AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

© 2024 American Psychological Association
ISSN: 0012-1649

2024, Vol. 60, No. 4, 729–746
https://doi.org/10.1037/dev0001672

Promoting Scientific Understanding and Conceptual Change in Young Children Using Explanations and Guidance

Vaunam P. Venkadasalam, Nicole E. Larsen, and Patricia A. Ganea
Applied Psychology and Human Development, University of Toronto

Evaluating evidence and restructuring beliefs based on anomalous evidence are fundamental aspects of scientific reasoning. These skills can be challenging for both children and adults, especially in domains where they possess inaccurate prior beliefs that can interfere with the acquisition of correct scientific information (e.g., heavier objects fall faster than light ones). Across two experiments, we examined the additive benefit of combining explanations with guided activities to promote conceptual change. In Experiment 1 ($N = 238$), 4- and 5-year-olds were randomly assigned to one of three conditions: guidance with explanations, guidance only, or baseline. The guided conditions varied only in the presence or absence of conceptual information (i.e., explanation about gravity). Pre- and posttest measures showed that children's predictions improved from both guided conditions compared to the baseline condition but did not significantly differ from each other. Experiment 2 ($N = 80$, 5-year-olds) included a delay test and assessed children's learning through the justification of their predictions. Although children's performance at the immediate posttest improved in both conditions, in the guidance only, children's performance returned to the pretest levels of understanding after the delay. Children in the guidance with explanations condition had greater understanding at posttest, retained this understanding long term, and transferred it to objects with the same weight. These findings highlight the role of explanations in aiding children's long-term learning from anomalous evidence in guided activities.

Public Significance Statement

This research shows that young children need supportive input from an adult to learn about complex science concepts for which they have incorrect beliefs (e.g., heavy objects always sink, or heavy objects fall faster than light ones). In this study, the adult provided support in the form of guidance and explanations. Guidance from an adult during the learning process helps children produce evidence that challenges their incorrect beliefs. Adults can also provide children with scientific explanations that they cannot produce on their own and thus facilitate the learning of correct scientific theories.

Keywords: conceptual change, science learning, young children

Supplemental materials: https://doi.org/10.1037/dev0001672.supp

Scientific thinking requires the ability to both differentiate between theories and coordinate theory with evidence (Kuhn & Pearsall, 2000; Zimmerman, 2007). These critical scientific skills are often influenced by the prior beliefs one holds, which can affect how evidence is perceived and interpreted (Chinn & Brewer, 2001). Young children have great difficulty evaluating evidence that conflicts with their prior beliefs (Butts et al., 1993; Chinn & Brewer, 1992, 1993; Chinn & Malhotra, 2002; Koerber et al., 2005). However, adult support, such as explanations and guidance, can clarify challenging aspects of scientific inquiry and help modify the learning process so that it matches the skill level of the learner (Lazonder & Harmsen, 2016). This study investigated whether explanations combined with guidance can help young children evaluate evidence that conflicts with prior beliefs and develop their conceptual understanding of a physical science concept.

Intuitive Theories, Conceptual Change, and Anomalous Evidence

By the time formal schooling begins, children have already developed intuitive conceptions of many aspects of the world (Lane &

Vaunam P. Venkadasalam https://orcid.org/0000-0001-5052-9461

We thank the children and families who participated in this research and the Language and Learning Lab members at the University of Toronto for their help with recruitment and coding. This research was supported by an Insight Grant from the Social Sciences and Humanities Research Council of Canada (SSHRC) to Patricia A. Ganea. Data presented in this article are publicly available at https://doi.org/10.17605/OSF.IO/MA3RT.

Vaunam P. Venkadasalam served as lead for data curation, formal analysis, visualization, writing–original draft, and writing–review and editing. Nicole E. Larsen contributed equally to data curation and served in a supporting role for formal analysis, writing–original draft, and writing–review and editing. Vaunam P. Venkadasalam and Nicole E. Larsen contributed equally to conceptualization. Patricia A. Ganea was involved in funding acquisition, supervision, writing–review and editing.

Correspondence concerning this article should be addressed to Patricia A. Ganea, Applied Psychology and Human Development, University of Toronto, 252 Bloor Street West, Office 9-276, Toronto, ON M5S 1V6, Canada. Email: patricia.ganea@utoronto.ca

Harris, 2014; Pine et al., 2001; Shtulman, 2017). Their intuitive beliefs can interfere with learning accurate scientific concepts, especially when beliefs are strongly ingrained (Gershman, 2019; Vosniadou, 2013). Children do not learn from anomalous evidence until they approach the approximate age where they begin to demonstrate the correct understanding of the target science concept. Previous research has found that 6- and 7-year-olds, but not 4- and 5-year-olds, learned to balance asymmetrical objects from observing anomalous evidence (Bonawitz et al., 2012). Evaluating and restructuring beliefs based on anomalous evidence alone is particularly difficult for young children because prior beliefs impose a strong bias on reasoning and the features of evidence they attend to and encode (Klahr & Li, 2005). Deeply entrenched beliefs are difficult to change because they are supported by evidence and, in some cases, apply across various domains (Chinn & Brewer, 1993). Although the specific mechanism responsible is not yet known, conceptual change is the process of restructuring erroneous knowledge to incorporate correct information (Shtulman & Lombrozo, 2016; Vosniadou, 2013) and learning to inhibit incorrect beliefs (Brault Foisy et al., 2015).

Early approaches to stimulating conceptual change in children included providing them with the opportunity to observe anomalous evidence–evidence that contradicts one's prior beliefs. In Posner et al.'s (1982) pioneering conceptual change model, anomalous evidence was used to induce dissatisfaction with one's beliefs and create cognitive conflict. As a result of this conflict, one should seek out or invent new theories to explain the anomalies. However, an abundance of research has demonstrated that children have great difficulty learning from anomalous evidence and instead default to their prior beliefs. For example, kindergarteners failed to overcome their beliefs about the role of weight in making objects sink and float following hands-on activities that generated anomalous evidence (Butts et al., 1993), even when their prior beliefs were evoked before demonstrations (Kloos & Van Orden, 2005). These difficulties have been found with older students as well. Even 9- to 11-year-olds failed to demonstrate conceptual change when observing anomalies across several science concepts (e.g., falling objects, electric current, combustion; Chinn & Malhotra, 2002). Finally, when children explored whether heavy objects sink faster than light ones, most 10-, 12-, and 14-year-olds tended to cling to their prior beliefs by designing experiments to confirm those beliefs (Penner & Klahr, 1996). Only a few older children successfully produced anomalies (e.g., displayed informative behaviors that adequately tested their beliefs).

The use of anomalous evidence to stimulate conceptual change can be impeded during four cognitive stages: observation, interpretation, generalization, and retention (Chinn & Malhotra, 2002). For example, people of various ages hold a common but incorrect belief that heavier objects always fall faster than lighter objects (Hast, 2014; Kavanagh & Sneider, 2006). In one study, after repeatedly observing two different weighted objects falling at the same rate, most 9- to 11-year-olds discounted (e.g., "the heavy one fell faster"), misinterpreted (e.g., "both objects have the same weight"), or reinterpreted (e.g., "the heavy one is slower because it has glue sticking out") the anomalous evidence (Chinn & Malhotra, 2002). Because of these errors in the observation and misinterpretation of anomalous evidence (Gershman, 2019), most children failed to integrate the anomalies into their knowledge—thus, preventing generalization and retention. Merely observing anomalous evidence usually leads children to ignore, neglect, or reject anomalies (Bonawitz et al., 2012; Butts et al., 1993; Chinn & Brewer, 1993; Chinn & Malhotra, 2002; Gershman, 2019; Hemmerich et al., 2016; Koslowski, 1996; Kuhn, 1989; Penner & Klahr, 1996; Potvin et al., 2015; Renken & Nunez, 2010; Zimmerman, 2007; Zimmerman & Klahr, 2018). Despite the difficulties with learning simply from anomalous evidence, prior research indicates two types of support that could aid children's learning from anomalies: providing explanations and guidance during the learning process. As discussed above, prior research indicates some limitations on relying on guidance alone for restructuring deeply entrenched beliefs (Butts et al., 1993; Chinn & Malhotra, 2002; Kloos & Van Orden, 2005; Penner & Klahr, 1996). The current research will examine whether adding explanations to guided activities promotes long-term learning from anomalous evidence.

Adult support during inquiry is a middle-ground approach that can help children learn scientific information better than traditional instruction or unstructured student-led inquiry (Furtak et al., 2012; Kuhn, 2007; Lazonder & Harmsen, 2016; Sweller et al., 2007; Vorholzer & von Aufschnaiter, 2019; Weisberg et al., 2016). Although explanations play a vital role in how evidence is evaluated (Koslowski, 1996; Sandoval et al., 2014), Chinn and Brewer (1998) argue that providing children with explanations to rationalize anomalous evidence is not sufficient for conceptual change because students need to be convinced that the scientific explanations are more useful and accurate than their own strongly ingrained conceptions. Thus, a guided activity where children are scaffolded to produce the anomalies and including conceptual explanations may facilitate better performance when learning from anomalies. The characteristics of the prior beliefs (i.e., strength, background knowledge), anomalous evidence (i.e., credibility, quality, frequency), and alternative explanation (i.e., availability and quality of explanations or theories that oppose the child's own) are all factors that impact how well children learn. Children in this research were provided with scientific explanations by a knowledgeable adult, in the context of a guided activity in which they were scaffolded to produce anomalous evidence. Young children in particular need support consolidating explanations and evidence (Kelemen et al., 2014). Although empirical research has shown the importance of individual factors of support (i.e., constrained environments, prompts, explanations) on scientific understanding, little empirical research has examined multiple factors in combination (Lazonder & Harmsen, 2016). The current research builds on research on the role of conceptual explanations in science learning by examining whether embedding explanations in guided activities impacts the revision of deeply entrenched beliefs over and above the effects of guidance alone.

The Importance of Explanations for Learning Science

As early as age 3, children are ready to receive and interpret age-appropriate causal explanations for physical events. Children are able to select or propose physical events to explain physical relations, biological patterns for biological relations, and psychological phenomena for psychological relations (Keil, 2006). Research has documented that young children can learn about complex science concepts from causal explanations provided in picture books, such as natural selection (Kelemen et al., 2014), balance (Larsen et al., 2020), sinking and floating (Ganea et al., 2021), and even the rate at which objects fall (Venkadasalam & Ganea, 2018). When children's current beliefs are incongruent with evidence, alternative explanations for anomalies provide children with a framework to interpret the evidence in inquiry activities (Brewer et al., 1998; Chinn & Malhotra, 2002), which can potentially enhance children's immediate (i.e., making correct observations and interpretations) and long-term (i.e., generalizations and retention) learning.

One study examined the impact of explanations combined with anomalous evidence embedded in guided activities that children experienced first-hand compared to picture books they listened to (Larsen et al., 2020). Five-year-olds learned to balance asymmetrical objects equally well from both interventions compared to a baseline activity with no guidance and a control book about plants. This research illustrates the important role of explanations when learning from anomalies. When anomalous evidence was combined with a refutation of the incorrect idea and an explanation, either through a picture book or a guided activity, 5-year-olds incorporated distance into their beliefs about balancing unevenly weighted objects.

Another study examined whether the order of explanations from picture books when combined with a guided anomalous evidence activity plays a role in the learning of a more complex and deeply entrenched concept (Ganea et al., 2021). During a posttest that immediately followed the intervention phase, children revised their beliefs about sinking and floating more successfully when the alternative scientific explanation preceded the anomalous evidence than when the explanation followed the anomalous evidence. Having access to an alternative explanation before anomalies helped children interpret the evidence that did not fit their intuitive beliefs. The reverse order, where they first interacted with anomalous evidence and then received a scientific explanation for it, impeded their ability to use the alternative explanation as well. Together these one-session interventions demonstrated that reasoning about anomalous evidence was difficult for 5-year-olds unless they received a plausible explanation for the anomalies. We do not currently know whether such explanations would have an impact on strongly engrained beliefs over longer periods of time.

Research with older children showed that providing students with an explanation before anomalous evidence only improved their ability to make the correct observations but provided no additional benefit in the interpretation, generalization, or retention stages (Chinn & Malhotra, 2002). Similarly, Hardy et al. (2006) found that when third graders were prompted to reflect on their beliefs about sinking and floating and consider a new explanation, they improved their understanding immediately and maintained their learning after a year. In contrast, children who were only provided with the correct explanation improved their understanding at posttest but did not maintain this knowledge at the delay test. This research highlights the importance of embedding scientific explanations in guided activities. This may be especially important when trying to address young children's entrenched beliefs.

The Role of Guidance in Learning

Research has demonstrated that for everyone but experts, minimal guidance during instruction is less effective than extensive guidance (R. Clark et al., 2012). Thus, for conceptual change to occur, children need more support than merely observing anomalous evidence (Butts et al., 1993; Chinn & Malhotra, 2002; Kloos & Van Orden, 2005). Beyond the provision of an alternative explanation, this study provided children with guidance to help them during the inquiry process because conceptual change requires diverse experiences with the target concept and a supportive environment for optimal learning (Duit et al., 2013). This work classified guidance as adult support and direction in play-based learning. Research has shown that adults can tailor the learning process to help the child learn as this approach takes inspiration from Vygotsky's "zone of proximal development" (Weisberg et al., 2016; Yu et al., 2018). Guidance was defined as dyadic interactions where adults scaffold the process to foster pedagogical objectives without interfering so that the activities remain child-led (Yu et al., 2018). This can include giving children clear objectives for an activity, such as to find out how objects fall when dropped at the same time.

Adult guidance, especially with younger children, can help make the evidence more explicit so it can be adequately integrated with children's prior beliefs and, as a result, successfully restructure their knowledge. For effective integration of evidence, children can be guided to systematically track and assess the anomalies across several instances (Tolmie et al., 2016), potentially highlighting the conflict between prior beliefs and evidence (Cheng & Brown, 2010; Zimmerman, 2007). If children receive both experimental and reflective support, they may be more likely to produce anomalous evidence and make accurate observations and interpretations. In this study, adults offered guidance to children during the inquiry process. Adults helped to create, observe, and correctly interpret anomalous evidence. This would allow children to be in a position in which they can possibly create their own explanations for the anomalies. The guidance condition would be more informative than the educational approach whereby children need to construct their knowledge and theories on their own (R. Clark et al., 2012).

The Current Research

The current work consisted of two experiments that focused on the concept of free fall. We build on previous work to examine how explanations can be used to improve young children's observation, interpretation, generalization, and retention of anomalous evidence and extend this to younger children. Recall that when older children were provided with explanations for anomalies, they only used explanations during the initial observation stage (Chinn & Malhotra, 2002). We aimed to improve children's ability to rely on explanations beyond the observation stage and to help them apply the explanations to the interpretation, generalization, and retention stages. Therefore, instead of providing explanations before the observation stage, in this study, children were first prompted to produce and observe anomalous evidence and then guided to draw the correct conclusion. Since research has shown that children struggle to produce anomalies (Penner & Klahr, 1996) and make accurate observations (Chinn & Malhotra, 2002; Tolmie et al., 2016), guidance was used for scaffolding the observation process. Children were asked to explain the anomaly during the guided activity with the goal of making them realize that their beliefs do not align with the evidence they just observed, thus highlighting a gap in knowledge. Following the guided observation of anomalous evidence, half of the children also received a conceptual explanation of the evidence. Unlike previous work, the explanation was included in the activity instead of being provided before or after the evidence. This manipulation allowed us to explore whether and how the provision of conceptually rich explanations improved children's ability to interpret, encode, and generalize their learning from anomalous evidence. Given existing evidence that explanations helped support older children's ability to make correct observations about anomalous evidence (Chinn & Malhotra, 2002), this research investigated the effect of explanations beyond the observations phase in young children with complex physical science concepts. The explanations in this study contained simple and straightforward causal information about a physical science concept. The explanation was also paired with a refutation of the incorrect belief as conceptual change research has shown that refutations can further promote conceptual change (Ganea et al., 2021; Kendeou et al., 2014; Larsen et al., 2020; Tippett, 2010; Venkadasalam & Ganea, 2018). The main research question was whether embedding an explanation, which included a refutation of incorrect prior beliefs and information about the observed phenomenon, in a guided activity aids in the interpretation and generalization of anomalous evidence and promotes conceptual change above and beyond the effect of guidance alone in young children. These experiments were not preregistered, but the data are available on the Open Science Framework website (Venkadasalam et al., 2023).

Experiment 1

The first experiment investigated whether the presence of an explanation embedded in a guided activity helps 4- and 5-year-olds revise their beliefs about falling objects. Research has shown that the ability to use evidence to make inferences develops significantly between the ages of 4–6 but supporting children in these years can help to improve later abilities (Tullos & Woolley, 2009). We build on previous work that found that 4- and 5-year-olds revised their predictions about falling objects after hearing an explanation from a picture book (Venkadasalam & Ganea, 2018). Since explanations with a refutation in a picture book alone were effective at promoting the revision of beliefs for this specific concept, we directly extend this work by embedding identical explanations into guided activities to examine whether explanations provide a benefit beyond guidance alone.

This experiment had three conditions: guidance only, guidance with explanation, and a baseline condition. In the guidance only condition, the experimenter scaffolded children to produce examples of anomalies and ensured that children made the correct observations by explicitly verbalizing that both objects hit the ground simultaneously. The guidance with explanation condition was the same, except the experimenter embedded conceptual information (e.g., “What goes up must come down. When an object is tossed into the air it will always come back down. If you hold an object and let it go, will it fall to the ground?”) in the activity, refuted incorrect beliefs (e.g., “Both jars reach the ground at the same time, even though one is heavier than the other, just like the boxes. Some people think heavier things reach the ground before light ones! Let’s find out if the same thing happens with buckets.” And explained the concept of gravity). Comparing the two guided conditions indicated whether explanations promoted the correct interpretation and generalization of the anomalous evidence above and beyond the effect of guidance alone. The comparison between the two guided conditions and the baseline condition, which included no guidance or explanation, provided insight into children’s ability to produce and evaluate anomalous evidence with and without support. The baseline condition parallels what frequently happens in the classrooms because many educators approach inquiry base science activities with the belief that children need to construct their own knowledge and theories (R. Clark et al., 2012).

Children’s conceptual understanding was measured before and after the intervention. Children were asked to predict whether same- and different-weight pairs of objects fall at the same rate. We expected more correct predictions for same-weight objects at posttest for all conditions. However, we expected differential response patterns for object pairs that differed in weight. For the baseline condition, we hypothesized no significant improvement from pre- to posttest because children would struggle to produce anomalous evidence. For the small number of children in the baseline condition who created examples of anomalies, we hypothesized that children would have difficulty with the observation stage (i.e., drawing correct conclusions about the outcome) because objects accelerate quickly to reach terminal velocity. Relative to the baseline condition, we hypothesized that children’s performance in both experimental conditions improved from pre- to posttest because they were guided to produce and observe anomalous evidence. Importantly, we hypothesized that children in the guidance with explanation condition would make more accurate predictions at posttest than children in the guidance only condition. These findings would support the conclusion that explanations embedded with guidance aid younger children’s observation of anomalous evidence beyond the observation stage.

Method

Participants

Two hundred and thirty-eight 4-year-olds ($n = 120$; $M = 4.53$; range: 4.01–4.98; 62 females, 58 males) and 5-year-olds ($n = 118$; $M = 5.46$; range: 5.00–5.99; 59 females, 59 males) participated in this study. Sixty-seven additional children were excluded because they had a perfect score on the pretest ($n = 25$), did not complete the protocol ($n = 15$), were exposed to English <50% of the time at home ($n = 22$), experienced parental interference ($n = 2$) or encountered experimenter error ($n = 3$). Approximately equal numbers of 4- and 5-year-olds were randomly assigned to one of three conditions: guidance with explanation ($n = 78$, $M_{\text{age}} = 4.96$; 41 females, 37 males), guidance only ($n = 79$, $M_{\text{age}} = 4.99$; 39 females, 40 males), and baseline ($n = 81$, $M_{\text{age}} = 5.03$; 41 females, 40 males).

Participants were recruited and tested at a science museum in a major metropolitan area of southern Ontario, Canada. The [university] Research Ethics Board (REB) approved this experiment. A female experimenter individually tested children in a designated testing area. The sample of children came from diverse ethnic backgrounds, including Asian (28%), White (26%), Latin American (1%), Black (1%), and mixed race (17%) children. An additional 26% of families declined to disclose ethnicity. Of those families who disclosed income, most children came from middle- and upper-class families (35%); however, 56% of families declined to disclose this information. A bachelor’s degree (15%) was the modal level of parental education, although 24% of families declined to disclose a level of education for either parent.

Materials

There were two object sets for the pre- and posttests, each of which consisted of four pairs of objects, for a total of eight object pairs (for pictures of each object set, see Supplementary Material A: Objects Sets for Test Phase in the online supplemental materials). All pairs were the same size and shape, but two were the same weight, and two were different weights. Half the pairs (one same weight and one different weight) were identical objects, whereas the remaining two were visually distinct. Object sets were counterbalanced across test phases (pre- and posttest). However, the order of presentation was the same for all children (identical same weight, nonidentical different weight, nonidentical same weight, and identical different weight) to avoid differences in response patterns between children.

Although children only completed a single activity, two types of activities were developed to ensure that any differences in learning were not a function of the type of activity. The fill and drop activity ($n = 118$) and a prediction with video activity ($n = 120$) each provided three examples of anomalous evidence, showing two differently weighted objects falling at the same rate (see Supplementary Material B: Activity Scripts and Materials in the online supplemental materials). For the fill and drop activity, children were given three pairs of containers (boxes, jars, and buckets), and within each pair, the containers were similar in size. During the activity, one of the containers from each pair was filled with light materials (feathers, pom-poms, yarn, or sponges) and one with heavy materials (marbles, rocks, crystals, or blunt-tipped nails). For the prediction with video activity, children were given three pairs of objects (blocks, balls, and animals) one pair at a time. Within each pair, objects were similar in size but varied in weight. In this activity, children recorded their predictions and the results on a worksheet. The prediction sheet had three rows with a picture of each pair of objects and two columns, one for their prediction and one for the outcome. Each column contained pictures of two balls falling at the same or different rates, which children could circle. After observing each pair of objects fall in person, children also watched a video clip of the objects being dropped in slow motion.

Procedure

This study used a between-subjects design, with four phases in the protocol: a weight test to ensure children understood weight; a pretest administered to determine children's prior beliefs; an intervention phase designed to target their prior beliefs; and a posttest to measure learning. During the intervention phase, children were randomly assigned to complete a single activity in one of three conditions: guidance with explanation, guidance only, or the baseline condition (see Table 1 for summary). The entire session was video recorded and lasted approximately 15–20 min.

Weight Test. First, children were asked to compare the weight of objects to ensure that they had a fundamental concept of weight. Two pairs of objects were used, one with the same and one with different weights. Children were asked, "Do these objects have the same weight or different weight?" followed by "How do you know?" Children who correctly justified their answers (e.g., "this one is heavy and this one is light") proceeded directly to the test phase. Children ($n = 165$) who answered incorrectly or incorrectly justified their answer (e.g., "this one has spots") were taught about weight. The experimenter asked children to think about how heavy and light objects feel in their hands. After the training, all children demonstrated an understanding of weight and continued to the pretest.

Test Phase. The pre- and posttest each followed the same procedure. Children held an object pair and were asked whether object pairs had the same size and weight, highlighting these two features. For different-weight pairs, children were also asked to identify the heavier object. Across the trials, children were adept at assessing whether the objects were equal in size (98.27%), whether the objects had the same or different weights (99.68%) and identifying the heavier object (99.79%). Most children answered correctly, but the few who answered incorrectly on some trials were not excluded and received feedback on their answers. Children were then asked the test question: "If I held the objects out like this and let them drop, do you think one of the two will fall faster, or do you think both will fall at the same time?" The sequence of the predictions (i.e., one or both) in the test question was counterbalanced. After answering the test question, children received neutral feedback ("Thank you").

Activity Intervention. Children in each condition completed one activity. In the fill and drop activity, children were given a pair of empty containers similar in size and a selection of heavy and light "filler" objects. They were told they would find out what happens when two objects were dropped together. Children were instructed to fill one container with lightweight materials and one with heavy materials. They were then given the containers and asked to compare their weight. The experimenter also explicitly told children which container was heavy and light to direct their attention to the difference in weight. The experimenter then dropped the containers, and the children observed that both fell at the same rate. If children made an incorrect observation and concluded that one fell faster, the containers were dropped again, and children's attention was directed to the fact that the objects fell simultaneously. This process was repeated with two remaining pairs of containers and different heavy and light materials.

In the prediction with video activity, children in both experimental conditions were shown a prediction sheet and told they would experiment like scientists. They were told they would make predictions about what happens when different objects are dropped together, test their predictions, and record the results. Children were given one pair of objects that varied in weight. They made predictions about how the pairs of objects would fall when dropped together (at the same time, or one faster than the other). They observed the experimenter drop the pair of objects and were immediately asked

Table 1
Design Summary for Experiment 1

Elements in each condition	Guidance with explanation	Guidance only	Baseline
Instructions highlight prior beliefs	☑	☑	☑
Summary of observations	☑	☑	
Guidance to create anomalous evidence	☑	☑	
Make correct interpretations	☑	☑	
Conceptual information throughout activity	☑		
Refutation before final anomalous evidence	☑		
Explanation at the end of the activity	☑		

for their observation. If children made an incorrect observation and concluded that one fell faster, the objects were dropped again, and children's attention was directed to the fact that the objects fell simultaneously. Next, the children watched a video clip of the objects being dropped in slow motion. The clip encouraged children to make accurate observations that the objects fell simultaneously. Finally, the children recorded their observations on the prediction sheet. This procedure was repeated with two more pairs of objects.

The activities were completed the same way across the two experimental conditions, which differed only in the conceptual information provided during the intervention. In the guidance only condition, children completed one activity following the procedure described above. They were guided to create and observe the anomalous evidence but did not receive conceptual information or an explanation about gravity. In the guidance with explanation condition, the experimenter provided conceptual information related to the concept in the activity (i.e., "What goes up must come down"), connected the evidence across trials (i.e., "Just like with the blocks"), and explained gravity at the end of the activity (i.e., "Gravity is the force that makes objects fall to the ground. Gravity affects things that are similar in size in the same way. When objects that are almost the same size are dropped together, they reach the ground at the same time, no matter what they weigh.").

In the baseline condition, children were given objects from the activities and asked to find out how objects of different weights fall when dropped together. Children received no guidance on how to find the answer and received no feedback about whether they made correct observations. The containers and materials that children in this condition received for the fill-and-drop activity were chosen to keep the task simple. They had two identical containers, rocks, and pom-poms, to prevent them from becoming overwhelmed or distracted, but could interact with the objects for as long as they wished (i.e., drop objects as many times as they liked). In the prediction with video activity, children were given all three pairs of objects and shown how to access the slow-motion video on an iPad. Since the iPad was locked on the video screen, children could only pause and play the video and had no trouble accessing them.

Coding

All sessions were live-coded and recorded. Two research assistants who were blind to the conditions and hypotheses coded children's responses to the pre- and posttest questions. Children's predictions were scored as 0 (*one object falls fasters*) or 1 (*both objects fall at the same time*). Most of the children's responses were coded from videos (90%), but live coding was used for 25 participants because the session was not recorded properly. There was high interrater reliability determined by Cohen's $\kappa = .94$, $p < .001$, a 96.86% agreement rate. A third coder resolved disagreements.

Results

For the hypothesis-driven analysis, we examined whether children's predictions changed after the intervention as a function of the condition. We also considered the effect of weight (same- vs. different-weight pairs) on predictions because children who believe that weight determines the rate at which objects fall would predict heavier (or in some cases the lighter) objects to fall faster, and that same-weight object pairs would fall at the same rate. Logistic mixed-effect modeling was used to determine whether children changed their predictions about the rate at which pairs of objects fall following an activity. This type of modeling was selected to account for random intercepts for participants and trials (Baayen et al., 2008), the presence of within- (test phase and weight) and between-subject (conditions) factors, as well as to accommodate the binary nature of the dependent variable. Random slopes were not included because we were interested in overall change across children rather than individual slopes for each participant. The models were fit using the lme4 package (Bates et al., 2015) in R (R Core Team, 2019). This analytic strategy was used in all analyses.

A preliminary analysis found that age (4- vs. 5-year-old), activity (fill and drop vs. prediction with video), and counterbalancing of object sets were not related to children's predictions, so these factors were excluded from the following analyses. A separate analysis indicated that pretest scores were not statistically different across the three conditions ($ps > .74$). Figure 1 displays the proportion of correct predictions for all three conditions at both test phases for same- and different-weight objects.

The model included fixed effects and the interaction with the variables: weight (same and different), test phase (pre- and posttest), and condition (guidance with explanation, guidance only, and baseline). For the fixed effects of the weight and test phases, simple planned contrasts (−.05, .05) were conducted. The fixed effect of the condition was analyzed with two simple contrasts comparing the baseline to the experimental conditions (c_1: −.66, .33, .33) and comparing the guidance only to guidance with explanation (c_2: 0, −.5, .5) condition. The model included random intercepts for participants and fully cross-random effects for the trials (i.e., all children encountered every pair). Parameter estimates for these models are summarized in Table 2.

There was a significant effect of weight, such that children made more correct predictions about the same-weight ($M = 0.72$, $SD = 0.45$) than the different-weight ($M = 0.33$, $SD = 0.47$) objects. There was a significant main effect of the test phase which indicated that predictions made at posttest ($M = 0.64$, $SD = 0.48$) were more accurate than those made at pretest ($M = 0.41$, $SD = 0.49$). There was a significant effect of condition which revealed that children in the experimental conditions ($M_{\text{GwE\&GO}} = 0.56$, $SD = 0.50$) made more correct predictions than those in the baseline condition ($M = 0.46$, $SD = 0.50$). There were significant interactions between weight and both condition contrasts (c_1: baseline vs. experimental conditions and c_2: guidance with explanation vs. guidance only). Follow-up tests for both types of weight separately indicated that children's predictions were similar across all conditions for same-weight objects ($M_{\text{GwE}} = 0.69$, $SD = 0.46$, $M_{\text{GO}} = 0.72$, $SD = 0.45$, $M_{\text{B}} = 0.73$, $SD = 0.44$), but children's predictions differed across all conditions for different-weight objects. Children in the guidance with explanation condition ($M = 0.46$, $SD = 0.50$) had the most accurate predictions, followed by those in the guidance only condition ($M = 0.34$, $SD = 0.48$) and children in the baseline condition had the least accurate responses ($M = 0.19$, $SD = 0.39$). There was also an interaction between test phase and condition (c_1: baseline vs. experimental). Follow-up tests for each test phase separately revealed that children's predictions were similar across conditions at pretest but differed as a function of condition at posttest. Children in the experimental conditions improved from pre- ($M_{\text{GwE\&GO}} = 0.40$, $SD = 0.49$) to posttest ($M_{\text{GwE\&GO}} = 0.71$, $SD = 0.45$), but children's predictions in the baseline condition did not differ at pre- ($M = 0.41$, $SD = 0.49$,) and posttest ($M = 0.80$, $SD = 0.40$).

Figure 1

Proportion of Correct Predictions for the Same- and Different-Weight Objects as a Function of Test Phase and Condition in Experiment 1

Note. Error bars represent standard errors.

The main effects and two-way interactions were superseded by a three-way interaction between weight, test phase, and condition. Follow-up tests were conducted for weight and test phase separately. For same-weight objects, there was a similar improvement in children's predictions from pre- to posttest for the baseline ($M_{\text{Pre}} = 0.64$, $SD = 0.48$, $M_{\text{Post}} = 0.83$, $SD = 0.38$) and experimental conditions ($M_{\text{Pre}} = 0.62$, $SD = 0.49$, $M_{\text{Post}} = 0.79$, $SD = 0.40$). For different-weight objects, children's predictions improved from pre- to posttest in the experimental conditions ($M_{\text{Pre}} = 0.19$, $SD = 0.39$, $M_{\text{Post}} = 0.19$, $SD = 0.39$), but there was no change for the baseline condition ($M_{\text{Pre}} = 0.19$, $SD = 0.39$, $M_{\text{Post}} = 0.62$, $SD = 0.48$). As expected, children's predictions in the baseline condition for different-weight objects did not improve, while children in the two guided conditions made more correct predictions after the intervention. However, there were no significant differences between the two experimental conditions at posttest.

Discussion

This experiment tested whether embedding explanations in guided activities promotes belief revision in children. For the same-weight object pairs, children's performance improved over time regardless of condition. This finding is consistent with previous research using this task, showing improvement in children's predictions for same-weight objects over time (Venkadasalam & Ganea, 2018). For different-weight objects, children in the guidance with explanation and guidance only conditions made more accurate predictions at posttest compared to children in the baseline condition. Thus, guidance helped children produce and correctly observe anomalies and prompted them to revise their predictions for different-weight objects. Despite the expected trend, there was no additional benefit of explanations on children's predictions.

Experiment 1 highlighted the independent role of guidance in helping children observe anomalies and revise their predictions about different-weight objects. This demonstrates the importance of scaffolding during the observation stage, which can further improve children's learning at the interpretation stage. Children were guided to create and assess three instances of anomalous evidence (Tolmie et al., 2016). Although children in the guidance with explanation condition trended toward having more accurate predictions than children in the guidance only condition, this pattern was not statistically significant. The absence of an additive benefit of explanations is not entirely surprising and is consistent with research with older children, showing that explanations only influence the observation stage of learning from anomalous evidence (Chinn & Malhotra, 2002).

Table 2

Parameter Estimates for the Correct Prediction Scores in Experiment 1

Effect	Estimate	*SE*	*z*	*p*
Intercept	0.10	0.13	0.81	.42
Weight (same vs. different)	**1.94**	**0.24**	**8.12**	**<.001**
Test phase (pre- vs. posttest)	**1.19**	**0.24**	**5.04**	**<.001**
Condition (experimental vs. baseline)	**0.43**	**0.15**	**2.92**	**.003**
Condition (guidance only vs. guidance with explanation)	0.24	0.17	1.45	.15
Weight (Same vs. Different) × Test Phase (Pre- vs. Posttest)	−0.47	0.47	−1.00	.32
Weight (Same vs. Different) × Condition (Experimental vs. Baseline)	**−1.22**	**0.25**	**−4.96**	**<.001**
Weight (Same vs. Different) × Condition (Guidance Only vs. Guidance With Explanation)	**−0.81**	**0.27**	**3.01**	**.003**
Test Phase (Pre- vs. Posttest) × Condition (Experimental vs. Baseline)	**0.98**	**0.24**	**−4.02**	**<.001**
Test Phase (Pre- vs. Posttest) × Condition (Guidance Only vs. Guidance With Explanation)	0.24	0.27	−0.89	.38
Weight (Same vs. Different) × Test Phase (Pre- vs. Posttest) × Condition (Experimental vs. Baseline)	**2.37**	**0.49**	**4.87**	**<.001**
Weight (Same vs. Different) × Test (Pre- vs. Posttest) × Condition (Guidance Only vs. Guidance With Explanation)	−0.41	0.54	0.44	0.44

Note. Significant parameter estimates are bolded.

It is unlikely that the explanations children received were insufficient because the explanations in this study have been shown to improve the predictions of 4- and 5-year-olds (Venkadasalam & Ganea, 2018). However, a few other limitations of the current experiment can account for the findings. First, it is impossible to determine whether explanations influence the retention stage because this was a single-session study. Since conceptual change is a gradual process (Vosniadou, 2013), without measuring children's learning after a delay, it is unknown whether they have undergone conceptual change or a short-term belief revision. Therefore, learning needs to be measured over an extended period to determine whether children have incorporated the new explanation into their knowledge. A second experiment included a delay test to explore this possibility.

Another consideration is that the dependent variable may not have been sensitive enough to capture whether children have undergone a conceptual change. Asking children to predict the rate at which objects fall without having them explain their reasoning may prevent them from thinking deeply about their answers, reducing the chance of dissatisfaction with and evaluation of their prior beliefs (Chinn & Malhotra, 2002; Posner et al., 1982). According to Li and Klahr (2005), predictions are unlikely to lead students to revise their beliefs because predictions are more likely to support a course of action that informs guessing by trial and pattern recognition, resulting in children being less likely to link their prediction to a testable hypothesis. However, for conceptual change to occur, children need to engage in deep-level processing. They not only have to know the correct explanation but also to recall and refute the incorrect belief (Tippett, 2010; van den Broek & Kendeou, 2008). Predictions capture an implicit level of engagement with the task, whereas asking children to explain their reasoning requires making explicit connections between the evidence and the conceptual information (Howe et al., 2013; Tolmie et al., 2016). Therefore, eliciting explanations may highlight gaps in children's knowledge that may not be apparent when using more implicit measures (Keil, 2006). Using a measure that elicits more explicit connections would allow us to determine whether children were merely engaging in heuristic pattern-based reasoning (i.e., picking the "same time" as the correct answer) or incorporating the evidence and explanation into their knowledge. In other words, asking children to explain their predictions can provide insight into whether they were parroting back what they heard from the intervention without updating their beliefs.

Furthermore, from their predictions, we cannot know whether children learned that even same-weight objects fall at the same rate because of gravity, rather than simply because they have a similar weight. It is possible that children might only be engaging in weak restructuring (Duit & Treagust, 2003) of their beliefs (i.e., accepting that different-weight objects also fall at the same time) without restructuring their entire theory (i.e., that weight does not affect the rate at which objects fall). Therefore, it is difficult to determine whether children underwent conceptual change or merely belief revision without knowing the reasoning behind their predictions. These questions were investigated in the following experiment.

Experiment 2

In Experiment 2, children were prompted to explain their predictions (referred to hereafter as justifications to differentiate from explanations provided to children) to evaluate whether there was an additive effect of explanations on children's learning; and a delay test was included to assess whether children underwent conceptual change or engaged in a weak reorganization of beliefs. We expected that children's justifications of their predictions would lead to differences in conceptual understanding. Justifications would prompt children to think more deeply about the information they received and how to apply it to new scenarios. The explanations provided by the experimenter could help children connect the anomalous evidence with their prior beliefs and offer a framework to interpret the anomalies (Brewer et al., 1998; Chinn & Malhotra, 2002). Without an explanation, evaluating anomalous evidence is difficult for children because they fail to create their own theory (Chinn & Brewer, 1993) and keep the evidence isolated from their prior beliefs (Zimmerman, 2007). If children use the explanations provided in their own justifications, this will indicate that they may be in the process of undergoing conceptual change. Thus, children in the guidance with explanation condition should learn and retain more than children in the guidance only condition.

Method

Participants

Eighty 5-year-olds ($M = 5.50$; range: 5.01–5.97, 40 females, 39 males, one gender nonbinary) participated in this study. Two additional children were excluded due to parental interference ($n = 2$). Equal numbers of children were randomly assigned to the two conditions of interest from Experiment 1: guidance with explanation ($n = 40$, $M_{\text{age}} = 5.51$, 20 females, 20 males), and guidance only ($n = 40$, $M_{\text{age}} = 5.50$, 20 females, 19 males, one gender nonbinary). Equal numbers of children were in each activity in each condition ($n = 20$).

Participants were recruited from a database of families who had expressed interest in participating in research. The [university] REB approved the experiment. A female experimenter individually tested children in a quiet room at a university in a major metropolitan area in southern Ontario, Canada. Most children identified as White (45%), but the sample also included children who identified as Asian (13%), Latin American (3%), Black (4%), and mixed race (33%). An additional 5% of families declined to disclose ethnic information. Most children came from middle- and upper-class families (69%), with 10% of families declining to disclose income level. The modal parental education level was a bachelor's degree (38%), with 8% of families declining to disclose a level of education for either parent.

Materials

The activity materials and test phase objects were the same as those used in Experiment 1. However, an additional set of objects containing four pairs was due to the addition of the delay test. The use of three object sets allowed the object sets to be counterbalanced between all test phases.

Procedure

The procedure for this experiment was similar to that of Experiment 1, albeit with several key additions. There were two sessions, both of which were video recorded. The first session lasted 15–20 min and was identical to Experiment 1, where children completed a weight test, pretest, intervention, and posttest. Children then returned 5–10 days ($M = 6.97$ days) later for a second session lasting approximately 5 min, where children completed a weight test followed by a delay test to measure retention. During the weight test, some children needed to be trained for the first ($n = 34$) and second ($n = 24$) sessions. Most of these children were trained during a single session (first: $n = 21$ or second: $n = 11$ only), although some required training before both sessions ($n = 13$).

The pre-, post-, and delay tests followed the same procedure as Experiment 1, except that there were two dependent variables: children's predictions and their justifications. After children were asked to make a prediction, justifications were obtained by asking "Why will this one fall faster?" or "Why will they both fall at the same time?" depending on the child's prediction. Justifications provided an explicit measure of children's reasoning and insight into how children's beliefs changed as a function of weight. As in Experiment 1, participants were adept at assessing whether the objects were equal in size (99.58%), comparing the weight of the objects (99.58%), and identifying the heavier object (98.75%). As in Experiment 1, most children answered correctly. However, children who answered incorrectly received feedback on the specific trial were included in the study.

Coding

There were two dependent variables: children's predictions and justifications. For each trial, children's predictions were coded in the same manner as in Experiment 1 (0 for *one object falling faster*; 1 for *both objects falling at the same rate*). Children's justifications were scored on a 0–2 scale for each trial. A score of 2 was assigned if children correctly identified that objects (both same- and different-weight pairs) would fall at the same rate and also correctly justified their predictions by reasoning about gravity and/or the similar size of the object pairs, or stated that they had learned a rule from the activity (e.g., "I remembered from the activity that they always fall at the same time"). The latter answer allowed children in the guidance only condition to readily score a 2 without referencing gravity or making inferences about shape. A score of 0 was assigned if children made an incorrect prediction (e.g., "the object pairs would fall at different rates"), or if they predicted that object pairs would fall at the same time but gave an incorrect justification referencing an erroneous belief (e.g., "they will fall at the same time because they have the same weight") or irrelevant feature (e.g., "they will fall at the same time because they are both blue"). The correctness of the predictions was considered in conjunction with the justifications because this was a more stringent measure of learning, such that children needed to apply the correct prediction and reasoning. A score of 1 was assigned if children correctly identified that object pairs would fall at the same rate and gave a correct justification in conjunction with an erroneous belief (e.g., "they will fall at the same time because of gravity, and they have the same weight").

Two research assistants blind to the hypotheses, test phase, and condition coded 100% of the children's responses. The coders were in 100% agreement with the prediction responses. There was high interrater reliability determined by Cohen's $\kappa = .91$, $p < .001$, a 96.04% agreement rate for the justification responses. The two coders resolved disagreements through discussion.

Results

The first two analyses assessed the dependent variables (children's predictions and justifications) as a function of weight, test phases, and condition. A final analysis explored the types of justifications children provided. Preliminary analyses revealed that fixed effects of age, the gap between Sessions 1 and 2 (in days), the counterbalanced order of test phase object sets, and the type of activity were not significant predictors of predictions or justifications[1] and were excluded in the following analysis.

Prediction Responses

Logistic mixed-effect modeling was used to analyze whether children revised their predictions (from one object falls faster to both falling at the same rate). A preliminary analysis indicated that pretest scores were similar across both conditions ($p > .46$). Figure 2 displays the proportion of correct predictions for both conditions at each test phase for same- and different-weight objects.

Children's predictions were analyzed with a model that included fixed effects and the interactions of weight (same and different), test phase (pre-, post-, and delay test), condition (guidance with explanation and guidance only), and random intercepts for trials and participants. For the fixed effects of weight and condition, planned contrasts (−.5, .5) were conducted. The fixed effect of the test phase was analyzed with simple contrasts comparing before (pretest) to after the intervention (post- and delay test; c_1: −.66, .33, 33) and then post- to delay test (c_2: 0, −.5, .5). Parameter estimates for the model are summarized in Table 3.

A significant main effect of weight indicated that children made more correct predictions about same-weight ($M = 0.82$, $SD = 0.39$) than different-weight objects ($M = 0.38$, $SD = 0.49$). A significant main effect of test phase (c_1: pre- vs. post- and delay test) revealed that predictions made after the intervention ($M_{\text{Post\&Delay}} = 0.70$, $SD = 0.46$) were significantly more accurate than those made before the intervention ($M_{\text{Pre}} = 0.40$, $SD = 0.49$). Finally, the main effect of condition showed that children in the guidance with explanation condition ($M = 0.67$, $SD = 0.47$) made more accurate predictions than children in the guidance only condition ($M = 0.53$, $SD = 0.50$).

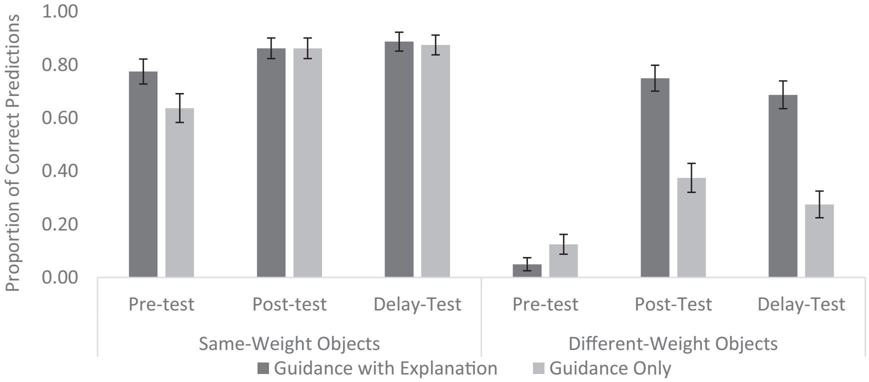
There was a two-way interaction between weight and test phase (c_1: pre- vs. post- and delay test). Follow-up tests for both types of weight separately indicated that the improvement in prediction accuracy after the intervention was greater for different-weight ($M_{\text{Pre}} = 0.09$, $SD = 0.28$, $M_{\text{Post\&Delay}} = 0.52$, $SD = 0.50$) than same-weight objects ($M_{\text{Pre}} = 0.71$, $SD = 0.46$, $M_{\text{Post\&Delay}} = 0.87$, $SD = 0.33$). There was an interaction between test phase (c_1: pre- vs. post- and delay test) and condition. Follow-up tests for each test phase separately showed that children's prediction only improved after the intervention for children in the guidance with explanation condition ($M_{\text{Pre}} = 0.41$, $SD = 0.49$, $M_{\text{Post\&Delay}} = 0.80$, $SD = 0.40$); this change was not significant for children in the guidance only condition ($M_{\text{Pre}} = 0.38$, $SD = 0.49$, $M_{\text{Post\&Delay}} = 0.60$, $SD = 0.49$).

---

[1] There was a significant effect for the type of activity in the preliminary model containing *only* the fixed effect of age, the gap between Sessions 1 and 2 (in days) and the counterbalanced order of test phase object sets. A subsequent analysis with the type of activity added to the justification hypothesis-driven model (with weight, test phase, and condition as fixed effects) revealed no significant main effect of activity nor any significant interactions. Furthermore, the addition of activity did not improve model fit that was based on a chi-square test of the change in −2 restricted log likelihood, $\chi^2(12) = 18.92$, $p = .09$, suggesting that with all main variables of interest included in the model children's justifications did not vary as a function of the type of activity. Thus, type of activity was not retained as a fixed effect in the final model.

Figure 2

Proportion of Correct Predictions for the Same- and Different-Weight Objects as a Function of Test Phase and Condition in Experiment 2



Note. Error bars represent standard errors.

Finally, these findings were superseded by a three-way interaction between weight, test phase, and condition. Follow-up tests were conducted for weight and test phase separately. For same-weight objects, there was a similar improvement in children's predictions after the intervention for both the guidance with explanation ($M_{\text{Pre}} = 0.78$, $SD = 0.42$, $M_{\text{Post\&Delay}} = 0.88$, $SD = 0.33$) and guidance only ($M_{\text{Pre}} = 0.64$, $SD = 0.48$, $M_{\text{Post\&Delay}} = 0.87$, $SD = 0.34$) conditions. For different-weight objects, children's predictions improved in both conditions from pre- to posttest. However, children in the guidance with explanation condition ($M_{\text{Pre}} = 0.05$, $SD = 0.22$, $M_{\text{Post\&Delay}} = 0.72$, $SD = 0.45$) made more accurate predictions compared to children in the guidance only condition ($M_{\text{Pre}} = 0.13$, $SD = 0.33$, $M_{\text{Post\&Delay}} = .33$, $SD = 0.47$). Notably, predictions at post- and delay test never differed across all fixed effects and interactions.

Justification Responses

Each justification was coded as *correct* (2), *incorrect* (0), or *mixed* (1; partially correct). A separate analysis showed pretest scores were similar across both conditions ($p = .98$). The proportion of children's justification scores across three test phases for both conditions is displayed in Figure 3 for same- and different-weight objects, respectively.

Mixed models with Poisson distributions and a log link function analyzed whether children could correctly justify that they thought the object pairs would fall at the same rate. A Poisson distribution was employed due to the ordinal nature of the justification scores (0–2). The models included the fixed effects of weight (same and different), test phase (pre-, post-, and delay test), condition (guidance with explanation and guidance only), the interaction between the three variables and the random intercept for participants. Planned contrasts (−.5, .5) were conducted for the fixed effect of weight and condition. The fixed effect of test phase was analyzed with simple contrasts comparing before (pretest) to after (post- and delay test) the intervention (c_1: −.66, .33, 33) and then post- to delay test (c_2: 0, −.5, .5). Parameter estimates for these models are summarized in Table 4.

As for predictions, there was a significant main effect of weight. However, children gave better justifications for different-weight ($M = 0.43$, $SD = 0.80$) than same-weight ($M = 0.41$, $SD = 0.72$)

Table 3

Parameter Estimates for the Correct Prediction Scores in Experiment 2

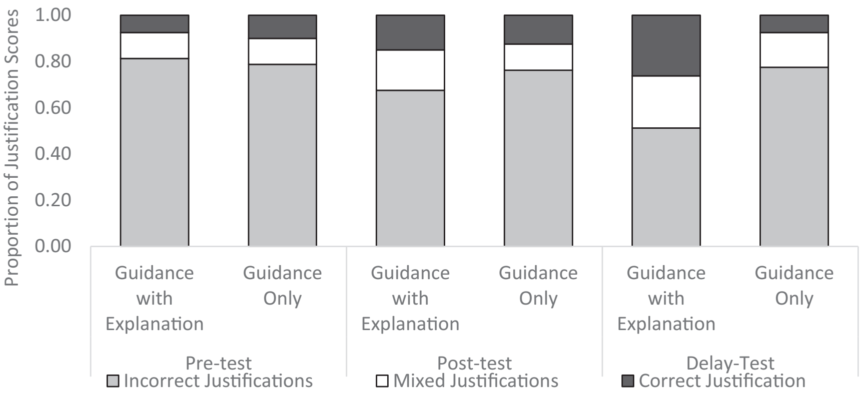
Effect	Estimate	*SE*	*z*	*p*
Intercept	**0.50**	**0.16**	**3.10**	**.002**
Weight (same vs. different)	**2.71**	**0.26**	**10.50**	**<.001**
Test phase (pre- vs. post- and delay test)	**2.06**	**0.28**	**7.45**	**<.001**
Test phase (post- vs. delay test)	−0.13	0.28	−0.46	.64
Condition (guidance only vs. guidance with explanation)	**−0.65**	**0.29**	**−2.24**	**.03**
Weight (Same vs. Different) × Test Phase (Pre- vs. Post- and Delay Test)	**−1.79**	**0.54**	**−3.30**	**.001**
Weight (Same vs. Different) × Test Phase (Post- vs. Delay Test)	0.65	0.57	1.14	.26
Weight (Same vs. Different) × Condition (Guidance Only vs. Guidance With Explanation)	0.68	0.40	1.72	.09
Test Phase (Pre- vs. Post- and Delay Test) × Condition (Guidance With Explanation vs. Guidance Only)	**−1.18**	**0.44**	**−2.65**	**.008**
Test Phase (Post- vs. Delay Test) × Condition (Guidance With Explanation vs. Guidance Only)	−0.16	0.44	−0.35	.73
Weight (Same vs. Different) × Test Phase (Pre- vs. Post- and Delay Test) × Condition (Guidance With Explanation vs. Guidance Only)	**3.85**	**0.89**	**4.33**	**<.001**
Weight (Same vs. Different) × Test (Post- vs. Delay Test) × Condition (Guidance With Explanation vs. Guidance Only)	0.05	0.88	−0.06	.95

Note. Significant parameter estimates are bolded.

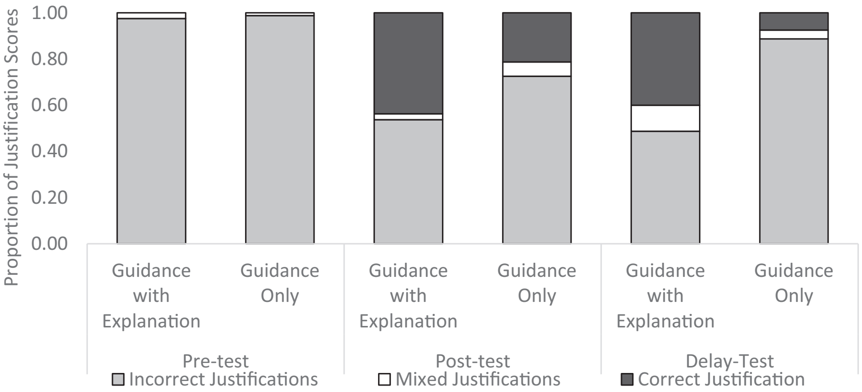
Figure 3

Proportion of Justification Scores (0, 1, or 2) in Experiment 2 as a Function of Test Phase and Condition

(a) *Same-weight objects*



(b) *Different-weight objects*



Note. (a) Same-weight objects; (b) different-weight objects.

objects. There was a significant main effect of test phase (c_1: pre- vs. post- and delay test) such that children's justification scores were higher after the intervention ($M_{\text{Post\&Delay}} = 0.55$, $SD = 0.83$) than before ($M_{\text{Pre}} = 0.16$, $SD = 0.48$). The effect of condition showed that children in the guidance with explanation condition ($M = 0.56$, $SD = 0.84$) had higher justification scores than children in the guidance only condition ($M = 0.28$, $SD = 0.64$).

Significant two-way interactions superseded these main effects. First, there was a significant interaction between weight and condition. Follow-up tests for both types of weight separately revealed that children's justifications for same- and different-weight objects scored higher for children in the guidance with explanation condition ($M_{\text{Same}} = 0.50$, $SD = 0.76$, $M_{\text{Diff}} = 0.62$, $SD = 0.90$) compared to the guidance only condition ($M_{\text{Same}} = 0.33$, $SD = 0.65$, $M_{\text{Diff}} = 0.23$, $SD = 0.61$). This finding indicated that the explanation fostered better justifications for different- and same-weight objects than guidance alone.

There were significant interactions between weight and both test phase contrasts (c_1: pre- vs. post- and delay test and c_2: post- vs. delay). Follow-up tests were conducted for both types of weight separately. For different-weight objects, justifications were scored higher during the post- ($M = 0.69$, $SD = 0.93$) and delay test ($M_{\text{Diff}} = 0.55$, $SD = 0.85$) compared to the pretest ($M = 0.03$, $SD = 0.24$). For same-weight objects, justifications were only scored higher during the delay test ($M = 0.53$, $SD = 0.77$) compared to the pretest ($M = 0.29$, $SD = 0.62$). However, the pre- and posttest ($M = 0.42$, $SD = 0.72$) did not differ. Together, this finding suggested that children were learning from the explanation and applying it to the different-weight anomalies at post- and delay test, but that children improved their justification for the same-weight objects only after approximately a 1-week delay.

Finally, there was an interaction between test phase (c_2: post- vs. delay test) and condition. Follow-up tests for each condition separately showed that the guidance with explanation scores at delay test ($M = 0.83$, $SD = 0.90$) were significantly higher than at posttest ($M = 0.69$, $SD = 0.90$), whereas the guidance only condition scores at delay test ($M = 0.24$, $SD = 0.58$) were significantly lower than at posttest ($M = 0.43$, $SD = 0.77$) and did not differ from pretest. Therefore, children who received guidance combined with explanations developed a more sophisticated understanding that was consolidated after a delay. In contrast, children who received only guidance improved their justification immediately after the intervention but

Table 4

Parameter Estimates for the Justification Scores in Experiment 2

Effect	Estimate	*SE*	*z*	*p*
Intercept	**−1.71**	**0.16**	**−10.49**	**<.001**
Weight (same vs. different)	**0.70**	**0.20**	**3.47**	**.001**
Test phase (pre- vs. post- and delay test)	**1.76**	**0.29**	**6.08**	**<.001**
Test phase (post- vs. delay test)	−0.17	0.12	−1.42	.15
Condition (guidance only vs. guidance with explanation)	**−0.92**	**0.31**	**−2.97**	**.003**
Weight (Same vs. Different) × Test Phase (Pre- vs. Post- and Delay Test)	**−2.63**	**0.58**	**−4.55**	**<.001**
Weight (Same vs. Different) × Test Phase (Post- vs. Delay Test)	**0.60**	**0.24**	**2.55**	**.01**
Weight (Same vs. Different) × Condition (Guidance Only vs. Guidance With Explanation)	**0.86**	**0.41**	**2.11**	**.03**
Test Phase (Pre- vs. Post- and Delay Test) × Condition (Guidance With Explanation vs. Guidance Only)	−0.24	0.58	−0.42	.68
Test Phase (Post- vs. Delay Test) × Condition (Guidance With Explanation vs. Guidance Only)	**−0.81**	**0.23**	**−3.41**	**.001**
Weight (Same vs. Different) × Test Phase (Pre- vs. Post- and Delay Test) × Condition (Guidance With Explanation vs. Guidance Only)	−1.07	1.16	−0.92	.36
Weight (Same vs. Different) × Test (Post- vs. Delay Test) × Condition (Guidance With Explanation vs. Guidance Only)	0.32	0.47	0.68	.50

Note. Significant parameter estimates are bolded.

did not maintain this understanding and returned to preintervention levels after a delay.

Types of Justifications

We briefly surveyed the responses where children achieved scores of 2 across both conditions to get a clearer picture of the types of justifications used by children (see Table 5). Justification scores were categorized as: (a) reasoning based on remembering what happened based on the outcomes (i.e., experiential knowledge such as, “I remember”; “I saw it in the activity”; “all the other ones did that”), (b) reasoning based on what they heard from the experimenter (i.e., testimony such as, “because you told me”; “I learned it”), (c) reasoning based on inferences children made about methodological soundness (timing, e.g., drop at the same time or property of the object, e.g., same size), or (d) reasoning based on scientific information (e.g., “because of gravity”). Cohen’s $\kappa = .98$, $p < .001$, a 98.82% agreement rate, determined high interrater reliability. The two coders resolved the two disagreements through discussion.

A chi-square goodness-of-fit test was used to determine whether the four types of reasonings (remembered, testimony, methodology, or scientific) were equally reported across the two conditions and two types of weight. Given the small proportion of scores of 2 (170 out of 960), the low pretest scores, and many cells with a frequency <5, we collapsed test phases and only examined the overall proportions across weight and condition. Thus, four tests were completed with a Bonferroni-adjusted alpha level of .0125 (.05/4). The reasonings across the four categories (remembered, testimony, methodology, or scientific) were not equally distributed for the guidance with explanation same-weight objects, $\chi^2(3,\ N = 42) = 15.33$, $p = .002$ or different-weight objects, $\chi^2(3,\ N = 66) = 16.06$, $p = .001$. The reasonings across the three categories (remembered, testimony, or methodology because there were no scientific reasons provided) were not equally distributed for the guidance only same-weight objects, $\chi^2(2,\ N = 32) = 21.06$, $p < .001$. However, reasonings across the three categories were equally distributed for the guidance only different-weight objects, $\chi^2(3,\ N = 30) = 7.20$, $p = .033$. The majority of children in both conditions reasoned based on the methodological properties of the experiment. However, almost 40% of responses in the guidance with explanation condition referenced scientific or learned information and most retained this level, but no children in the guidance only condition did so.

Discussion

The inclusion of justifications and a delay test in this experiment provided a more robust measure of conceptual change. Justifications differentiated children who engaged in heuristic pattern-based reasoning (i.e., merely predicting that all objects fall at the same rate) from children who revised their beliefs and generalized their knowledge (i.e., making correct predictions and demonstrating an understanding of why by applying the explanation to a new context). The addition of the delay test allowed us to assess the extent to which children retained this new conceptual understanding. Thus,

Table 5

Experiment 2 Correct Justification Scores (2 Out of 2) Categorized Into Types of Reasoning Across Test Phases and Conditions

	Guidance with explanation							Guidance only						
	Same weight			Different weight				Same weight			Different weight			
Reasoning	Pretest	Posttest	Delay test	Pretest	Posttest	Delay test	Total	Pretest	Posttest	Delay test	Pretest	Posttest	Delay test	Total
Experiential	2	1	3	0	7	6	19	2	3	4	0	11	5	25
Testimony	0	1	2	0	3	2	8	0	0	1	0	2	2	5
Methodological	3	6	5	2	12	12	40	10	7	5	1	6	3	32
Learned	0	10	9	0	12	10	41	0	0	0	0	0	0	0
Grand total	5	18	19	2	34	30	108	12	10	10	1	19	10	62

Experiment 2 provided three important findings: one regarding children's predictions and two regarding children's justification responses.

First, Experiment 2 replicated the findings from Experiment 1, in which there was a three-way interaction between weight, test phase, and condition for prediction responses. As before, children in both conditions had more accurate predictions for same-weight objects after the interventions. However, at posttest, children in the guidance with explanation condition had more accurate predictions for different-weight objects than the guidance only condition, which was not significant in Experiment 1. This comparison highlighted the independent role of correctly observing anomalous evidence and the additive effect of conceptual explanations in revising predictions that conflict with prior beliefs. Asking children for justifications did not result in more children changing their predictions, as the proportion of children who changed their minds in Experiments 1 (1.67%) and 2 (2.71%) were similar. Thus, we can rule out that the addition of justifications had an impact on children's predictions. Prompting children to justify their predictions appeared to account for the difference in results at posttest between the two guided conditions in Experiment 2, which was not apparent in Experiment 1.

Children's justification responses mirrored the prediction results, such that children in the guidance with explanation condition revised their understanding at a higher rate than those in the guidance only condition. A critical finding was that children's justifications for different-weight objects improved immediately following the intervention and were maintained after a delay. In contrast, children's justifications for same-weight objects only improved after a delay. Finally, there was a significant change from post- to delay test for both conditions. After a delay, children in the guidance with explanation condition demonstrated superior understanding, whereas children in the guidance only condition showed deterioration in understanding. An examination of the type of justifications that children provided supports this conclusion because there were higher frequencies of justifications that stated learned information in the guidance with explanation condition compared to the guidance only condition. These results collectively suggest that children in the guidance with explanation condition incorporated conceptually rich explanations about why objects fall at the same rate into their knowledge. This demonstrates that young children are capable of learning complex scientific explanations.

Furthermore, providing children with conceptual information not only impacted retention but also improved generalization to an analogous context: same-weight objects. Children in the guidance with explanation had competing explanations (i.e., whether weight or gravity affected the rate that the same-weight objects fell), and they likely discounted the former in favor of the latter. The mechanism that led to this change after a delay cannot be determined based on the current data. However, researchers have advocated that causal discounting of explanations may not necessarily require explicit comparison of the two competing explanations but may be more automatic and outside of awareness (Keil, 2006). When given processing time, children incorporated the evidence and explanation for anomalies (different-weight objects) and applied it to the conventional scenario (same-weight objects). This finding suggests that children coordinated competing theories, probably unintentionally.

General Discussion

This research examined the effect of embedding an explanation in a guided activity on young children's evaluation of anomalous evidence. Both experiments showed that guidance improved the observation and interpretation of anomalous evidence for both outcome measures: predictions and justification. The benefit of explanations was especially evident in Experiment 2 when children were asked to justify their predictions. Children generalized the new schema acquired from explanations from different- to same-weight objects and retained it long term when tested after a 1-week delay. As a result, children evaluated and learned from anomalous evidence more effectively when they received adequate guidance with conceptually rich explanations. Children in this study started with similar levels of prior beliefs[2] across conditions, and because the improvement in children's understanding was only found in the experimental conditions, but not the control, these findings speak to the effects of the guided conditions on children's learning. The benefits of guidance and explanations and the combined effect on children's long-term learning will be discussed below.

The Impact of Guidance

First, guidance improved children's learning from anomalous evidence. When children are confronted with evidence that conflicts with prior beliefs, they have difficulty restructuring their knowledge and achieving conceptual change because this process requires the coordination of evidence and theory. Conceptual change can be impeded at several stages. Children can: make incorrect observations (e.g., "the heavy one fell faster"); reinterpret the observation (e.g., "the heavy one is slower because it has glue sticking out"); reject the outcome on methodological grounds (e.g., "they weren't dropped at the same time") preventing generalization; or have difficulty with retention by subsequently reverting to prior beliefs (Chinn & Malhotra, 2002). Children's observations and interpretation of anomalous evidence improved more when they were guided than when they were required to produce the anomalous evidence in the baseline condition.

Guidance can scaffold the inquiry process and foster learning in several ways. Children can have difficulty producing and reflecting on the relevant concept if the task is too complex, even with materials closely related to the target concept. Adults can provide experimental support by setting up the environment so that children do not become overwhelmed by the stimuli and remain on task. The materials selected were uniform across as many dimensions as possible, so that children could focus on the more relevant features. In the fill and drop activity, the containers were identical boxes, and jars in the first two trials and the buckets were similar in size for the final trial. In the prediction with video activity, the object pairs were from the same category (e.g., both were blocks, balls, or animals) and the weight became more different with each trail. These nuances in the design helped children focus on the weight differences between the pairs.

Guidance may also reduce the cognitive load placed on children (Kirschner et al., 2018). Adults can support children in producing, observing, and interpreting anomalies, making the task appropriate.

---

[2] Preexisting knowledge was similar across both experiments. Children's predictions at pretest were similar for Experiments 1 and 2 across all conditions (*M*s = 0.38–0.41, *SD*s = .49).

Making sure that children make the correct observation and interpretation (i.e., both objects fall at the same time, even though one is heavier than the other), can help to reduce the demands of the task. Without guidance, children may not have enough cognitive resources to correctly observe the anomalies, successfully track the evidence, and integrate it with prior beliefs. With guidance, children can systematically track and assess anomalous evidence across several trials (Tolmie et al., 2016), which can emphasize the conflict between prior beliefs and evidence (Cheng & Brown, 2010; Zimmerman, 2007). Reflective support can also make the evidence more explicit, therefore, allowing children to use more cognitive resources to integrate the evidence with their beliefs and come to the correct interpretation of anomalous evidence.

Our findings show that the guidance only condition was more informative than the baseline condition which is a common approach in education settings where children need to construct their knowledge and theories on their own (R. Clark et al., 2012). Despite the immediate benefits of guidance, children did not retain this understanding after the delay. Furthermore, few had justifications that received maximum score and most of the reasons children provided were categorized as experiential or methodological. No children provided correct scientific justifications in the guidance only condition. Together, these findings show that only providing guidance during inquiry is insufficient for long-term understanding—especially about complex science concepts.

The Impact of Explanations

This research adds to a body of literature showing that explanations can facilitate conceptual change (Ganea et al., 2021; Kelemen et al., 2014; Kendeou et al., 2014; Larsen et al., 2020; Tippett, 2010; Venkadasalam & Ganea, 2018). In Experiment 2, children who learned that gravity affects the rate at which objects fall were better able to generalize and retain this understanding than children who did not hear an explanation. While anomalous evidence may have led to an immediate revision of prior beliefs, as seen in children's justification responses at posttest for the guidance only condition, this understanding was not retained long term, likely because the naïve theory remained entrenched. This idea is in line with research by Hardy et al. (2006), in which an extensive in-class intervention addressed beliefs about sinking and floating. Children who were given a more explicit comparison between conceptual information and naïve beliefs retained learned explanations at a higher rate than children who were not encouraged to contrast their naïve beliefs with the new information and evidence. These two instructional groups are analogous to the guidance with explanation and guidance only conditions used in the current research.

In this research, the experimental conditions only differed in the presence or absence of a scientific explanation with refutation of an incorrect belief. In other respects, the sessions were very similar in length, such that the guidance with explanation condition was only about a minute longer. Nevertheless, in the absence of explanations, children achieved only short-term learning. Future research should further investigate whether giving children different types of information about a domain (e.g., factual knowledge vs. explanations) would lead to similar improvements. We expect that such an effect might show in the justifications that children give, with children given explanations giving better justifications compared to children who receive more facts about a domain.

Research has shown that scientific explanations can aid in the restructuring of knowledge in several ways. First, explanations can highlight the difference between children's naïve beliefs and anomalous evidence. Students often fail to develop the accepted scientific theory on their own (Chinn & Brewer, 1993). Without a viable alternative theory, it is difficult for them to connect the evidence with prior beliefs (Cheng & Brown, 2010) because they often view these as isolated from each other (Zimmerman, 2007). As a result, there is no direct challenge to children's current beliefs, so they never reach a state of cognitive conflict (Posner et al., 1982). In such a circumstance, anomalous evidence serves both as a contradiction to children's naïve beliefs and as evidence for creating an alternative target theory, requiring children to consider a single event from multiple perspectives (Howe et al., 2013). Thus, providing an explanation likely aided the interpretation of anomalous evidence and helped children restructure their prior beliefs (Ganea et al., 2021; Potvin et al., 2015).

Furthermore, research has shown that even if children notice that anomalies are inconsistent with their prior beliefs, they do not reevaluate these beliefs unless the discrepancy is made explicit (Li & Klahr, 2005; Zimmerman, 2007). In the current work, explanations may have provided a level of meta-conceptual awareness that made coordinating evidence and theory more explicit (Kuhn, 2010; Zimmerman, 2007; Zimmerman & Klahr, 2018). In comparison, in the absence of an explanation, children in the guidance only condition may have created only a temporary justification, reverting to their prior beliefs a week later, and did not revise their explanation for the same-weight objects. Although children as young as age 3 offer causal accounts for physical events using physical causal relations (Keil, 2006), most of the children in the guidance only condition were not able to produce an alternative physical causal explanation on their own to account for the anomalous evidence. Finally, in Experiment 2, justifications revealed insight into children's understanding and a delay measure provided evidence of how this understanding changed over time. There were two key findings; first, children's retention of understanding varied for the guided conditions (i.e., after a delay children's justification scores increased in the guidance with explanation condition but decreased in the guidance only condition). That is, in the absence of an explanation, children's ability to reflect on anomalies and generate an ad hoc justification based solely on anomalous evidence was limited to the period immediately following the observations. Under these circumstances, children reasoned about the same- and different-weight objects independently and preserved their core belief that the rate of free fall is affected by weight. Without an explanation, it is likely that children were subtyping the different-weight objects into a separate category and only engaged in peripheral changes to their prior beliefs (Chinn & Brewer, 1993). This was supported by the fact that the number of children in the guidance with explanation condition who provided justifications with a maximum score (indicating the inclusion of correct scientific information) was nearly double compared to the guidance only condition. Furthermore, nearly half of the children in this condition had a reasoning that referenced the learned scientific information.

The second key finding was that in the guidance with explanation condition children's justifications improved for same-weight objects, but only after a delay. Immediately after the intervention, children in this condition began by treating same- and different-weight objects separately, similar to children's responses in the guidance only condition. They used the anomalies containing heavy and light objects to update their beliefs about different-weight objects without modifying their beliefs about same-weight objects. However, after a delay, children who heard an explanation applied their understanding broadly, indicating that they unified their beliefs by consolidating their schemas and achieved conceptual change (Hemmerich et al., 2016). The process of selecting the competing explanations may be an automatic process instead of an explicit one (Keil, 2006). Nonetheless, this finding supports the view that conceptual change is a gradual process of knowledge restructuring (Hardy et al., 2006; Shtulman & Lombrozo, 2016; Vosniadou, 2013) and points to the dual modality of knowledge that exists both as individual elements and theories (Brown, 2014; Hast, 2016). Initially, children seemed to be treating same- and different-weight objects independently (i.e., as fragmented pieces of knowledge: D. B. Clark, 2006; diSessa & Sherin, 1998). Given time to consolidate their understanding, children unified same- and different-weight objects under a single governing rule (i.e., a coherent theory: Carey, 2009; Vosniadou & Skopeliti, 2014). Although prior beliefs continue to coexist with scientific conceptions (Potvin & Cyr, 2017; Shtulman & Valcarcel, 2012), these findings are consistent with the notion that conceptual knowledge exists both in fragments and theories.

Limitations and Future Research

This study adds to the growing body of research on how to successfully foster learning from anomalous evidence. Research with older children suggests that 6- and 7-year-olds revise their beliefs about balancing objects after viewing anomalous evidence and engaging in exploratory play (Bonawitz et al., 2012). However, 5-year-olds only learned to balance objects when anomalous evidence was combined with explanations (Larsen et al., 2020). With more complex concepts, 5-year-olds could not learn from anomalous evidence presented in guided activities and only revised their beliefs about why some objects sink and float using explanations presented through picture books (Ganea et al., 2021). In contrast, children in this study revised their predictions from anomalies in a guided activity and even developed possible justifications at posttest; however, this was temporary because they reverted to their justifications scores similar to the pretest scores after a delay.

Differences between this study and previous findings may be due to the concept addressed. Balancing objects is a concept that children naturally learn around 8 years of age (Karmiloff-Smith & Inhelder, 1974) through an unambiguous outcome (an object balancing or not). In contrast, erroneous beliefs about free fall persist into adulthood (Kavanagh & Sneider, 2006), perhaps because observations of falling objects can be ambiguous (Chinn & Malhotra, 2002). Furthermore, this belief is perpetuated by the familiar bowling ball and feather exemplar. Weight is a salient difference between a bowling ball and a feather and supports the argument that heavy objects fall faster. However, it is not weight but an invisible force—air resistance—that slows the feather down as it descends. We simplified the concept of free fall by using identical objects for the test phase and objects similar in size during the intervention to reduce the effect of air resistance. Thus, children did not have to evaluate an object's aerodynamics and could focus on the critical variable: weight.

However, isolating children's focus on one variable is impossible for all concepts. For instance, with the concept of sinking and floating, weight and density are conflated, with no viable means of isolating these variables. As in the current study, previous research showed that kindergarten students could understand predictions and effects in single-variable experiments (Siegler & Chen, 1998) but have difficulty when the number of variables increases (Ganea et al., 2021; Larsen et al., 2020). Indeed, anomalous evidence may not be salient enough to promote any belief revision for complex concepts. Future studies are needed to further examine children's ability to update their beliefs because the conditions under which children restructure their knowledge can vary across domains and concepts. Additionally, an important avenue for future research is to determine developmentally appropriate interventions over a longer period to improve children's scientific understanding in naturalistic settings.

One final consideration is that the guidance only condition did not have any filler information for the conceptual information (brief introduction, refutation of the incorrect idea, and explanation about gravity) that was added to the guidance with explanation condition. Future studies on the effects of different kinds of information on children's conceptual knowledge can examine more comparative conditions such as adding a circular explanation or implying it to be a well-known fact (e.g., "everyone knows that objects fall at the same time even heavier ones"). Prior research (Danovitch & Mills, 2018) suggests that even 5-year-olds would find a circular explanation less useful when pitted against a good explanation, but would they learn from it when combined with guidance? Or would simply telling them what everyone believes prompt them to revise their belief in the face of anomalous evidence? The current research shows that the simple delivery of conceptual information by a trusted source, such as a teacher or a book, can help children learn complex concepts that run counter to their beliefs and help them retain this understanding. This research together with a body of prior work (Chinn & Malhotra, 2002; Ganea et al., 2021; Kelemen et al., 2014; Larsen et al., 2020; Tippett, 2010; van den Broek & Kendeou, 2008; Venkadasalam & Ganea, 2018) highlights the importance of adding conceptual information to instructional strategies focused on conceptual change.

Conclusion

The interrelation between theory and evidence is complex. Mature scientific thinking develops slowly beginning in early childhood (Sandoval et al., 2014) and reflects the consideration of patterns of evidence, making judgments based on existing knowledge, as well as knowledge about causal mechanisms, including alternative explanations (Koslowski, 1996). Individuals need to evaluate the plausibility between their own prior beliefs, the evidence they witness, and alternative explanations to build a cognitive representation of a concept (Chinn & Brewer, 2001). Our research expands on the current understanding of children's responses to and interpretations of anomalous evidence, by showing that children are more likely to update their prior beliefs in the face of anomalous evidence when such evidence is presented with an alternative explanatory framework. Children are also capable of extrapolating information from patterns of anomalous evidence, particularly when the evidence is supported with adult guidance as in this research; however, their ability to do so long term is challenged in the absence of conceptual information. These findings add to the growing body of research regarding evidence evaluation and the conditions under which children engage in conceptual change.

Our work supports advocacy for early science education and science literacy (Bowman et al., 2001; Duschl et al., 2007; Eshach & Fried, 2005; Gelman & Brenneman, 2004; Morgan et al., 2016).

Science knowledge in kindergarten is the strongest predictor of knowledge in Grade 1, which in turn is the strongest predictor of science achievement from Grades 3 to 8 (Morgan et al., 2016). A recent meta-analysis has shown that individual differences in knowledge are stable throughout learning, and prior knowledge is essential for later performance, by emphasizing the long-term nature of knowledge acquisition (Simonsmeier et al., 2021). This supports the view that conceptual change is a gradual process of knowledge restructuring (Shtulman & Valcarcel, 2012; Vosniadou, 2013), and the current findings suggest that belief change can begin early in development with adequate support. Orienting children's attention toward basic physical concepts in the early years will lay the foundation for more advanced understanding throughout their development (Hardy et al., 2006). Although prior beliefs continue to coexist into adulthood, especially with this particular concept (i.e., falling objects), disrupting the reliance on incorrect intuitive beliefs can benefit long-term understanding and conceptual change.

Adequate support is required to direct children's attention to accurate scientific theories in a developmentally appropriate way. This study shows that prior beliefs can be addressed through scaffolds, such as guided presentation of anomalous evidence and the addition of conceptual information. When adults scaffold the process by guiding children to make the correct observations that counter the child's beliefs, they make the activities age-appropriate and foster pedagogical objectives while allowing the child to maintain agency over their activities (Weisberg et al., 2016; Yu et al., 2018). Conceptual explanations provide an alternative theory for children to help explain anomalous evidence, leading to increased domain knowledge and scientific literacy. This is important since research shows that children's knowledge in the preschool years is one of the strongest predictors of their later science learning (Morgan et al., 2016) and that their prior knowledge constrains their interpretation of evidence (Bonawitz et al., 2012; Penner & Klahr, 1996). Thus, addressing prior beliefs and beginning to build a strong foundation of knowledge in the early years is imperative.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bonawitz, E. B., van Schijndel, T. J. P., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*(4), 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002

Bowman, B. T., Donovan, M. S., & Burns, M. S. (2001). *Eager to learn: Educating our preschoolers* (Issue 5). National Academy Press. https://doi.org/10.17226/9745

Brault Foisy, L. M., Potvin, P., Riopel, M., & Masson, S. (2015). Is inhibition involved in overcoming a common physics misconception in mechanics? *Trends in Neuroscience and Education*, *4*(1–2), 26–36. https://doi.org/10.1016/j.tine.2015.03.001

Brewer, W. F., Chinn, C. A., & Samarapungavan, A. (1998). Explanation in scientists and children. *Minds and Machines*, *8*(1), 119–136. https://doi.org/10.1023/A:1008242619231

Brown, D. E. (2014). Students' conceptions as dynamically emergent structures. *Science & Education*, *23*(7), 1463–1483. https://doi.org/10.1007/s11191-013-9655-9

Butts, D. P., Hofman, H., & Anderson, M. (1993). Is hands-on experience enough? A study of young children's views of sinking and floating objects. *Journal of Elementary Science Education*, *5*(1), 50–64. https://doi.org/10.1007/BF03170644

Carey, S. (2009). The origin of concepts—Beyond core cognition: Natural number. *Journal of Cognition and Development*, *1*(1), 37–41. https://doi.org/10.1207/S15327647JCD0101N_3

Cheng, M.-F., & Brown, D. E. (2010). Conceptual resources in self-developed explanatory models: The importance of integrating conscious and intuitive knowledge. *International Journal of Science Education*, *32*(17), 2367–2392. https://doi.org/10.1080/09500690903575755

Chinn, C. A., & Brewer, W. F. (1992). Psychological responses to anomalous data. In J. K. Kruschke (Ed.), *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 165–170). Lawrence Erlbaum Associates.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*(1), 1–49. https://doi.org/10.3102/00346543063001001

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching*, *35*(6), 623–654. https://doi.org/10.1002/(SICI)1098-2736(199808)35:6<623::AID-TEA3>3.0.CO;2-O

Chinn, C. A., & Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction*, *19*(3), 323–393. https://doi.org/10.1207/S1532690XCI1903_3

Chinn, C. A., & Malhotra, B. A. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology*, *94*(2), 327–343. https://doi.org/10.1037/0022-0663.94.2.327

Clark, D. B. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. *Cognition and Instruction*, *24*(4), 467–563. https://doi.org/10.1207/s1532690xci2404_3

Clark, R., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning: The case for fully guided instruction. *American Educator*, *36*(1), 5–11. https://eric.ed.gov/?id=EJ971752

Danovitch, J. H., & Mills, C. M. (2018). Understanding when and how explanation promotes exploration. In M. Saylor, P. A. Ganea (Eds.), *Active learning from infancy to childhood* (pp. 95–112). Springer International Publishing. https://doi.org/10.1007/978-3-319-77182-3_6

diSessa, A. A., & Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education*, *20*(10), 1155–1191. https://doi.org/10.1080/0950069980201002

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education*, *25*(6), 671–688. https://doi.org/10.1080/09500690305016

Duit, R., Treagust, D. F., & Widodo, A. (2013). Teaching science for conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (2nd ed., pp. 487–503). Routledge.

Duschl, R. A., Richard, A., Schweingruber, H. A., Shouse, A. W., National Research Council (U.S.). Committee on Science Learning, K. T. E. Grade, National Research Council (U.S.). Board on Science Education, & National Research Council (U.S.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. National Academies Press.

Eshach, H., & Fried, M. N. (2005). Should science be taught in early childhood? *Journal of Science Education and Technology*, *14*(3), 315–336. https://doi.org/10.1007/s10956-005-7198-9

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching. *Review of Educational Research*, *82*(3), 300–329. https://doi.org/10.3102/0034654312457206

Ganea, P. A., Larsen, N. E., & Venkadasalam, V. P. (2021). The role of alternative theories and anomalous evidence in children's scientific belief revision. *Child Development*, *92*(3), 1137–1153. https://doi.org/10.1111/cdev.13481

Gelman, R., & Brenneman, K. (2004). Science learning pathways for young children. *Early Childhood Research Quarterly*, *19*(1), 150–158. https://doi.org/10.1016/j.ecresq.2004.01.009

Gershman, S. J. (2019). How to never be wrong. *Psychonomic Bulletin & Review*, *26*(1), 13–28. https://doi.org/10.3758/S13423-018-1488-8

Hardy, I., Jonen, A., Möller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking". *Journal of Educational Psychology*, *98*(2), 307–326. https://doi.org/10.1037/0022-0663.98.2.307

Hast, M. (2014). Exploring the shift in children's incline motion predictions: Fragmentation and integration of knowledge as possible contributors. *Journal of Educational and Developmental Psychology*, *4*(2), 74–81. https://doi.org/10.5539/jedp.v4n2p74

Hast, M. (2016). Children's reasoning about rolling down curves: Arguing the case for a two-component commonsense theory of motion. *Science Education*, *100*(5), 837–848. https://doi.org/10.1002/sce.21237

Hemmerich, J. A., Van Voorhis, K., & Wiley, J. (2016). Anomalous evidence, confidence change, and theory change. *Cognitive Science*, *40*(6), 1534–1560. https://doi.org/10.1111/cogs.12289

Howe, C., Devine, A., Taylor Tavares, J., & Tavares, J. T. (2013). Supporting conceptual change in school science: A possible role for tacit understanding. *International Journal of Science Education*, *35*(5), 864–883. https://doi.org/10.1080/09500693.2011.585353

Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition*, *3*(3), 195–212. https://doi.org/10.1016/0010-0277(74)90008-0

Kavanagh, C., & Sneider, C. (2006). Learning about gravity I. Free fall: A guide for teachers and curriculum developers. *Astronomy Education Review*, *5*(2), 21–52. https://doi.org/10.3847/AER2006018

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, *57*(1), 227–254. https://doi.org/10.1146/annurev.psych.57.102904.190100

Kelemen, D., Emmons, N. A., Seston Schillaci, R., & Ganea, P. A. (2014). Young children can be taught basic natural selection using a picture-storybook intervention. *Psychological Science*, *25*(4), 893–902. https://doi.org/10.1177/0956797613516009

Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014). Knowledge revision processes in refutation texts. *Discourse Processes*, *51*(5–6), 374–397. https://doi.org/10.1080/0163853X.2014.913961

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano, R. J. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning*, *13*(2), 213–233. https://doi.org/10.1007/s11412-018-9277-y

Klahr, D., & Li, J. (2005). Cognitive research and elementary science instruction: From the laboratory, to the classroom, and back. *Journal of Science Education and Technology*, *14*(2), 217–238. https://doi.org/10.1007/s10956-005-4423-5

Kloos, H., & Van Orden, G. C. (2005). Can a preschooler's mistaken belief benefit learning? *Swiss Journal of Psychology*, *64*(3), 195–205. https://doi.org/10.1024/1421-0185.64.3.195

Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, *64*(3), 141–152. https://doi.org/10.1024/1421-0185.64.3.141

Koslowski, B. (1996). Theory and evidence: The development of scientific reasoning. In *Choice reviews online* (Vol. 34, No. 7). MIT Press. https://doi.org/10.5860/choice.34-3823

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*(4), 674–689. https://doi.org/10.1037/0033-295X.96.4.674

Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist*, *42*(2), 109–113. https://doi.org/10.1080/00461520701263376

Kuhn, D. (2010). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 497–523). Wiley-Blackwell. https://doi.org/10.1002/9781444325485.ch19

Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development*, *1*(1), 113–129. https://doi.org/10.1207/S15327647JCD0101N_11

Lane, J. D., & Harris, P. L. (2014). Confronting, representing, and believing counterintuitive concepts: Navigating the natural and the supernatural. *Perspectives on Psychological Science*, *9*(2), 144–160. https://doi.org/10.1177/1745691613518078

Larsen, N. E., Venkadasalam, V. P., & Ganea, P. A. (2020). Prompting children's belief revision about balance through primary and secondary sources of evidence. *Frontiers in Psychology*, *11*, Article 1503. https://doi.org/10.3389/fpsyg.2020.01503

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, *86*(3), 681–718. https://doi.org/10.3102/0034654315627366

Li, J., & Klahr, D. (2005). The psychology of scientific thinking: Implications for science teaching and learning. In J. Rhoton & P. Shane (Eds.), *Teaching science in the 21st century* (pp. 307–328). NSTA Press.

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. https://doi.org/10.3102/0013189X16633182

Penner, D. E., & Klahr, D. (1996). The interaction of domain-specific knowledge and domain-general discovery strategies: A study with sinking objects. *Child Development*, *67*(6), 2709–2727. https://doi.org/10.2307/1131748

Pine, K., Messer, D., & St. John, K. (2001). Children's misconceptions in primary science: A survey of teachers' views. *Research in Science & Technological Education*, *19*(1), 79–96. https://doi.org/10.1080/02635140120046240

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66*(2), 211–227. https://doi.org/10.1002/sce.3730660207

Potvin, P., & Cyr, G. (2017). Toward a durable prevalence of scientific conceptions: Tracking the effects of two interfering misconceptions about buoyancy from preschoolers to science teachers. *Journal of Research in Science Teaching*, *54*(9), 1121–1142. https://doi.org/10.1002/tea.21396

Potvin, P., Sauriol, É., & Riopel, M. (2015). Experimental evidence of the superiority of the prevalence model of conceptual change over the classical models and repetition. *Journal of Research in Science Teaching*, *52*(8), 1082–1108. https://doi.org/10.1002/tea.21235

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Renken, M. D., & Nunez, N. (2010). Evidence for improved conclusion accuracy after reading about rather than conducting a belief-inconsistent simple physics experiment. *Applied Cognitive Psychology*, *24*(6), 792–811. https://doi.org/10.1002/acp.1587

Sandoval, W. A., Sodian, B., Koerber, S., & Wong, J. (2014). Developing children's early competencies to engage with science. *Educational Psychologist*, *49*(2), 139–152. https://doi.org/10.1080/00461520.2014.917589

Shtulman, A. (2017). Scienceblind: Why our intuitive theories about the world are so often wrong. *Science*, *356*(6336), Article 385. https://doi.org/10.1126/science.aan4200

Shtulman, A., & Lombrozo, T. (2016). Bundles of contradiction: A coexistence view of conceptual change. In D. Barner & A. S. Baron (Eds.), *Core knowledge and conceptual change* (pp. 53–71). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190467630.003.0004

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*(2), 209–215. https://doi.org/10.1016/j.cognition.2012.04.005

Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analysis. *Cognitive Psychology*, *36*(3), 273–310. https://doi.org/10.1006/cogp.1998.0686

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2021). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist*, *57*(1), 31–54. https://doi.org/10.1080/00461520.2021.1939700

Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational Psychologist*, *42*(2), 115–121. https://doi.org/10.1080/00461520701263426

Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education*, *8*(6), 951–970. https://doi.org/10.1007/s10763-010-9203-x

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. https://doi.org/10.1111/bjep.12119

Tullos, A., & Woolley, J. D. (2009). The development of children's ability to use evidence to infer reality status. *Child Development*, *80*(1), 101–114. https://doi.org/10.1111/j.1467-8624.2008.01248.x

van den Broek, P., & Kendeou, P. (2008). Cognitive processes in comprehension of science texts: The role of co-activation in confronting misconceptions. *Applied Cognitive Psychology*, *22*(3), 335–351. https://doi.org/10.1002/acp.1418

Venkadasalam, V. P., & Ganea, P. A. (2018). Do objects of different weight fall at the same time? Updating naive beliefs about free-falling objects from fictional and informational books in young children. *Journal of Cognition and Development*, *19*(2), 165–181. https://doi.org/10.1080/15248372.2018.1436058

Venkadasalam, V. P., Larsen, N. E., & Ganea, P. A. (2023). *Promoting scientific understanding and conceptual change in young children using explanations and guidance*. https://doi.org/10.17605/OSF.IO/MA3RT

Vorholzer, A., & von Aufschnaiter, C. (2019). Guidance in inquiry-based instruction—An attempt to disentangle a manifold construct. *International Journal of Science Education*, *41*(11), 1562–1577. https://doi.org/10.1080/09500693.2019.1616124

Vosniadou, S. (2013). Conceptual change in learning and instruction: The framework theory approach. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 11–30). Routledge. https://doi.org/10.4324/9780203154472

Vosniadou, S., & Skopeliti, I. (2014). Conceptual change from the framework theory side of the fence. *Science & Education*, *23*(7), 1427–1445. https://doi.org/10.1007/s11191-013-9640-3

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

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

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223. https://doi.org/10.1016/j.dr.2006.12.001

Zimmerman, C., & Klahr, D. (2018). Development of scientific thinking. In J. T. Wixted (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience* (4th ed., pp. 1–25). John Wiley & Sons. https://doi.org/10.1002/9781119170174.epcn407

Received May 26, 2022
Revision received September 30, 2023
Accepted October 12, 2023 ■