Early Childhood Research Quarterly 46 (2019) 275–284

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Exploring potential cognitive foundations of scientific literacy in preschoolers: Causal reasoning and executive function

Jessie-Raye Bauer*, Amy E. Booth

Department of Psychology, The University of Texas at Austin, Austin, TX, 78712, United States

ARTICLE INFO

Article history:
Received 24 May 2017
Received in revised form 1 September 2018
Accepted 17 September 2018
Available online 16 November 2018

Keywords:
Causal reasoning
Executive function
Science
Education

ABSTRACT

Despite increasing emphasis in the United States on promoting student engagement and achievement in science, technology, engineering, and mathematics (STEM) fields, the origins of scientific literacy remain poorly understood. We begin to address this limitation by considering the potential contributions of two distinct domain-general skills to early scientific literacy. Given their relevance to making predictions and evaluating evidence, we consider the degree to which causal reasoning skills relate to scientific literacy (as measured by an adaptive standardized test specifically designed for preschoolers). We also consider executive function (EF) as a potentially more fundamental contributor. While previous research has demonstrated that EF is predictive of achievement in other core academic domains like reading and math, its relationship to scientific literacy, particularly in early childhood, has received little attention. To examine how causal reasoning and EF together potentially relate to the development of scientific literacy in young children, we recruited 125 3-year-olds to complete three causal reasoning tasks, three EF tasks, and the aforementioned measure of scientific literacy. Results from a series of hierarchical regressions revealed that EF, and one measure of causal reasoning (causal inferencing) were related to scientific literacy, even after controlling for age, ethnicity, maternal education, and vocabulary knowledge. Moreover, causal inferencing ability was a significant partial mediator between EF and scientific literacy. Although additional research will be required to further specify the nature of these relationships, the current work suggests that EF has the potential to support scientific literacy, perhaps in part, by scaffolding causal reasoning skills.

1. Introduction

Despite research and policy recommendations highlighting the importance of science to early childhood education, preschool classrooms typically do not focus on promoting learning in this domain (Gopnik, 2012; Nayfeld, Fuccillo, & Greenfield, 2013; National Research Council [NRC], 2012; Next Generation Science Standards [NGSS] Lead States, 2013; U.S. Department of Health and Human Services (U.S. DHHS), 2015). Unfortunately, this is especially true of preschools that primarily serve low-income families (Greenfield et al., 2009; Nayfeld, Brenneman, & Gelman, 2011). In the rare instances in which curricula aimed at improving the scientific skills and knowledge of young children are implemented, results have been mixed (Greenfield et al., 2009; Klahr, 2005; Klahr, Zimmerman, & Jirout, 2011). This is concerning in light of recent slippage in our nation's standing as a world leader in scientific innovation, as well as persistent achievement gaps within the United States. One reason for our slow progress in developing effective early science programming might have to do with our limited understanding of the developmental foundations of scientific literacy.

One promising contributor to early scientific literacy is causal reasoning. Evidence suggests that even toddlers are capable of some forms of causal reasoning (see Gopnik, & Schulz, 2007). Considering how central causal discovery is to hypothesis testing and theory development, it seems likely that early emerging causal reasoning capabilities scaffold the development of more formal and sophisticated scientific inquiry skills and knowledge. That said, the emergence of scientific literacy surely cannot rest on the development of causal reasoning skills alone. Indeed, executive function (EF) may be an even more fundamental contributor. Not only might EF directly facilitate the acquisition of factual and conceptual knowledge about science by optimally directing attention during the learning process, it might also indirectly support the development of scientific literacy by pro-

* Corresponding author at: 108 E. Dean Keeton, Stop A8000, Austin, TX 78712, United States.
E-mail address: jessie.raye@utexas.edu (J.-R. Bauer).

https://doi.org/10.1016/j.ecresq.2018.09.007
0885-2006/

moting successful and efficient causal reasoning ability. Several studies conducted with elementary school-aged children have already shown that EF is strongly related to academic achievement in other domains like reading and mathematics (Best, Miller, & Naglieri, 2011; Blair & Razza, 2007; Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009; Clark et al., 2014; Nathanson, & Grimm, 2009). However, potential relationships between EF, causal reasoning, and scientific literacy remain largely unspecified.

Our goal in this paper is to determine whether causal reasoning and EF skills are related to early scientific literacy. Although we take a contemporaneous correlational approach, thereby precluding any definitive causal conclusions, this study will help to clarify whether the former have the potential to lay the foundation for development of the latter. To address this goal, we examined EF and three different measures of causal reasoning as they relate to a standardized measure of scientific literacy in children who are just beginning preschool.

1.1. Scientific literacy in preschoolers

Current national preschool and kindergarten science standards conceptualize scientific literacy in accordance with a three-dimensional integrated framework that includes disciplinary content, scientific practices, and crosscutting concepts (NRC, 2012). With respect to disciplinary content, current standards emphasize four broad conceptual areas: life sciences (e.g., animals, organisms, evolution), earth and space sciences (e.g., weather, the earth's composition), physical sciences (e.g., object properties, states and changes), and engineering and technology (e.g., structures and tools) (NGSS, 2013; NRC, 2012; U.S. DHHS, 2015). In the rare case when a preschool classroom emphasizes science (Brenneman, 2011; Mantzicopoulos, Patrick, & Samarapungavan, 2013), these content areas are evident when children learn about gravitational, and other, forces (e.g., why balls roll down hills), what humans, plants and animals need to live and grow (e.g., water, food and sunlight), object properties and state changes (e.g., why ice melts), weather (e.g., how the clouds make rain), and properties of structures (e.g., stable and unstable buildings).

With respect to scientific practices, young children are expected to develop abilities foundational to scientific experimentation and evidence-based reasoning (Klahr, 2005; NGSS, 2013; NRC, 2012; U.S. DHHS, 2015; Zimmerman, 2007). For example, children might be encouraged to begin observing the world around them by focusing on what they see and hear, and by learning to use tools like magnifying glasses and rulers to enhance their observations. Children might further learn how to describe, record, and reflect on what they observe (e.g., "this block is heavy and smooth") in ways that can help them figure out how the world works. These skills are thought in turn to scaffold the ability to make comparisons (e.g., a block sinks, but a leaf floats), generate questions (e.g., "*why* does this block sink but this leaf float?"), make predictions (e.g., "this ball is heavy like the block so I think it will sink"), and execute experiments (e.g., drop the ball in water to see if it will float) (e.g., Duschl, Schweingruber, & Shouse, 2007).

Current standards further advocate for a special focus on crosscutting concepts that bridge key disciplinary areas in science and engineering. Some examples include understanding systems and models, structure and function, and stability and change (NRC, 2012). In some respects, understanding cause and effect might be considered the most fundamental crosscutting concept of all because it is required for a full appreciation of each of the others, as well as for understanding specific disciplinary content.

1.2. Causal reasoning as a potential foundation for scientific literacy

Causal reasoning is the process by which children make sense of relationships between objects and events that specify the rules of interaction governing the world around them (Kushnir, Gopnik, Schulz, & Danks, 2003). Although the componential structure of this process has yet to be clearly articulated in the literature, tasks used to assess causal reasoning in children typically involve tracking probabilistic contingencies and drawing on existing knowledge to consider or make inferences about alternative causal mechanisms. In this study, we focus on three such tasks, chosen based on their prominence in the literature, and their potential for tapping distinct aspects of causal reasoning. First, we evaluate children's ability to diagnose causes from observed spatiotemporal contingencies between objects and outcomes (Gopnik, Sobel, Schulz, & Glymour, 2001). Second, we evaluate children's ability to infer causes from snapshots of the start and end states of transformative events (Das Gupta & Bryant, 1989; Gelman, Bullock & Meck, 1980). Finally, we measure children's ability to reason counterfactually about alternative chains of events in imagined stories (e.g., Guajardo & Turley-Ames, 2004; Rafetseder, Renate, & Perner, 2010; Rafetseder, Schwitalla, & Perner, 2013).

In line with current National Research Council guidelines, we propose that causal reasoning likely plays a foundational role in supporting the development of scientific literacy (NRC, 2012). This idea is further reflected in the preliminary framework of scientific thinking recently advanced by Tolmie, Ghazali, and Morris (2016), in which causal reasoning takes center stage. Specifically, Tolmie et al. (2016) argue that scientific literacy is supported primarily by the ability to extract, reason about, and explain observed causal phenomena. This framework, however, was developed to account for scientific thinking in primary and secondary school-aged children who were learning chemistry, biology, and advanced domain-specific vocabulary. Thus, these children were likely building upon an already extensive base of relevant content knowledge. The question remains as to whether these relationships between causal reasoning and scientific literacy hold in younger children when the latter is just beginning to take shape.

Converging evidence suggests that causal reasoning is indeed a promising foundation for early scientific literacy. It is well established that young children are inherently interested in trying to understand the causal structure of the world around them (e.g., Alvarez & Booth, 2014; Gopnik, 2000; Schulz & Bonawitz, 2007). For example, children will persist in asking questions about novel objects until their functions or other causal powers are revealed (Greif, Kemler Nelson, Keil, & Gutierrez, 2006). Children are also capable of extracting causal information from the environment on their own, for example, by tracking the spatiotemporal movements of objects and associated events (Gopnik et al., 2001; Schulz & Gopnik, 2004; Sobel & Kirkham, 2006). In one relevant study, Gopnik et al. (2001), first introduced children to a "blicket detector," a box that activated (e.g., lit up or played music) when some objects, but not others, came in contact with it. Children as young as three were, as a group, able to correctly identify which objects were causal and demonstrate this understanding by producing a novel event (i.e., making the box turn off even though they were never shown how).

In addition to tracking novel probabilistic contingencies, young children are also able to draw on existing knowledge to infer causal agents. For example, they are able to infer the causes of events when presented with only a sequence of still-frame pictures. In one foundational study demonstrating this capability in 3 and 4 year olds, Gelman et al. (1980) began by showing children pictures of events in which objects underwent changes (e.g., an intact cup next to the same cup in a broken state). When asked to select from an array of pictured alternatives, children were able to choose the correct causal instrument (e.g., hammer).

The ability of young children to infer a potential cause is also evident in more demanding tasks that require the consideration of counterfactuals, or situations that are inconsistent with a given reality. For example, in Guajardo & Turley-Ames (2004), children were told a story about a child playing in a muddy backyard. The child in the story becomes thirsty, and then goes inside to get some juice. Consequently, the floor inside gets muddy. When asked what they could have done so that the floor would *not* have gotten dirty, 3-, 4-, and 5-year-old children were able to spontaneously generate and consider counterfactual solutions (e.g., the child could have taken off their shoes before entering the house) (Drayton, Turley-Ames, & Guajardo, 2011; Guajardo & Turley-Ames, 2004). In order to answer correctly, these young children had to engage in a complex reasoning process in which they held certain event features constant, while systematically manipulating others to determine the causally relevant components. Together, the work reviewed here suggests that causal reasoning skills are in place early enough to potentially lay the foundation for scientific literacy. Causal reasoning, however, is not the only potential contributor to this foundation.

1.3. Executive function as a potential foundation for scientific literacy

Executive function (EF) refers to a set of skills that help control and regulate attention, action, and behavior (e.g., Best et al., 2011; Huizinga, Dolan, & van der Molen, 2006; Lehto, Juujarvi, Kooistra, & Pulkkinen, 2003; Miyake et al., 2000). It has been well established that EF undergoes significant development during the preschool years (e.g., Carlson, 2005; Diamond, 2013; Huizinga et al., 2006; Zelazo, Müller, Frye, & Marcovitch, 2003). Results from confirmatory factor analyses suggest that EF initially emerges in preschool as a unitary construct, wherein a variety of different EF tasks load onto one common latent factor (Wiebe, Espy & Charak, 2008; Wiebe et al., 2011). As overall skill levels improve throughout preschool, EF appears to grow more differentiated, such that by school-age, three-factor models (corresponding to working memory, inhibition, and cognitive flexibility) best account for children's pattern of performance across tasks (Best & Miller, 2010; Brocki & Bohlin, 2004; Huizinga et al., 2006; Lehto et al., 2003).

We propose that EF likely contributes to early scientific literacy, just as it has been shown to contribute to success in other academic domains, including language, literacy, and math (Best et al., 2011; Blair & Razza, 2007; Brock et al., 2009; Clark et al., 2014). In part, it is likely to do so by directly optimizing the processing and acquisition of relevant content knowledge. EF also might contribute indirectly to the development of scientific literacy by supporting fundamental causal reasoning skills. Gropen, Clark-Chiarelli, Hoisington, and Ehrlich (2011) offer a theoretical framework articulating this latter possibility in greater detail. Specifically, they argue that working memory supports the ability to hold causal rules or actions in mind (e.g., dropping a typical object from a tower will cause it to fall straight down), while inhibition supports the ability to revise hypotheses by comparing prior beliefs or predictions from what was observed (e.g., unlike most objects, dropping a feather from a tower will cause it to float along a circuitous trajectory) (Gropen et al., 2011). It is not hard to imagine how cognitive flexibility might also contribute by supporting children's ability to consider multiple possible courses of action and potential outcomes in the process of evaluating causal contingencies.

Consistent with Gropen et al.'s (2011) proposal, relationships between EF and the general reasoning abilities of older children and young adults have been documented repeatedly in the literature (Handley, Capon, Beveridge, Dennis, & Evans 2004; McCormack, Simms, McGourty, & Beckers, 2013; Simoneau & Markovits, 2003; Markovits & Doyon, 2004). Rhodes et al. (2016) also recently reported that working memory significantly predicted gains in conceptual knowledge of chemistry (but not simpler factual knowledge). This suggests that EF skills may scaffold students' ability to adopt, apply, and extend learned causal principles to novel academic domains. But like Tolmie et al.'s (2016) proposal regarding the relationship between causal reasoning and scientific literacy, Gropen et al.'s (2011) incorporation of EF into their theoretical framework was developed with school-aged children in mind. Indeed, although the role of EF in facilitating academic achievement broadly speaking is well established for this older age group, little research has been conducted in younger children, especially with respect to science.

1.4. The current study

The goal of the current study is to begin specifying candidate foundations of scientific literacy in children. Specifically, we evaluate whether causal reasoning and EF are related to young children's scientific literacy. With respect to the latter, we further evaluate whether EF might relate to early scientific literacy not only directly, but also indirectly through its relationship with causal reasoning. In addition to the theory and evidence outlined above, a small number of studies with preschoolers support affirmative predictions. For example, Nayfeld et al. (2013) recently reported that EF predicted gains in low income preschooler's academic achievement in science, and that this effect outstripped those observed for math or reading literacy. Other relevant work highlights early relationships between different aspects of EF and causal (specifically, counterfactual) reasoning (e.g., Guajardo, Parker & Turley-Ames, 2009; Beck, Riggs, & Gorniak, 2009; Riggs & Beck, 2007). Importantly, in the current work, we incorporate measures of causal reasoning, EF and scientific literacy into a unitary design in order to gain a more complete picture of the relationship between these key skills in young children.

2. Methods

2.1. Participants

Data for this study were collected as part of a larger longitudinal study investigating the development of children's interest in causal information. Our study sample included 125 children (65 = female) from a major southern metropolitan area. Participating children were three years of age at the first session (M = 41.04 months; SD = 3.12 months, range = 36.12–46.92 months).

Children did not have any diagnosed developmental delays or disorders, and they understood English "well" or "very well" as reported by a parent. The sample was racially, ethnically, and socioeconomically representative of our recruitment area. Based on parent report, 10.4% of participating children were African American, 70.4% were White, 3.2% were Asian or Pacific Islander, and 16% identified as having a mixed racial background. In addition, 31.2% of these children (primarily from the White, mixed racial categories) were also identified by their parents as Hispanic or Latino. With respect to maternal education, 1.6% of mothers reported not completing high school, 17.6% held a high school degree, 8% completed some college or additional training beyond high school, 44% had a four-year bachelor's degree, and 28.8% held a master's degree or higher.

2.2. Procedure

Data for this study were collected over three sessions lasting approximately 45–60 min each. The first session took place at a local children's museum. The second and third sessions took place in our laboratory. Although the precise timing between sessions varied with participant schedules, all were completed within a 6-month window. At the first session, parent consent was obtained and children were assessed on their receptive language (NIH-Toolbox Picture Vocabulary Test). Most children were also tested on their scientific literacy (Lens on Science) during this session. However, due to technical difficulties, some (n = 25) had to be retested at a later date. Parents also completed a demographic interview as part of the larger longitudinal study that included questions concerning parental education, race and ethnicity, as well as household composition, income, and literacy environment. Children completed EF and causal reasoning measures during the two subsequent laboratory sessions, which were scheduled an average of 38 days (*sd* = 71) after initial enrollment. Children were tested in a colorfully decorated room at a child-sized Table Sessions were audio-visually recorded for offline coding of participant responses and verification of protocol fidelity.

2.3. Measures

2.3.1. Measuring scientific literacy

The Lens on Science was presented on a 13.9″ touch screen Lenovo Yoga laptop. During the task, children were adaptively presented with 35–40 questions (from a bank of 499 possibilities) over the course of approximately 15 min. The Lens on Science targets all elements of early scientific literacy as currently conceived (NGSS, 2013; NRC, 2012). Specifically, it goes beyond factual disciplinary knowledge (of earth and space, life, physics and energy, and engineering and technology) and taps into both scientific practices and crosscutting concepts as defined by the National Research Council's framework (NRC, 2012). When the Lens on Science is administered, an algorithm selects an equal proportion of items from each content area and question type (detailed below). The question bank items were calibrated using the dichotomous Rasch model, and scaled to have a mean item difficulty of zero. As reported by Greenfield (2015), the outcome variable is a continuous item response theory (IRT) ability score ranging from −3 to 3. Based on a sample of 1753 students, the average standard error of the Rasch ability estimate was 0.31, which corresponds to a reliability of 0.87. The Lens on Science also correlates highly with vocabulary, $r = 0.55$ (Greenfield et al., 2009).

Prior to the administration of the Lens on Science, children were shown an instructional video and completed a readiness screening to ensure they could follow the instructions for three different question formats used in the assessment. The first question type required children to select their answers by touching one of three pictured alternatives. For example, one item shows a picture of a girl examining a flower and asks the child to point to which of three tools (a spatula, magnifying glass, and ruler) the girl should use to see the flower's tiny parts. An example of the second question format displays a large picture of a sea turtle. The child is told, "This is a turtle. Touch the part of the turtle that helps it swim." Boxes then appear around the head, arm, and shell of the turtle and the child must touch the box around the arm of the turtle to score correctly. The third question format displays several pictures on the screen and requires that the child choose multiple answers. For example, a child is shown six pictures of different shells. The child is then told, "Some of these shells have the same shape, and some do not. Touch all the shells that have the same shape. When you are done, touch the gold star." Boxes then appear around the six shells, specifying the potential choices.

2.3.2. Measuring causal reasoning

Causal reasoning was assessed with three measures that tapped cause-effect association, causal inference, or counterfactual reasoning.

2.3.2.1. Tracking cause-effect associations. Four boxes ("blicket detectors") measuring $8 \times 4 \times 4$ inches were constructed out of cardboard (Gopnik & Sobel, 2000; Gopnik et al., 2001). The first box was painted blue and had a gear toy affixed to the top that spun when the box was activated. The second box, painted yellow with blue polka dots, produced a continuous playful noise when activated. The third box was painted green and had a laminated yellow cartoon cat figure mounted on the front. When activated, a small light bulb inside the box illuminated, making it appear as though the cat's nose lit up. The fourth box was painted silver and had a puppy cartoon on the front. When activated, a motor within the box slowly turned the puppy back and forth.

The experimenter introduced the task to the children by saying, "I have some special boxes with me! Some blocks make the boxes go, and some blocks don't do anything to the boxes. They don't turn them on at all! Do you think you can help me figure out how the boxes work?"

A single block was then introduced with a unique novel name and was placed on the box to no effect (e.g., "This one is a duffin. This one doesn't do anything to our boxes."). The first block ("duffin") was then removed and the second block (e.g., a "buttle") was introduced and then placed on the box, which led to its activation. Finally, while the box was still activated, the experimenter also placed the non-causal block (the "duffin") back on the box while simultaneously saying, "let's put both on together!" The box remained active while both blocks were on the box. The experimenter then asked the child, "Now I need your help! How can we make the box stop?" In order to respond correctly, children had to keep track of which object was more reliably associated with activation of the box and remove it. If a response was not immediately forthcoming, children were further prompted (e.g., "Can you tell me which one?" or "Which one do you think?") until they offered an answer. This same procedure was repeated for the remaining three trials. Children were assigned a score of 1 if they chose the correct block and a score of 0 if they chose the incorrect block on each of four trials. Although the experimenter discouraged children from choosing both blocks by holding both blocks down with their hands and asking the child to choose just one, children did not always comply, therefore leading to the occasional uncodable trial. In these rare circumstances, we excluded the trial from analysis.

2.3.2.2. Generating causal inferences. This task, modeled after previous studies (Das Gupta & Bryant, 1989) involved eight trials (see Table 1). On each trial, children were shown a timeline of photographs in which objects underwent changes (e.g., a broken flowerpot next to a whole flowerpot). The experimenter first described the pictures in general terms ("First it looked like Tatziki *this*, then I did something to it, and now it looks like *this*."). Then, children were shown photographs of four possible instruments (e.g., hammer, light bulb, paintbrush, glue) and were asked to choose the one that caused the change. The outcome variable is the total number of trials answered correctly out of eight.

2.3.2.3. Counterfactual reasoning. In this task, modeled after Guajardo & Turley-Ames (2004), children were presented with four vignettes and asked to reason counterfactually about each situation. For example, in one story, children were asked to imagine that they are playing outside in a muddy yard. Then they imagine

Table 1
Stimuli used in the causal inference task.

Stimuli	Choices
Pre → post pictures	
Tomato → sliced tomato	Whisk, spatula, measuring cups, knife
Sweater with hole → sewn sweater	Teapot, roller-skate, mug, needle and thread
Test items (forwards)	
Spilled dirt → swept dirt	Chair, tissue box, clock, broom
Raw egg → cooked egg	Blender, drying rack, napkin holder, stove
Messy hair → brushed hair	Sponge, toothbrush, rolling pin, hairbrush
Torn paper → taped paper	Keys, toy blocks, crayons, tape
Test items (backwards)	
Wet pan → dry pan	Sink, microwave, calculator, tape
Chalkboard with writing → erased chalkboard	Chalk, scissors, stapler, eraser
Broken flowerpot → glued flowerpot	Hammer, light bulb, paintbrush, glue
Stained shirt → clean shirt	Ketchup, purse, iron, detergent

Note: Based on Das Gupta and Bryant (1989).

that they get really thirsty, so they go inside their house for a drink and dirty the clean kitchen floor in the process. Children were then asked, “What could you have done so the floor would *not* have gotten dirty?” In order to respond correctly, children had to mentally represent what would have had to have happened in the past to produce a different outcome. Children were assigned a score of 1 if they generated a counterfactual statement (e.g., take off my shoes) on each of four trials. They were otherwise assigned a score 0 for each trial. The outcome variable is therefore a score with a range of 0–4.

2.3.3. *Measuring executive function*

EF was assessed using measures of inhibitory control, working memory, and cognitive flexibility. However, because previous work suggests that EF is best represented by a unitary factor structure in preschool aged children (e.g., Garon, Bryson & Smith, 2008), we created a composite EF score for each child by standardizing and averaging across the three measures (for a possible score ranging from −3 to 3).

2.3.3.1. Inhibitory control. The NIH-Toolbox Flanker Inhibitory Control and Attention Test (NIH-TB Flanker; Zelazo et al., 2013) was used to evaluate children's inhibitory control of visual attention (Gershon et al., 2013). During each trial of this task, which was presented on an iPad, a central target fish was flanked by similar stimuli on the left and right. Children were instructed to indicate the direction of the central stimulus by pressing a left or right arrow button on the screen. On congruent trials, the flanker fish faced the same direction as the target fish. On incongruent trials, they faced the opposite direction, requiring the child to inhibit their attention to the flanker fish. The NIH-TB Flanker task was normed on a sample of 174 children. Zelazo et al. (2013) reported the NIH-TB Flanker positively correlated with the WPPSI-III Block Design in 3–6 year olds ($r = 0.60$) and D-KEFS Inhibition raw scores in 8–15 year olds ($r = 0.34$). The outcome variable reported is a standard score based on either accuracy or, if a child achieved greater than 80% accuracy, a combination of accuracy and reaction time ($M = 100$, $SD = 15$).

2.3.3.2. Working memory. The count and label task was used to assess working memory, or the ability to hold and manipulate information in one's mind (Baddeley & Hitch, 1994; Gordon & Olson, 1998). We chose this particular test because it is one of few options available that is developmentally appropriate for very young children (Carlson, 2005). In this task, children were shown three objects (e.g., a key, a cup, and a toy dog) and asked to label them. Then the experimenter suggested they count the objects (e.g., “one”, “two”, “three”). Then they demonstrated how to count the objects and label them each in turn (e.g., “one is a key, two is a cup, three is a dog”) and asked the child to try. Children were assigned either a score of either 0 (for incorrect responses) or a score of 1 (for correct responses) on each of two trials. The outcome variable is a raw score ranging from 0 to 2.

2.3.3.3. Cognitive flexibility. The NIH-TB Dimensional Change Card Sort Test (NIH-TB DCCS; Zelazo et al., 2013) was used to evaluate cognitive flexibility, or the ability to adjust to new tasks and demands (Deák & Bauer, 1995; Diamond, 2013; Gershon et al., 2013; Zelazo, Frye, & Rapus, 1996). On each trial of the NIH-TB DCCS, a target visual stimulus must be matched to one of two stimuli according to shape or color. Children first received a block of trials in which one dimension (e.g., shape) was relevant to this decision, and then a second block (switch) in which the other dimension (e.g., color) was critical. Those who succeeded following the switch also received a mixed block, in which shape was relevant on most trials, with occasional and unpredictable shifts to color. The relevant criterion word, “color” or “shape,” was simultaneously presented in visual and auditory form. This task was normed on a sample of 166 children, and Zelazo et al. (2013) reported significant correlations with the WPPSI-III Block Design in 3–6 year olds ($r = 0.69$), and D-KEFS Inhibition raw scores in 8–15 year olds ($r = 0.64$). Similar to the NIH-TB Flanker task, if children achieve 80% accuracy or higher, the algorithm weights accuracy and reaction time in the standard score ($M = 100$, $SD = 15$).

2.3.4. *Measuring receptive vocabulary*

The NIH-TB Picture Vocabulary Test (NIH-TB PVT; Gershon et al., 2013) is a measure of receptive vocabulary administered in a computerized adaptive format on an iPad. During this task, children were presented with an audio recording of a word while four pictures simultaneously appeared on the screen. On each trial, children were asked to select the picture that most closely matched the meaning of the word that was said. The child was presented with two practice trials with feedback, followed by 25 test trials. Total administration time for the adaptive test is approximately 5 min. Initial item-level calibration was conducted online with 3190 children ages 3–17 (Gershon et al., 2013). The version of the NIH-PVT used in this study was normed on a sample of 120 3–6-year-old children and correlates strongly ($r = 0.9$, $p < 0.001$) with the Peabody Picture Vocabulary Test 4th Edition (PPVT-IV) (Gershon et al., 2013; Weintraub et al., 2013). As is the case for the other NIH-TB tasks, the NIH-PVT also yields a standard score ($M = 100$, $SD = 15$).

2.4. *Coding*

Trained researchers scored the cause-effect association, causal inference, and counterfactual reasoning tasks offline using video recordings. For both the cause-effect association and causal inference tasks, a second coder assessed all files to ensure reliability of coding. For the counterfactual reasoning task, each video was initially transcribed and coded by a blind researcher, then a second researcher coded the transcript. If the two coders disagreed, a third researcher coded the video and transcript. All disagreements were discussed and resolved. The coders also assessed the videos for fidelity to protocol on the bases of general procedures

Table 2
Rates of missingness for each study variable.

Variable	% Missing
Lens on Science	4.8
NIH-TB PVT	0.8
Count and label	13.6
NIH-TB Flanker	8.8
NIH-TB DCCS	23.2
Cause-effect association	5.6
Causal inference	7.2
CFR	14.2

Note: NIH-TB = NIH Toolbox, PVT = Picture Vocabulary Test, DCCS = Dimensional Change Card Sort, CFR = counterfactual reasoning.

Table 3
Descriptive statistics for imputed data (N = 125).

Variables	Mean	SD	Min	Max
Age	3.42	0.26	3.01	3.91
Lens on Science	0.22	1.14	−1.40	3.00
NIH-TB PVT	64.33	9.36	54	88
Count and label	0.55	0.81	0	2
NIH-TB Flanker	37.74	11.90	25	64
NIH-TB DCCS	47.08	13.62	34	73
Cause-effect association	0.65[a]	0.35	0	1
Causal inference	4.86	1.71	0	8
CFR	0.98	1.17	0	4

Note: NIH-TB = NIH Toolbox, PVT = Picture Vocabulary Test, DCCS = Dimensional Change Card Sort, CFR = counterfactual reasoning.
[a] Cause-effect association score is proportion correct.

and adherence to a script (if applicable). Coders agreed 100% of the time on all measures. All four measures from the NIH-TB were automatically scored. Participant data were coded and managed using REDCap (Research Electronic Data Capture) (Harris et al., 2009), a secure, web-based application designed to support data capture for research studies.

3. Results

3.1. Missing data

Rates of missing data ranged from less than 1% to 23.2% (see Table 2) across tasks. Individual scores were excluded due to poor attention as rated by another researcher on a 1–5 Likert scale. We excluded data from individual tasks completed by a total of 16 children because they received an attention score of a 1 or 2: Lens on Science ($n = 1$), count and label ($n = 9$), NIH-TB Flanker ($n = 2$), causal inference ($n = 2$), and counterfactual reasoning ($n = 2$). Twelve children also failed to pass the training trials on either the NIH-Flanker ($n = 4$) or the NIH-DCCS ($n = 8$). An additional ten children failed to complete either the NIH-Flanker (n = 4) or the NIH-DCCS ($n = 6$). Additional individual task data were lost to experimenter error (NIH-PVT: $n = 1$, cause-effect association: $n = 2$, CFR: $n = 1$), and attrition ($n = 17$).

All data analyses were conducted using RStudio (R version 1.0.136; RStudio Team, 2016; R Core Team, 2016). To handle missing data, we employed multiple imputation (MI) using the "mice" package that runs in RStudio (Brand, 1999; van Buuren & Groothuis-Oudshoorn, 2011). MI (Brand, 1999; Rubin & Schenker, 1986; Schafer, 1997) allowed us to retain as many subjects as possible by replacing missing values with unbiased estimates based on the observed data (Little & Rubin, 2002). We imputed ten datasets with 30 iterations based on recommended procedures (Bodner, 2008; White, Royston, & Wood, 2011). We then used the imputed datasets to analyze and pool the results.

3.2. Descriptive statistics

The means, standard deviations, and ranges for age, vocabulary, EF, and causal reasoning measures are displayed in Table 3. Simple bivariate correlations between these variables of interest are presented in Table 4. Please recall that, because children occasionally did not provide a clear response on a single trial during the cause-effect association task, we use proportion correct (ranging from 0 to 1) as our dependent variable for this task. Because sex did not correlate with any of the tasks, it is not included in this table and was not considered in further analyses. Children's receptive vocabulary was significantly correlated with scientific literacy, thereby necessitating that we control for its influence in our analyses. Correlations between our key measures of interest (i.e., EF and causal reasoning) and children's emergent scientific literacy were also significant. The only key variable that did not correlate with scientific literacy was cause-effect association. Performance on this task also failed to correlate with our other measures of causal reasoning (i.e., causal inference and counterfactual reasoning).

3.3. Hierarchical regression analyses

Hierarchical regressions were used to determine whether causal reasoning or EF accounted for significant variance in children's early scientific literacy (see Table 5). Child age, ethnicity (Hispanic or Latino vs. Non-Hispanic or Latino), vocabulary, level of maternal education (No High School diploma, High School diploma, Technical or Associates degree, Bachelor's degree, Advanced degree), and the average delay window between initial and subsequent testing sessions were entered as covariates for all analyses.

Because we hypothesized that EF might contribute to the development of causal reasoning skills, we entered the composite EF score alone in the first block. Results revealed that EF was significantly related to scientific literacy, $B = 0.27$, $t(99.53) = 2.30$, $p = 0.02$,

Table 4
Correlations between study variables for imputed data (N = 125).

Variables	1	2	3	4	5	6	7	8	9
1. Child gender	–								
2. Child age	0.03	–							
3. Maternal education	−0.04	0.02	–						
4. Child ethnicity	0.04	0.05	−0.12	–					
5. Lens on Science	−0.16	0.35***	0.17	−0.3**	–				
6. NIH-PVT	−0.01	0.21*	0.17	−0.2*	0.49***	–			
7. EF composite score	−0.01	0.32***	0.12	−0.08	0.39***	0.22*	–		
8. Cause-effect association	0.12	0.08	−0.14	−0.02	−0.07	−0.16	0.06	–	
9. Causal inference	−0.05	0.28**	0.01	−0.11	0.49***	0.37***	0.35***	0.02	–
10. CFR	−0.03	0.38***	0.08	0.14	0.33***	0.15	0.27**	−0.21*	0.24**

Note: Simple correlation coefficients are shown. NIH-TB PVT = NIH-Toolbox Picture Vocabulary Test, CFR = counterfactual reasoning.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Table 5
Hierarchical regression results predicting scientific literacy performance (N = 125).

	Block 1		Block 2	
Variable	*B*	*SE B*	*B*	*SE B*
EF composite score	0.27**	0.11	0.25*	0.14
Cause-effect association			−0.15	0.24
Causal inference			0.15**	0.05
Counterfactual reasoning			0.06	0.07
R^2	0.42		0.48	
ΔR^2	*n/a*		0.06	

Note: $^*p < 0.05$; $^{***}p < 0.001$.
** $p < 0.01$.

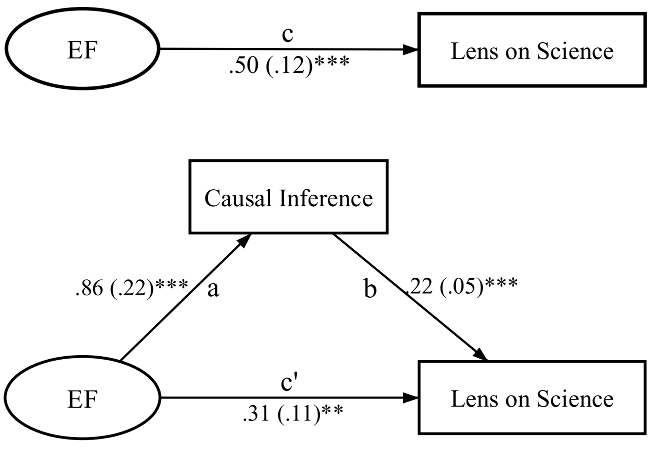

Fig. 1. Model for working causal inference as a mediator between executive function and scientific literacy (N = 125).
Note. Unstandardized coefficients are shown; standard errors are in parentheses. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

accounting for 2.89% unique variance in Lens on Science performance, over and above our covariates.

All three measures of causal reasoning were entered simultaneously in the second block. Considered individually, the cause-effect association and counterfactual reasoning tasks were not significant contributors, but causal inference was, $B = -0.15$, $t(95.41) = -0.60$, $p = 0.55$, $B = -0.06$, $t(105.75) = -0.94$, $p = 0.35$, and $B = 0.15$, $t(52.98) = 3.01$, $p = 0.004$, respectively. Overall, causal reasoning accounted for 6.05% additional variance above and beyond EF (and the covariates). A Wald test confirmed the additional variance accounted for by causal reasoning in block two was significant compared to block one, Wald $F(3, 1126.28) = 3.69$, $p = 0.01$.

3.4. Mediation analysis

Interestingly, when the second block of causal reasoning measures was added into the hierarchical regression model described above, EF was reduced to a statistcially non-significant level, $B = 0.25$, $t(104.54) = -1.74$, $p = 0.08$. To test whether this might be explained by potential contributions of EF to causal reasoning ability, we conducted an exploratory mediation analysis using the "lavaan" structural equation modeling (SEM) and the "mice" and "mitools" imputed data packages in RStudio (Lumley, 2010; Rosseel, 2012; van Buuren & Groothuis-Oudshoorn, 2011). The mediation model estimated the effect of the independent variable of children's EF skills on the mediator, causal reasoning (Path a), the effect of the mediator on children's scientific literacy (Path b), and the effect of EF on scientific literacy (Path c′). Finally, the original total observed effect of EF on scientific literacy is represented by path c (see Fig. 1 for reference).

We selected causal inferencing ability as the mediating variable for this analysis because the results from our regression models indicated that it explained the most variance in scientific literacy. The mediation analysis confirmed that EF was significantly related to both causal inference performance, $B = 0.86$, $SD = 0.22$, $p < 0.001$ (Path a), and scientific literacy, $B = 0.5$, $SD = 0.12$, $p < 0.001$ (Path c). Children who had greater EF composite scores performed better on our measures of both causal inference and scientific literacy. Moreover, causal inference was a significant positive associate of scientific literacy, $B = 0.22$, $SD = 0.05$, $p < 0.001$ (Path b). To test if causal inference mediated the relationship between EF and scientific literacy, we calculated the indirect effect of causal inference ability. The direct association between EF and scientific literacy was reduced, $B = 0.31$, $SD = 0.11$, $p = 0.005$ (path c′), and the indirect effect was significant after controlling for causal inference ability, $B = 0.19$ $SD = 0.06$, $p = 0.002$, suggesting causal inference ability is indeed a significant partial mediator.

4. Discussion

In this study, we examined whether causal reasoning and EF have the potential to support the emergence of scientific literacy. As expected, both sets of skills accounted for unique variance in preschoolers' scores on a standardized measure of scientific content, crosscutting concepts, and practice knowledge. Of the specific measures of causal reasoning evaluated in our models, causal inference ability emerged as a significant associate of scientific literacy. Although this task required some general knowledge about objects (e.g., scissors, glue), its relationship to scientific literacy held after controlling for vocabulary. We therefore suggest that children's ability to access, flexibly consider, and apply their knowledge to new contexts might better account for these relationships.

Notably, despite replicating previously reported levels of performance on our measure of cause-and-effect associations (e.g., Gopnik et al., 2001), this task was not associated with scientific literacy in any of our analyses. Neither did it positively correlate with our other measures of causal reasoning (i.e., causal inference or counterfactual reasoning). This is surprising given "blicket detector" tasks like this are often highlighted as a classic demonstration of children's early causal reasoning abilities (e.g., Gopnik & Sobel, 2000; Kloos & Sloutsky, 2005; Sobel & Legare, 2014). One reason for the dissociation observed here might be that the causal association task merely requires children to isolate arbitrary relationships between readily observable physical objects and events, while our causal inference and counterfactual reasoning tasks require that children draw on specific mechanisms to reason about meaningful causal systems. The current results suggest that the ability to detect cause-effect relations on the basis of arbitrary spatiotemporal contingencies might not be integral to emerging scientific literacy. Rather, higher-level causal reasoning skills (i.e., causal inference) appear to be stronger candidates. This idea is also consistent with Tolmie et al.'s (2016) model of scientific thinking in school age children, in which practicing higher-level causal reasoning skills like these is essential for scaffolding core scientific practices (e.g., hypothesis generation and testing).

The current results, however, also highlight the importance of keeping more fundamental cognitive processes like EF in mind when considering the potential origins of scientific literacy. Working memory and inhibition might be particularly important for helping young children to focus on, and keep track of, relevant information in the context of self-guided exploration and play, as well as in structured science activities orchestrated by parents or teachers. In this way, children with strong EF skills might more efficiently acquire scientific content knowledge.

Recall, however, that the relationship between EF and scientific literacy was partially mediated by causal inference, suggesting the possibility of indirect effects as well. Gropen et al.'s (2011) theory regarding the foundations of school-age scientific literacy might therefore be relevant to understanding earlier development as well. As argued therein, working memory contributes to causal reasoning by allowing us to bring together incoming perceptual information and past knowledge to resolve, and think creatively about, novel causal systems (Byrne, 2007). At the same time, inhibition is thought to play a critical role by helping to focus attention on relevant dimensions of the causal system and by suppressing, and allowing the revision of, prior beliefs (Gopnik et al., 2001; Legare, Gelman, & Wellman, 2010; Schulz & Bonawitz, 2007; Wilkening & Sodian, 2005).

As these possibilities are explored in future research, it will be important to keep in mind that major developmental changes in EF are just beginning to unfold in the young population studied here (Anderson & Reidy, 2012; Carlson, 2005; Garon et al., 2008; Nelson, James, Chevalier, Clark, & Espy, 2016). It is an open question whether EF, or its increasingly differentiable components, will remain key associates of causal reasoning and scientific literacy in older children. It will also be important to keep in mind the possibility of reciprocal influences. In particular, as highlighted by Nayfeld et al. (2013), practice applying the complex problem-solving skills commonly encountered in scientific activities might well exercise and strengthen both causal reasoning and EF capabilities. Longitudinal and intervention work will be necessary to clarify these relationships.

4.1. Limitations

Although this study advances the field by highlighting two potential contributors to emerging scientific literacy in preschoolers, there are clearly some limitations, including the notable amount of missing data. We addressed this issue to the best of our ability by imputing complete and unbiased datasets, but replication will be important for confirming the resulting analyses. Replication will also be helpful in addressing limitations of the testing schedule implemented in this first iteration. Because the current study was conducted as a supplement to an ongoing longitudinal investigation, the scheduling of experimental sessions was significantly constrained, leading to lengthier and more variable data collection windows than was ideal.

Another limitation to the current work was that, because our participants were so young, the developmentally appropriate options available to us for testing their EF and causal reasoning skills were quite limited. As a result, some of our measures (e.g., counterfactual reasoning and count and label) produced less than ideal score ranges and observed distributions, and contributed to the aforementioned missing data. Future studies with older children will be crucial to further specify the relationship between more complex, and potentially more sensitive, measures of causal reasoning, EF, and scientific literacy. With this goal in mind, we are currently collecting data from a cross-sectional sample of 5-year-olds. We also intend to follow the 3-year-olds tested here longitudinally through preschool.

This approach will also allow us to address yet another limitation of the current work, which is that it took only a static snapshot of EF and causal reasoning skills at an age associated with the onset of rapid developmental change. Studies comparing relationships among EF, causal reasoning, and scientific literacy throughout this transition will be essential to fully specify the extent to which these aspects of cognitive development are intertwined. Finally, the current study, and those in progress, are also limited by their correlational design. To the extent that EF and causal reasoning are malleable skills, intervention studies targeting these skills will be helpful in more definitively testing whether they have a causal influence on the development of scientific literacy.

5. Conclusion

In closing, this study further adds to a growing body of research on relationships between domain-general cognitive skills and academic achievement by documenting associations between both causal reasoning and EF and scientific literacy in young children. In this way, the current work is entirely consistent with Snow's (2007) suggestion that children's school readiness is likely dependent upon interactions between sets of skills across multiple domains. Longitudinal and intervention studies will of course be essential for establishing whether the relationships observed in the current study are causal in nature. In the long run, our hope is that this line of investigation might ultimately have practical relevance for fostering early scientific literacy in young children. As noted in the introduction, preschools have been slow to adopt science-focused curricula, and have failed to promote widespread success in this domain. Moreover, significant achievement gaps in STEM fields are early emerging and persistent (Morgan, Farkas, Hillemeier, & Maczuga, 2016). Thus, translational research in this area is critical. To the extent that we can specify the earliest foundations of scientific literacy, the better equipped we will be to design targeted and efficient interventions.

Conflicts of interest

None.

Acknowledgements

This research was generously supported by a grant from the National Science Foundation [Grant number 1535102] to the second author. The opinions expressed are those of the authors and do not represent views of the funding agencies. The authors thank the research assistants and participating families who made this research possible.

References

Alvarez, A. L., & Booth, A. E. (2014). Motivated by meaning: Testing the effect of knowledge-infused rewards on preschoolers' persistence. *Child Development*, *85*(2), 783–791. http://dx.doi.org/10.1111/cdev.12151

Anderson, P., & Reidy, N. (2012). Assessing executive function in preschoolers. *Neuropsychology Review*, *22*(4), 345–360. http://dx.doi.org/10.1007/s11065-012-9220-3

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology*, *8*(4), 485–493. http://dx.doi.org/10.1037/0894-4105.8.4.485

Beck, S. R., Riggs, K. J., & Gorniak, S. L. (2009). Relating developments in children's counterfactual thinking and executive functions. *Thinking & Reasoning*, *15*(4), 337–354.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*(6), 1641–1660. http://dx.doi.org/10.1111/j.1467-8624.2010.01499.x

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, *21*(4), 327–336. http://dx.doi.org/10.1016/j.lindif.2011.01.007

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663. http://dx.doi.org/10.1111/j.1467-8624.2007.01019.x

Bodner, T. E. (2008). What improves with increased missing data imputations? *Structural Equation Modeling*, *15*(4), 651–675. http://dx.doi.org/10.1080/10705510802339072

Brand, J. P. L. (1999). *Development, implementation and evaluation of multiple imputation strategies for the statistical analysis of incomplete data sets. Unpublished doctoral dissertation*. Rotterdam, The Netherlands: Erasmus University.

Brenneman, K. (2011). Assessment for preschool science learning and learning environments. *Early Childhood Research and Practice*, *13*(1).

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of 'hot' and 'cool' executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly*, *24*(3), 337–349. http://dx.doi.org/10.1016/j.ecresq.2009.06.001

Byrne, R. M. (2007). *The rational imagination: How people create alternatives to reality*. MIT Press.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology*, *28*(2), 595–616.

Clark, C. A. C., Nelson, J. M., Garza, J., Sheffield, T. D., Wiebe, S. A., & Espy, K. A. (2014). Gaining control: Changing relations between executive control and processing speed and their relevance for mathematics achievement over the course of the preschool period. *Frontiers in Psychology*, *5*, 1–15. http://dx.doi.org/10.3389/fpsyg.2014.00107

Das Gupta, P., & Bryant, P. E. (1989). Young children's causal inferences. *Child Development*, *60*(5), 1138–1146. http://dx.doi.org/10.2307/1130788

Deák, G., & Bauer, P. J. (1995). The effects of task comprehension on preschoolers' and adults' categorization choices. *Journal of Experimental Child Psychology*, *60*(3), 393–427. http://dx.doi.org/10.1006/jecp.1995.1047

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. http://dx.doi.org/10.1146/annurev-psych-113011-143750

Drayton, S., Turley-Ames, K. J., & Guajardo, N. R. (2011). Counterfactual thinking and false belief: The role of executive function. *Journal of Experimental Child Psychology*, *108*(3), 532–548. http://dx.doi.org/10.1016/j.jecp.2010.09.007

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Taking science to school. Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), 31–60. http://dx.doi.org/10.1037/0033-2909.134.1.31

Gelman, R., Bullock, M., & Meck, E. (1980). Preschoolers' understanding of simple object transformations. *Child Development*, *51*(3), 691–699. http://dx.doi.org/10.2307/1129454

Gershon, R. C., Slotkin, J., Manly, J. J., Blitz, D. L., Beaumont, J. L., Schnipke, D., . . . & Weintraub, S. (2013). IV. NIH Toolbox Cognition Battery (CB): Measuring language (vocabulary comprehension and reading decoding). *Monographs of the Society for Research in Child Development*, *78*(4), 49–69. http://dx.doi.org/10.1111/mono.12034

Gopnik, A., & Schulz, L. E. (2007). *Causal learning: Psychology, philosophy, and computation*. New York: Oxford University Press.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*(5), 1205–1222. http://dx.doi.org/10.1111/1467-8624.00224

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two, three, and four year olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*(5), 620–629. http://dx.doi.org/10.1037/0012-1649.37.5.620

Gopnik, A. (2000). Explanation as orgasm and the drive for causal knowledge: The function, evolution, and phenomenology of the theory formation system. In F. C. Keil, & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 299–323). Cambridge, MA: MIT Press.

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, *337*(6102), 1623–1627. http://dx.doi.org/10.1126/science.1223416

Gordon, A. C. L., & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology*, *68*(1), 70–83. http://dx.doi.org/10.1006/jecp.1997.2423

Greenfield, D. B., Jirout, J., Dominguez, X., Greenberg, A., Maier, M., & Fuccillo, J. (2009). Science in the preschool classroom: A programmatic research agenda to improve science readiness. *Early Education and Development*, *20*(2), 238–264. http://dx.doi.org/10.1080/10409280802595441

Greenfield, D. B. (2015). Assessment in early childhood science education. In K. Trundle, & M. Saçkes (Eds.), *Research in early childhood science education* (pp. 353–380). New York, NY: Springer Publishing.

Greif, M. L., Kemler Nelson, D. G., Keil, F. C., & Gutierrez, F. (2006). What do children want to know about animals and artifacts? Domain-specific requests for information. *Psychological Science*, *17*(6), 455–459. http://dx.doi.org/10.1111/j.1467-9280.2006.01727.x

Gropen, J., Clark-Chiarelli, N., Hoisington, C., & Ehrlich, S. B. (2011). The importance of executive function in early science education. *Child Development Perspectives*, *5*(4), 298–304. http://dx.doi.org/10.1111/j.1750-8606.2011.00201.x

Guajardo, N. R., Parker, J., & Turley-Ames, K. (2009). Associations among false belief understanding, counterfactual reasoning, and executive function. *British Journal of Developmental Psychology*, *27*(3), 681–702.

Guajardo, N. R., & Turley-Ames, K. J. (2004). Preschoolers' generation of different types of counterfactual statements and theory of mind understanding. *Cognitive Development*, *19*(1), 53–80. http://dx.doi.org/10.1016/j.cogdev.2003.09.002

Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning*, *10*(2), 175–195.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) — A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–381. http://dx.doi.org/10.1016/j.jbi.2008.08.010

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017–2036. http://dx.doi.org/10.1016/j.neuropsychologia.2006.01.010

Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational interventions to advance children's scientific thinking. *Science*, *333*(6045), 971–975. http://dx.doi.org/10.1126/science.1204528

Klahr, D. (2005). Early science instruction: Addressing fundamental issues. *Psychological Science*, *16*(11), 871–873. http://dx.doi.org/10.1111/j.1467-9280.2005.01629.x

Kloos, H., & Sloutsky, V. M. (2005). Do preschoolers understand causality? A critical look. In B. G. Bara, L. Basalou, & M. Bucciarelli (Eds.), *Proceedings fo the XXVII Annual Conference of the Cognitive Science Society* (pp. 1154–1159). Mahwah, NJ: Erlbaum.

Kushnir, T., Gopnik, A., Schulz, L., & Danks, D. (2003). Inferring hidden causes. In R. Alterman, & D. Kirsh (Eds.), *Proceedings of the 25th annual conference of the Cognitive Science Society* (pp. 699–703). Boston, MA: Cognitive Science Society.

Legare, C. H., Gelman, S. A., & Wellman, H. M. (2010). Inconsistency with prior knowledge triggers children's causal explanatory reasoning. *Child Development*, *81*(3), 929–944.

Lehto, J. E., Juujarvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*(1), 59–80. http://dx.doi.org/10.1348/026151003321164627

Little, R. J., & Rubin, D. B. (2002). Single imputation methods. In *Statistical analysis with missing data* (second edition, pp. 59–74). http://dx.doi.org/10.1002/9781119013563

Lumley, T. (2010). *mitools: Tools for multiple imputation of missing data. R package version 2.0.1.*. Retrieved from http://CRAN.R-project.org/package=mitools

Mantzicopoulos, P., Patrick, H., & Samarapungavan, A. (2013). Science literacy in school and home contexts: Kindergarteners' science achievement and motivation. *Cognition and Instruction*, *31*(1), 62–119. http://dx.doi.org/10.1080/07370008.2012.742087

Markovits, H., & Doyon, C. (2004). Information processing and reasoning with premises that are empirically false: Interference, working memory, and processing speed. *Memory & Cognition*, *32*, 592–601.

McCormack, T., Simms, V., McGourty, J., & Beckers, T. (2013). Blocking in children's causal learning depends on working memory and reasoning abilities. *Journal of Experimental Child Psychology*, *115*(3), 562–569. http://dx.doi.org/10.1016/j.jecp.2012.11.016

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. http://dx.doi.org/10.1006/cogp.1999.0734

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher*, *45*(1), 18–35. http://dx.doi.org/10.3102/0013189X16633182

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Committee on a Conceptual Framework for New K-12 Science Education Standards*. Washington, DC: The National Academies Press.

Nayfeld, I., Brenneman, K., & Gelman, R. (2011). Science in the classroom: Finding a balance between autonomous exploration and teacher-led instruction in preschool settings. *Early Education and Development*, *22*(6), 970–988. http://dx.doi.org/10.1080/10409289.2010.507496

Nayfeld, I., Fuccillo, J., & Greenfield, D. B. (2013). Executive functions in early learning: Extending the relationship between executive functions and school readiness to science. *Learning and Individual Differences*, *26*, 81–88. http://dx.doi.org/10.1016/j.lindif.2013.04.011

Nelson, J. M., James, T. D., Chevalier, N., Clark, C. A. C., & Espy, K. A. (2016). Structure, measurement, and development of preschool executive function. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.), *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research* (pp. 65–89). Washington, DC: American Psychological Association.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at https://www.R-project.org/

RStudio Team. (2016). *RStudio: Integrated development for R*. Boston, MA: RStudio, Inc. Available at http://www.rstudio.com/

Rafetseder, E., Renate, C. V., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of nearest possible world. *Child Development*, *81*(1), 376–389. http://dx.doi.org/10.1111/j.1467-8624.2009.01401.x

Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology*, *114*(3), 389–404. http://dx.doi.org/10.1016/j.jecp.2012.10.010

Rhodes, S. M., Booth, J. N., Palmer, L. E., Blythe, R. A., Delibegovic, M., & Wheate, N. J. (2016). Executive functions predict conceptual learning of science. *British Journal of Developmental Psychology*, *34*(2), 261–275. http://dx.doi.org/10.1111/bjdp.12129

Riggs, K. J., & Beck, S. R. (2007). Thinking developmentally about counterfactual possibilities. *Behavioral and Brain Sciences*, *30*(5-6), 463.

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to

children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology*, *45*(4), 958–972.
Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/
Rubin, D. B., & Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association*, *81*, 366–374. http://dx.doi.org/10.2307/2289225
Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.
Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*(4), 1045–1050. http://dx.doi.org/10.1037/0012-1649.43.4.1045
Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, *40*(2), 162–176. http://dx.doi.org/10.1037/0012-1649.40.2.162
Simoneau, M., & Markovits, H. (2003). Reasoning with premises that are not empirically true: evidence for the role of inhibition and retrieval. *Developmental Psychology*, *39*(6), 964–975.
Snow, K. L. (2007). Integrative views of the domains of child function: Unifying school readiness. In R. C. Pianta, M. J. Cox, & K. L. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability*. Baltimore, MD: Paul H. Brookes Publishing Co.
Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology*, *42*(6), 1103–1115. http://dx.doi.org/10.1037/0012-1649.42.6.1103
Sobel, D. M., & Legare, C. H. (2014). Causal learning in children. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(4), 413–427. http://dx.doi.org/10.1002/wcs.1291
Tolmie, A. K., Ghazali, Z., & Morris, S. (2016). Children's science learning: A core skills approach. *British Journal of Educational Psychology*, *86*(3), 481–497. http://dx.doi.org/10.1111/bjep.12119
U.S. Department of Health and Human Services. (2015). *Head Start early learning outcomes framework*. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/hs/sr/approach/pdf/ohs-framework.pdf.
van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Sotware*, *45*(3), 1–67. http://dx.doi.org/10.18637/jss.v045.i03
Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, *44*(2), 575–587.
Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, *108*(3), 436–452.
Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., . . . & Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*(11 Supplement 3), S54–64. http://dx.doi.org/10.1212/WNL.0b013e3182872ded
White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–399. http://dx.doi.org/10.1002/sim.4067
Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*(1), 37–63. http://dx.doi.org/10.1016/S0885-2014(96)90027-1
Zelazo, P. D., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood: I. The development of executive function. *Monographs of the Society for Research in Child Development*, *68*(3), 1–27. http://dx.doi.org/10.1111/j.1540-5834.2003.06803002.x
Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development*, *78*(4), 16–33. http://dx.doi.org/10.1111/mono.12032
Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223. http://dx.doi.org/10.1016/j.dr.2006.12.001