engage with physical materials in the same way as toddlers in Chapter 2 or 6-year-old children in the prior in-person version of the problem-solving task (Radovanovic et al., in preparation). Physical objects may help or hinder learning depending on the information to be learned and the maturity level of the learner. When children learn abstract concepts such as mathematics, physical manipulatives can sometimes support learning (Laski et al., 2015), but in other cases, manipulatives may be distracting (Bruner, 1966), particularly for younger learners (ages 3-6; Carbonneau et al., 2013). Manipulatives can be used in specific ways to improve abstract learning (McNeil & Uttal, 2009), but may hinder learning if the objects have features that do not align with the information to be learned (Beilock & Goldin-Meadow, 2010).

In Chapter 2, toddlers learned which actions should be performed with physical objects; their learning was relatively concrete and bounded to the materials with which they were working. Yet, they effectively transferred the actions they had learned to similar toys comprised of pieces of different colors and textures, suggesting that children had abstracted the actions they had learned on the taught toys. The particular physical features of the taught items did not impede children's relatively close abstraction of learned information. In addition, children remembered the goal states of the objects they had assembled one year later when tested with 2D representations of the toys. This level of information transfer, from a 3D environment to a 2D representation, demonstrated that physical activity can have lasting effects on children's visual recognition memory.

In Chapters 3 and 4, children experienced a 2D online environment rather than a physical environment. They learned about the particular lock-key matches but also generalized the abstract unlocking rule to novel locks and keys. Children generalized the rule at higher levels in the online environment than children had previously in the same task using physical locks and keys (Radovanovic et al., in preparation). Perhaps acting in a less concrete system (online) allowed children to abstract the rule more easily than when children acted in a physical environment. The physical materials may have distracted children from learning the abstract rule. In addition, the physical world has more degrees of freedom relative to an online environment. In general, the material to be learned (concrete versus abstract information) affected the environment in which that learning could occur best (physical versus online). Future research should continue to investigate when physical materials or 2D virtual representations of materials support different learning and transfer outcomes in learners of different ages.

Independent Success

Successful action performance supported children's generalization in Chapters 3 and 4. Children who achieved more successes than failures while attempting to unlock the locks generalized the unlocking rule more accurately. This contrasts with literature on productive failure (Loibl & Rummel, 2013), which suggests that when children make mistakes, they compare errors to later instruction and learn more effectively. Instead, children who achieved more successes when acting independently generalized the rule more accurately. The mistakes that children did make when they explored suggested that children engaged in trial-and-error hypothesis testing. Specifically, children who received instruction were the most directed in their hypothesis-testing behavior. In turn, successful hypothesis-testing improved children's generalization.

Children's success rates supported generalization under the specific learning conditions of the problem-solving task. Children had a limited amount of time to actively explore the problem space; when learning time was limited, children learned best from successful active experience guided by instruction. It is possible that when learners have more time to engage with problems over the course of a classroom lesson or over weeks or months of learning, productive failure and other methods such as discovery learning (Alfieri et al., 2011; Dean & Kuhn, 2007) may be effective. The success of children's actions also depends on how informative each action was for learning. Informativeness depends on the constraints or degrees of freedom in the problem space. The online environment had fewer degrees of freedom than the in-person Lock and Keys Task, which made children's actions highly informative for learning. In addition, learners' maturity levels affected their action successes. Indeed, adults in Study 6 were more successful in unlocking the locks compared to children; more advanced learners may encounter more success while acting, which can in turn support learning outcomes.

It is possible that performing actions successfully contributed to toddlers' action learning in Chapter 2 as well. All children were highly successful in performing taught actions under the guidance of an experimenter. Indeed, prior research (Brezack et al., 2021) found that when caregivers guided toddlers' actions to be more successful, toddlers demonstrated better learning of the taught actions. This suggests that the effect of active experience on memory in Chapter 2 might be due to children's high levels of action success in addition to physical activity. To test whether children's successes while acting causally improved learning, children's successes could be directly manipulated. The constraints on the problem space could be changed to make children's attempts more or less likely to yield success. For example, the actions children performed could be made more or less physically challenging. This same question could be addressed in the Lock and Keys Task; the mouse movements in the problem-solving game could be slowed or made more effortful to examine whether action success directly supported learning.

The Role of Instruction

Across the chapters of this dissertation, different types of instruction were coupled with children's active engagement to support their learning. In Chapter 2, guidance was provided concurrently with children's actions: the experimenter instructed each action children performed. In contrast, instruction was provided separately from active experience in Chapters 3 and 4: Children saw an instructional demonstration before or after they had a chance to act independently. Future research could examine whether instruction provided concurrently with or separately from active engagement would be more supportive of children's learning. Importantly, the instructions in both the action learning task and the problem-solving task made children's independent actions more successful. Concurrent guidance ensured that nearly all toddlers performed the taught actions perfectly in Chapter 2. After children viewed the instructional demonstration in Chapters 3 and 4, their actions were more successful compared to children's actions prior to receiving instruction.

Is instruction necessary for learning? In Studies 3 and 5, given enough time, children likely would have figured out how to unlock the locks in the problem-solving task. When exploration time was limited, active experience guided by instruction supported generalization. In contrast, toddlers likely would not have figured out the correct way to assemble the objects in Chapter 2 without support. Prior research (Brezack et al., 2021) demonstrated that when children were not instructed to act on similar toys, they rarely stumbled upon the correct actions independently. Conventional information, such as the correct uses for artifacts, can be challenging or impossible for children to discover without guidance (Tomasello, 2001), and conventional systems like mathematics could not be discovered without help. Indeed, when infants are left to freely explore novel objects, observing another person's actions supports action learning more than unguided activity (Meltzoff, 1985; Fagard & Lockman, 2010; Somogyi et al., 2015). Though children can and do learn well from observing, the results of Chapter 2 suggest there may be a specific memory benefit associated with guided active experience. In contexts where instruction was crucial (Chapter 2), or at least made learning more efficient (Chapters 3 and 4), there was still a role for children's active experience in supporting their learning.

In addition, instruction content can support children's learning of more complex systems. For example, conceptual instructions about the equals sign provided prior to children's independent exploration improved their learning of math equivalence (Fyfe et al., 2014). In Chapters 3 and 4, the instruction demonstrations were procedural rather than conceptual; children were shown the locks unlocking, but were not explicitly told the rule. If children had been told the rule with linguistic instructions (e.g., "There are two parts of the rule: The colors on the keys have to match the locks and the keys have to be round"), children would surely have learned the rule and would have likely generalized to new examples. In addition, children were only shown successful unlocking; they were not shown the counterfactual evidence or the critical key comparisons that would have disambiguated the unlocking rule. Such evidence would likely also led to effective learning and generalization. Even so, it is possible that guided, successful active experience would generate better learning than conceptual instructions or fully disambiguating evidence. For example, successful active experience might support children's long-term memory more effectively than the clearest instructions.

Learning from Observing

In this dissertation, active experience did not uniformly improve learning compared to observational experience. In Study 1, toddlers learned and generalized taught actions just as accurately from acting as they did from observing. In addition, children in Study 5 learned equally well from a double-dose of activity as they did from overserving another child's actions. Why was observation in some cases as effective as active experience? All of the information necessary to learn the taught material was provided in the matched observational conditions in Chapter 2 and Chapter 4 as was present in the active conditions. It therefore makes sense that children could glean the same information from watching as from doing. Active experience, however, improved children's memory (Study 2) and generalization (marginally, Study 5). It is possible that active engagement during learning deepens what children learn or makes learning "stick" over time.

It is also possible that children learned well from observing because they were mentally active. In both the action learning (Chapter 2) and problem-solving (Chapters 3 and 4) studies, children practiced performing actions that were similar to those they were later taught. Priming children with an action improved their action memory (Howard et al., 2020). Practicing could have also increased the relevance of the observed actions for children (Meyer et al., 2022; 2011) since they likely assumed they would be performing those actions as well. Children may have also interpreted the observational condition as collaborative. During the practice phase in Study 1, children took turns with the experimenter to build a block tower. Turn-taking could have created a collaborative context in which children were more likely to overclaim observed actions as their own (the "I did it" bias; Sommerville & Hammond, 2007). It was clearer in the observation condition in the problem-solving task (Study 5) that children were not acting since they could see another child’s mouse moving on the screen. Even so, it is possible that children felt as though they were collaborating with an invisible person as they watched another child’s actions. Children may have even clicked and dragged along with the mouse on the screen, which could have increased children’s mental activity and sense of collaboration.

To tease apart whether children were mentally engaged when observing, and whether such mental engagement improved learning outcomes, future work could more directly manipulate children’s mental activity. Children could observe actions in contexts that are more or less collaborative, similar to work by Sommerville and Hammond (2007). In addition, future research could continue to measure neural correlates of active engagement (the mirror neuron system; Marshall & Meltzoff, 2014; theta oscillations; Begus & Bonawitz, 2020) to examine whether more neural activity induced by different contexts improved children’s observational learning (e.g., Meyer et al., 2022; 2011).

Limitations and Future Directions

One limitation of this dissertation is that we did not measure children’s long-term memory in Chapters 3 and 4. We therefore do not know whether instructed exploration would improve children’s memory relative to uninstructed exploration, two opportunities to explore, or a matched observational experience. Instructed active experience improved generalization relative to uninstructed active experience, but was only marginally better than two opportunities to explore or observing another child’s exploration. Future research should continue measuring varied learning outcomes, including long-term memory, to examine whether and how instructed active experience improves learning.

In addition, the generalization measures used in these studies were relatively close levels of transfer from the taught information. In Chapter 2, the generalization toys were highly similar to the taught toys. In Chapters 3 and 4, the unlocking rule was applied to novel locks that had different colors than the taught locks. Further transfer could be measured with the types of tested items, such as more dissimilar toys or locks and keys. Cross-modal generalization could also be measured in future research. Toddlers in Study 1 were taught to perform actions on physical objects, but their memory was tested with 2D images in Study 2. Children in the in-person problem-solving task (Radovanovic et al., in preparation) learned to unlock physical locks, but generalization was tested with 2D images. It would be interesting to examine whether the actions children learned in Study 1 could be transferred to actions on digital representations of the objects, or whether the rule learned online in Studies 3 and 5 could be transferred to physical locks and keys.

The instructions in this dissertation were either offered concurrently with children's action (Chapter 2) or separately from children's exploration (Chapters 3 and 4). Future studies could manipulate the types of guidance offered before, after, and while children are actively engaging with the information to be learned. For example, the action learning task could have included separate verbal instructions before children performed actions, and the problem-solving task could have included guidance responsive to children's actions. As mentioned previously, the content of the instructions can also influence children's activity and learning. Linguistic instructions that disambiguate the information to be learned could be equally or more effective than children's active experience, though more research is necessary to understand which learning outcomes would be affected by different types of instruction. The effectiveness of different forms of instruction also likely vary with children's maturity; linguistic instructions might be effective for 6-year-olds, but would be less effective for toddlers.

Conclusion

Across six preregistered experiments, young children and adults learned novel conventional information and rules through instruction that either concurrently guided their activity, or instruction that preceded or followed exploration. In addition, different learning outcomes were measured: children were tested on their learning of taught information, generalization to new contexts, and long-term memory. The results showed that physically-active experience guided by instruction supported toddlers' memory. In contrast, successful active exploration guided by prior instruction supported 6-year-old children's rule generalization. In sum, even when children were instructed, children's active engagement with the conventional information to be learned improved children's learning outcomes.

REFERENCES

Adams, F. (2010). Embodied cognition. *Phenomenology and the Cognitive Sciences*, *9*(4), 619-628. https://doi.org/10.1007/s11097-010-9175-x

Adolph, K., & Joh, A. S. (2007). Motor development: How infants get into the act. In A. Slater, & M. Lewis (Eds.), *Introduction to infant development* (2 ed., pp. 63-80). Oxford University Press.

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning?. *Journal of educational psychology*, *103*(1), 1. https://doi.org/10.1037/a0021017

Amso, D., & Scerif, G. (2015). The attentive brain: insights from developmental cognitive neuroscience. *Nature Reviews Neuroscience, 16*(10), 606-619. https://doi.org/10.1038/nrn4025

Baek, S., Jaffe-Dax, S., & Emberson, L. L. (2020). How an infant's active response to structured experience supports perceptual-cognitive development. *Progress in Brain Research*, *254*, 167-186. https://doi.org/10.1016/bs.pbr.2020.05.015

Bandura, A. (1977). Social Learning Theory. New York, NY: General Learning Press.

Barr, R., Dowden, A., & Hayne, H. (1996). Developmental changes in deferred imitation by 6-to 24-month-old infants. *Infant behavior and development, 19*(2), 159-170. https://doi.org/10.1016/S0163-6383(96)90015-6

Barrett, T. M., Traupman, E., & Needham, A. (2008). Infants' visual anticipation of object structure in grasp planning. *Infant Behavior and Development, 31*(1), 1-9. https://doi.org/10.1016/j.infbeh.2007.05.004

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Bauer, P. J. (1996). What do infants recall of their lives? Memory for specific events by one- to two-year-olds. *American Psychologist, 51*(1), 29–41. https://doi.org/10.1037/0003-066X.51.1.29

Bauer, P. J., Hertsgaard, L. A., & Dow, G. A. (1994). After 8 months have passed: Long-term recall of events by 1-to 2-year-old children. *Memory, 2*(4), 353-382. https://doi.org/10.1080/09658219408258955

Bayley, N. (2006). Bayley Scales of Infant and Toddler Development – Third Edition. San Antonio, TX: Harcourt Assessment.

Begus, K., Gliga, T., & Southgate, V. (2014). Infants learn what they want to learn: Responding to infant pointing leads to superior learning. *PloS one*, *9*(10), e108817. https://doi.org/10.1371/journal.pone.0108817

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological science*, *21*(11), 1605-1610. https://doi.org/10.1177/0956797610385353

Bonawitz, E., Bass, I., & Lapidow, E. (2018). Choosing to learn: Evidence evaluation for active learning and teaching in early childhood. In *Active learning from infancy to childhood* (pp. 213-231). Springer, Cham.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322-330. https://doi.org/10.1016/j.cognition.2010.10.001

Boudreau, J. P., & Bushnell, E. W. (2000). Spilling thoughts: Configuring attentional resources in infants' goal-directed actions. *Infant behavior and Development, 23*(3-4), 543-566. https://doi.org/10.1016/S0163-6383(01)00059-5

Brezack, N., Radovanovic, M., & Woodward, A. (2021). Everyday interactions support toddlers' learning of conventional actions on artifacts. *Journal of Experimental Child Psychology.* https://doi.org/10.1016/j.jecp.2021.105201

Bruner, J. (1993). Do we "acquire" culture or vice versa?. *Behavioral and Brain Sciences*, *16*(3), 515-516. https://doi.org/10.1017/S0140525X00031290

Bruner, J. S. (1966). *Toward a theory of instruction* (Vol. 59). Harvard University Press.

Callanan, M. A., Legare, C. H., Sobel, D. M., Jaeger, G. J., Letourneau, S., McHugh, S. R., ... & Watson, J. (2020). Exploration, explanation, and parent–child interaction in museums. Monographs of the Society for Research in Child Development, 85(1), 7-137. https://doi.org/10.1111/mono.12412

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology*, *105*(2), 380. https://doi.org/10.1037/a0031084

Clark, R. E. (2009). How much and what type of guidance is optimal for learning from instruction. In S. Tobias & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure?* (pp. 158–183). Routledge/Taylor & Francis Group.

Cohen, R. L. (1989). Memory for action events: The power of enactment. *Educational Psychology Review, 1*(1), 57-80. https://doi.org/10.1007/BF01326550

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. *Processes of change in brain and cognitive development. Attention and performance XXI*, *21*, 249-274.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Danovitch, J. H., & Mills, C. M. (2018). Understanding when and how explanation promotes exploration. In *Active learning from infancy to childhood* (pp. 95-112). Springer, Cham.

Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education, 91*(3), 384–397. https://doi.org/10.1002/sce.20194

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology, 113*(3), 552–568. https://doi.org/10.1016/j.jecp.2012.06.009

Fagard, J., & Lockman, J. J. (2010). Change in imitation for object manipulation between 10 and 12 months of age. *Developmental Psychobiology, 52*(1), 90–99. https://doi.org/10.1002/dev.20416

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics, 21*(1), 95–115. https://doi.org/10.1017/S0142716400001053

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, *28*(4), 717-741. https://doi.org/10.1007/s10648-015-9348-9

Fisher, K. R., Hirsh-Pasek, K., Newcombe, N., & Golinkoff, R. M. (2013). Taking shape: Supporting preschoolers' acquisition of geometric knowledge through guided play. *Child development*, *84*(6), 1872-1878. https://doi.org/10.1111/cdev.12091

Fyfe, E. R., DeCaro, M. S., & Rittle-Johnson, B. (2014). An alternative time for telling: When conceptual instruction prior to problem solving improves mathematical knowledge. British journal of educational psychology, 84(3), 502-519. https://doi.org/10.1111/bjep.12035

Gaskins, S., & Paradise, R. (2010). Learning through observation in daily life. In D. F. Lancy, J. Bock, & S. Gaskins (Eds.), The Anthropology of Learning in Childhood (pp. 85–117). Rowman & Littlefield Publishers, Inc.

Gerson, S. A., & Woodward, A. L. (2014). Learning from their own actions: The unique effect of producing actions on infants' action understanding. *Child Development, 85*(1), 264-277. https://doi.org/10.1111/cdev.12115

Gerson, S. A., Schiavio, A., Timmers, R., & Hunnius, S. (2015). Active drumming experience increases infants' sensitivity to audiovisual synchrony during observed drumming actions. *PLoS One*, *10*(6), e0130960. https://doi.org/10.1371/journal.pone.0130960

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological bulletin*, *138*(6), 1085. https://doi.org/10.1037/a0028044

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron*, *84*(2), 486-496. https://doi.org/10.1016/j.neuron.2014.08.060

Gruber, Matthias J., Valji, Ashanti and Ranganath, Charan 2019. *Curiosity and learning: A neuroscientific perspective.* In: Renninger, K. Ann and Hidi, Suzanne E. eds. *The Cambridge Handbook of Motivation and Learning*, Cambridge: Cambridge University Press, pp. 397-417.

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science*, *7*(5), 464-481. https://doi.org/10.1177/1745691612454304

Han, M., Moore, N., Vukelich, C., & Buell, M. (2010). Does play make a difference? How play intervention affects the vocabulary learning of at-risk preschoolers. *American Journal of Play*, *3*(1), 82-105.

Hayne, H., Barr, R., & Herbert, J. (2003). The effect of prior practice on memory reactivation and generalization. *Child Development, 74*(6), 1615-1627. https://doi.org/10.1046/j.1467-8624.2003.00627.x

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99-107. https://doi.org/10.1080/00461520701263368

Howard, L. H., Riggins, T., & Woodward, A. L. (2020). Learning from others: The effects of agency on event memory in young children. *Child Development, 91*(4), 1317-1335. https://doi.org/10.1111/cdev.13303

Hutmacher, F., & Kuhbandner, C. (2018). Long-term memory for haptically explored objects: fidelity, durability, incidental encoding, and cross-modal transfer. *Psychological science*, *29*(12), 2031-2038. https://doi.org/10.1177/0956797618803644

James, K. H., & Swain, S. N. (2011). Only self-generated actions create sensori-motor systems in the developing brain. *Developmental science*, *14*(4), 673-678. https://doi.org/10.1111/j.1467-7687.2010.01011.x

Jirout, J., & Klahr, D. (2012). Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental review*, *32*(2), 125-160. https://doi.org/10.1016/j.dr.2012.04.002

Johnson, C. J., Paivio, A. U., & Clark, J. M. (1989). Spatial and verbal abilities in children's crossmodal recognition: A dual coding approach. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *43*(3), 397. https://doi.org/10.1037/h0084229

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science*, *40*(4), 651-672. https://doi.org/10.1007/s11251-012-9209-6

Keen, R. (2011). The development of problem solving in young children: A critical cognitive skill. Annual Review of Psychology, 62, 1–21. https://doi.org/10.1146/annurev.psych.031809.130730

Kinder, K. T., & Buss, A. T. (2021). The effect of motor engagement on memory: Testing a motor-induced encoding account. *Memory & Cognition*, *49*(3), 586-599. https://doi.org/10.3758/s13421-020-01113-6

Kirschner, P., Sweller, J., & Clark, R. E. (2006). Why unguided learning does not work: An analysis of the failure of discovery learning, problem-based learning, experiential learning and inquiry-based learning. *Educational Psychologist, 41*(2), 75-86. https://doi.org/10.1207/s15326985ep4102_1

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological science*, *15*(10), 661-667. https://doi.org/10.1111/j.0956-7976.2004.00737.x

Knoblich, G., & Sebanz, N. (2006). The social nature of perception and action. *Current Directions in Psychological Science, 15*(3), 99-104. https://doi.org/10.1111/j.0963-7214.2006.00415.x

Kontra, C., Goldin-Meadow, S., & Beilock, S. L. (2012). Embodied learning across the life span. *Topics in cognitive science*, *4*(4), 731-739. https://doi.org/10.1111/j.1756-8765.2012.01221.x

Kushnir, T., Wellman, H. M., & Gelman, S. A. (2009). A self-agency bias in preschoolers' causal inferences. *Developmental Psychology*, *45*(2), 597. https://doi.org/10.1037/a0014727

Laski, E. V., Jor'dan, J. R., Daoust, C., & Murray, A. K. (2015). What Makes Mathematics Manipulatives Effective? Lessons From Cognitive Science and Montessori Education. *SAGE Open*. https://doi.org/10.1177/2158244015589588

Legare, C. H. (2019). The development of cumulative cultural learning. *Annual Review of Developmental Psychology*, 1, 119-147. https://doi.org/10.1146/annurev-devpsych-121318-084848

Lie, E., & Newcombe, N. S. (1999). Elementary school children's explicit and implicit memory for faces of preschool classmates. *Developmental Psychology, 35*(1), 102. https://doi.org/10.1037/0012-1649.35.1.102

Lillard, A. S. (2016). Montessori: The science behind the genius. Oxford University Press.

Loibl, K. & Rummel, N. (2013). Delaying Instruction Alone Doesn't Work: Comparing and Contrasting Student Solutions is Necessary for Learning from Problem-Solving Prior to Instruction. In Rummel, N., Kapur, M., Nathan, M., & Puntambekar, S. (Eds.), To See the World and a Grain of Sand: Learning across Levels of Space, Time, and Scale: CSCL 2013 Conference Proceedings Volume 1. https://doi.dx.org/10.22318/cscl2013.1.296

Lubin, A., Rossi, S., Poirel, N., Lanoë, C., Pineau, A., & Houdé, O. (2015). The role of self-action in 2-year-old children: An illustration of the arithmetical inversion principle before formal schooling. *Child Development Research*, *2015*. https://doi.org/10.1155/2015/879258

Lucca, K., & Wilbourn, M. P. (2018). Communicating to learn: Infants' pointing gestures result in optimal learning. *Child development*, *89*(3), 941-960. https://doi.org/10.1111/cdev.12707

Mangold (2017): INTERACT User Guide. Mangold International GmbH (Ed.) www.mangold-international.com

Markant, D. B., Ruggeri, A., Gureckis, T. M., & Xu, F. (2016). Enhanced memory as a common effect of active learning. *Mind, Brain, and Education, 10*(3), 142-152. https://doi.org/10.1111/mbe.12117

Marshall, P. J., & Meltzoff, A. N. (2014). Neural mirroring mechanisms and imitation in human infants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1644), 20130620. https://doi.org/10.1098/rstb.2013.0620

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning?. American psychologist, 59(1), 14. https://doi.org/10.1037/0003-066X.59.1.14

McCarty, M. E., Clifton, R. K., & Collard, R. R. (1999). Problem solving in infancy: the emergence of an action plan. *Developmental psychology*, *35*(4), 1091. https://doi.org/10.1037/0012-1649.35.4.1091

McDonough, L., & Mandler, J. M. (1994). Very long-term recall in infants: Infantile amnesia reconsidered. *Memory, 2*(4), 339-352. https://doi.org/10.1080/09658219408258954

McNeil, N. M., & Uttal, D. H. (2009). Rethinking the use of concrete materials in learning: Perspectives from development and education. *Child development perspectives*, *3*(3), 137-139. https://doi.org/10.1111/j.1750-8606.2009.00093.x

Meltzoff, A. N. (1985). Immediate and deferred imitation in fourteen-and twenty-four-month-old infants. *Child Development*, 62-72. https://doi.org/10.2307/1130174

Meltzoff, A. N. (1988). Infant imitation and memory: nine-month-olds in immediate and deferred tests. *Child Development, 59*(1), 217–225. https://doi.org/10.1111/j.1467-8624.1988.tb03210.x

Meyer, M., Chung, H., Debnath, R., Fox, N., & Woodward, A. L. (2022). Social context shapes neural processing of others' actions in 9-month-old infants. *Journal of Experimental Child Psychology, 213*, 105260. https://doi.org/10.1016/j.jecp.2021.105260

Meyer, M., Hunnius, S., Van Elk, M., Van Ede, F., & Bekkering, H. (2011). Joint action modulates motor system involvement during action observation in 3-year-olds. *Experimental Brain Research, 211*(3), 581-592. https://doi.org/10.1007/s00221-011-2658-3

Mix, K. S., Moore, J. A., & Holcomb, E. (2011). One-to-one play promotes numerical equivalence concepts. *Journal of cognition and development*, *12*(4), 463-480. https://doi.org/10.1080/15248372.2011.554928

Möhring, W., & Frick, A. (2013). Touching up mental rotation: Effects of manual experience on 6-month-old infants' mental object rotation. *Child development*, *84*(5), 1554-1565. https://doi.org/10.1111/cdev.12065

Montessori, M. (1991/1918) Spontaneous Activity in Education (reprinted as The Advanced Montessori Method I). Oxford: Clio Press

Needham, A., Barrett, T., & Peterman, K. (2002). A pick-me-up for infants' exploratory skills: Early simulated experiences reaching for objects using "sticky mittens" enahnces young infants object exploration skills. *Infant Behavior and Development, 25*(3), 279–295. https://doi.org/10.1016/S0163-6383(02)00097-8

Nilsson, L. G. (2000). Remembering actions and words. The Oxford handbook of memory, 137-148.

Pereira, A. F., Shen, H., Smith, L. B., & Yu, C. (2009). A first-person perspective on a parent-child social interaction during object play. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 31, No. 31).

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: movement influences infant rhythm perception. *Science*, *308*(5727), 1430-1430. https://doi.org/10.1126/science.1110922

Piaget, J. (1953). The origin of intelligence in the child. London: Routledge.

Piaget, J. (1964). Piaget rediscovered: A report on the conference on cogntive studies and curriculum development. In R. E. Ripple & V. N. Rockcastle (Eds.), A report of the Jean Piaget Conferences at Cornell University and the University of California (pp. 7-20). Cornell University.

Pouw, W. T., Van Gog, T., & Paas, F. (2014). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review*, *26*(1), 51-72. https://doi.org/10.1007/s10648-014-9255-5

Radovanovic, Brezack, Shneidman, & Woodward. (in preparation). Unlocking learning: Exploration before instruction supports children's problem-solving.

Rochat, P. (1989). Object manipulation and exploration in 2-to 5-month-old infants. *Developmental Psychology, 25*(6), 871. https://doi.org/10.1037/0012-1649.25.6.871

Rogoff, B. (1990). Apprenticeship in thinking: Cognitive development in social context. Oxford University Press.

Savery, J. R., & Duffy, T. M. (June 2001). Problem Based Learning: An instructional model and its constructivist framework (technical report no. 16–01). Center for Research on Learning and Technology.

Saylor, M., & Ganea, P. (2018). *Active Learning from Infancy to Childhood*. Springer International Publishing AG.

Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental science*, *10*(3), 322-332. https://doi.org/10.1111/j.1467-7687.2007.00587.x

Schulz, L., Kushnir, T., & Gopnik, A. (2007). Learning from doing: Intervention and causal inference. *Causal learning: Psychology, philosophy, and computation*, 67-85. https://doi.org/10.1093/acprof:oso/9780195176803.003.0006

Shneidman, L., & Woodward, A. L. (2015). Are child-directed interactions the cradle of social learning? *Psychological Bulletin, 142*(1), 1-17. https://doi.org/10.1037/bul0000023

Sobel, D. M., & Letourneau, S. M. (2018). Curiosity, exploration, and children's understanding of learning. In *Active learning from infancy to childhood* (pp. 57-74). Springer, Cham.

Sobel, D. M., & Sommerville, J. A. (2010). The importance of discovery in children's causal learning from interventions. *Frontiers in Psychology*, 1, 1–7. https://doi.org/10.3389/fpsyg.2010.00176

Sommerville, J. A., & Hammond, A. J. (2007). Treating another's actions as one's own: Children's memory of and learning from joint activity. *Developmental Psychology, 43*(4), 1003–1018. https://doi.org/10.1037/0012-1649.43.4.1003

Sommerville, J. A., Hildebrand, E. A., & Crane, C. C. (2008). Experience matters: the impact of doing versus watching on infants' subsequent perception of tool-use events. *Developmental Psychology, 44*(5), 1249. https://doi.org/10.1037/a0012296

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. Cognition, 96(1), B1-B11. https://doi.org/10.1016/j.cognition.2004.07.004

Somogyi, E., Ara, C., Gianni, E., Rat-Fischer, L., Fattori, P., O'Regan, J. K., & Fagard, J. (2015). The roles of observation and manipulation in learning to use a tool. *Cognitive Development*, 35, 186-200. https://doi.org/10.1016/j.cogdev.2015.06.002

Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in development: Motor skill acquisition facilitates three-dimensional object completion. *Developmental Psychology, 46*(1), 129–138. https://doi.org/10.1037/a0014618

Stockard, J., Wood, T. W., Coughlin, C., & Rasplica Khoury, C. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half century of research. *Review of Educational Research*, *88*(4), 479-507. https://doi.org/10.3102/0034654317751919

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*(3), 251-296. https://doi.org/10.1023/A:1022193728205

Thelen, E. (2000). Grounded in the world: Developmental origins of the embodied mind. *Infancy*, *1*(1), 3-28. https://doi.org/10.1207/S15327078IN0101_02

Tomasello, M. (1999). The Cultural Origins of Human Cognition. Harvard University Press.

Tomasello, M. (2001). Cultural transmission: A view from chimpanzees and human infants. *Journal of Cross-Cultural Psychology, 32*(2), 135–146. https://doi.org/10.1177/0022022101032002002

Trewartha, K. M., Case, S., & Flanagan, J. R. (2015). Integrating actions into object location memory: A benefit for active versus passive reaching movements. *Behavioural Brain Research*, 279, 234-239. https://doi.org/10.1016/j.bbr.2014.11.043

Vygotsky, L. S. (1980). Mind in Society: The development of higher psychological processes. Harvard University Press.

Wakefield, E. M., Hall, C., James, K. H., & Goldin-Meadow, S. (2018). Gesture for generalization: Gesture facilitates flexible learning of words for actions on objects. *Developmental science*, *21*(5), e12656. https://doi.org/10.1111/desc.12656

Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Kittredge, A. K., & Klahr, D. (2016). Guided play: Principles and practices. *Current directions in psychological science*, *25*(3), 177-182. https://doi.org/10.1177/0963721416645512

Wickham, H., François, R., Henry, L., & Müller, K. (2018). dplyr: A Grammar of Data Manipulation. R package version 0.7.6. https://CRAN.R-project.org/package=dplyr

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review, 9*(4), 625-636. https://doi.org/10.3758/BF03196322

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *The Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Yang, S. H. (2017, January). Teaching versus active learning: A computational analysis of conditions that affect learning. In *Proceedings of the Cognitive Science Society*.

Yu, Y., Shafto, P., Bonawitz, E., Yang, S. C. H., Golinkoff, R. M., Corriveau, K. H., ... & Xu, F. (2018). The theoretical and methodological opportunities afforded by guided play with young children. *Frontiers in psychology*, *9*, 1152. https://doi.org/10.3389/fpsyg.2018.01152

Zosh, J. M., Hirsh-Pasek, K., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., ... & Whitebread, D. (2018). Accessing the inaccessible: Redefining play as a spectrum. *Frontiers in psychology*, *9*, 1124. https://doi.org/10.3389/fpsyg.2018.01124