

Learning the control-of-variables strategy during an informal science lesson on popping popcorn

Bradley J. Morris, Jacob Cason, Katie Asaro, Yin Zhang, Michelle Rivers, Whitney Owens & John Dunlosky

To cite this article: Bradley J. Morris, Jacob Cason, Katie Asaro, Yin Zhang, Michelle Rivers, Whitney Owens & John Dunlosky (08 Feb 2024): Learning the control-of-variables strategy during an informal science lesson on popping popcorn, International Journal of Science Education, DOI: [10.1080/09500693.2024.2310526](https://doi.org/10.1080/09500693.2024.2310526)

To link to this article: <https://doi.org/10.1080/09500693.2024.2310526>



View supplementary material [↗](#)



Published online: 08 Feb 2024.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



Learning the control-of-variables strategy during an informal science lesson on popping popcorn

Bradley J. Morris ^a, Jacob Cason ^b, Katie Asaro ^a, Yin Zhang ^a, Michelle Rivers ^c, Whitney Owens ^b and John Dunlosky ^d

^aLearning Sciences and Educational Psychology, Kent State University, Kent, OH, USA; ^bCincinnati Museum Center, Cincinnati, OH, USA; ^cDepartment of Psychology, Texas Christian University, Fort Worth, TX, USA; ^dPsychological Sciences, Kent State University, Kent, OH, USA

ABSTRACT

Understanding experimental design (e.g. control of variable strategy or CVS) is foundational for scientific reasoning. Previous research has demonstrated that demonstrations with cognitive conflict (e.g. asking students to evaluate and explain different experimental designs) are effective in promoting children's scientific reasoning, however, the implementation of this approach often requires significant instructional time and resources. This study reports the impact of a brief, scalable intervention on one component of scientific reasoning, understanding experimental design, by providing brief instruction on the control-of-variable strategy (CVS), embedded in a food science activity (popping popcorn). Three-hundred and seven (307) 3rd-5th graders in the midwestern US participated in either a CVS intervention or a demonstration on the science of popcorn without a CVS intervention. Performance on a pre-activity test (involving identification of good and bad experiments) did not differ between conditions. By contrast, post-activity performance was significantly greater for classes who received the CVS intervention. Thus, a brief discussion of the CVS embedded within a food-science demonstration can have a meaningful impact on children's understanding of conducting a quality experiment. Our results demonstrate the efficacy of a simple, low-cost intervention for CVS that is potentially scalable.

ARTICLE HISTORY

Received 30 May 2023
Accepted 23 January 2024

KEYWORDS

Cognitive development;
control of variables strategy;
informal STEM learning;
classroom-based
interventions; scalable
interventions

Introduction

Science education is critically important for twenty-first century education, science literacy, and STEM careers (NASEM, 2018). Most young children begin their educational journeys interested in and curious about science (Jirout, 2020), however, this interest and engagement often declines by middle school (Bonnette et al., 2019). Some students who disengage from science begin to believe that they are not suited for science inquiry or science careers (Vincent-Ruz & Schunn, 2021), which often reduces the likelihood of obtaining a high paying STEM career (Goff et al., 2020). In addition, students who disengage in science during childhood are less likely to be scientifically literate adults who

CONTACT Bradley J. Morris  bmorri20@kent.edu

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/09500693.2024.2310526>.

© 2024 Informa UK Limited, trading as Taylor & Francis Group

engage with, understand, and apply science in their lives as adults (Howell & Brossard, 2021; NASEM, 2016; Sharon & Baram-Tsabari, 2020). One way to support children's emergent scientific reasoning is through informal activities that provide authentic learning opportunities that connect science to children's everyday experience (Rogoff, Dahl, & Callanan, 2018). This is particularly important for helping children acquire foundational scientific reasoning skills, such as how to design a good (i.e. unconfounded) experiment, which provides further benefits for acquiring science knowledge (Klahr & Nigam, 2004). A key concept for understanding experimental design is the Control of Variable strategy (hereafter CVS), in which only one variable at a time is compared in a 'good' experiment (Schwichow et al., 2016; Schwichow, Osterhaus, & Edelsbrunner, 2020). This paper describes an experiment in which 7–12 year old children participated in an informal science activity to set up and recognise good experiments in the familiar context of popping popcorn. This paper describes a study that leveraged the children's curiosity and interest in a familiar cooking activity, popping popcorn, to help them learn about setting up and recognising good (i.e. unconfounded) scientific experiments. Our approach was novel because (a) the activity was constructed to include two evidence-based instructional factors, demonstrations and creating cognitive conflict, to support learning and transfer (Schwichow et al., 2016), (b) the activity provides learning opportunities through a familiar, everyday activity, cooking, and (c) unlike most CVS instruction, the activity could be conducted in large groups in a variety of settings such as intact classrooms, community centres, or museums.

Scientific reasoning

Scientific reasoning is a specific type of information seeking that includes science processes (e.g. setting up and conducting experiments, engaging in data analysis) as well as science concepts (e.g. scientific theories and concepts; Schwichow et al., 2020), which are deeply connected in authentic science practice (Klahr et al., 2011). Children engage in scientific reasoning through learning experiences that can be formal (e.g. lessons in science classrooms) and informal (e.g. activities in museums or science camps). These learning experiences assist children in the acquisition of the socio-cultural tools of science (e.g. science terminology, knowledge of experimental design, etc.) (Morris, Croker, Masnick, & Zimmerman, 2012).

Young children engage in information seeking that is rooted in their curiosity about the world around them (Jirout, 2020). This curiosity is foundational to early scientific reasoning skills that begin with remarkably sophisticated cognitive abilities and information seeking strategies that allow children to learn about and make sense of their world (Weisberg & Sobel, 2022). Cognitive abilities such as curiosity, generating hypotheses, and seeking evidence underlie early learning and conceptual development (Legare, 2014), in stark contrast to early theories of cognitive development that suggested that such skills were only available later in adolescence (Inhelder & Piaget, 1958). Many components of scientific reasoning, including generating hypotheses, setting up good experiments, and even seeking disconfirming evidence are spontaneously generated by young children (Weisberg & Sobel, 2022).

While early scientific reasoning is fostered through everyday life and open-ended questions, formal scientific reasoning is characterised by logic and a systematic approach

to validate conclusions (Teig & Scherer, 2016). Formal scientific reasoning that is practiced within scientific communities also begins within the surprisingly powerful cognitive capacities demonstrated by young children as they explore their world (Köksal-Tuncer & Sodian, 2018; Legare, 2014; Weisberg & Sobel, 2022). These cognitive abilities provide the foundation through which children acquire formal scientific reasoning skills (Klahr et al., 2011; Morris et al., 2012; Zimmerman, 2007).

As noted above, young children seek information to test hypotheses using informative comparisons and can draw valid inferences from these comparisons (Köksal-Tuncer & Sodian, 2018). In the domain of causal reasoning, even young children use informative testing to determine causal relations between variables (Sobel & Kirkham, 2006) and diagnostic causal reasoning (i.e. reasoning from effects to causes; Fernbach & Erb, 2013). It is important to note that these early scientific reasoning skills have limits. For example, children often fail to test combinations of variables or attend to more informative variables over less informative ones (Köksal-Tuncer & Sodian, 2018). Additionally, children sometimes have difficulty testing hypotheses in novel contexts suggesting difficulty with transfer (Chen & Klahr, 1999) and when drawing inferences (Schwchow et al., 2020). These limitations suggest that, although powerful, children's scientific reasoning benefits from, and becomes formalised through, learning opportunities and instruction on systematic approaches to science.

One example of the formalisation of scientific reasoning is learning the logic of experimentation or setting up unconfounded experiments. The process of setting up an unconfounded experiment is foundational for testing scientific hypotheses and having a firm grasp of the logic of experimentation is important in understanding how to obtain evidence, evaluate the quality of evidence, and draw conclusions from experiments and data (Klahr, 2000). For children, the ability to recognise good and bad experiments is a key strategy for understanding a fundamental process in science: formal information seeking through experimentation (Zimmerman, 2007). While children have intuitions about setting up good experiments (Bramley et al., 2022) and even young children can learn how to construct simple, unconfounded experiments (van der Graaf et al., 2015), children sometimes err in applying CVS in the absence of instruction until later in childhood (Schwchow et al., 2016, p. 2020). Knowledge of experimentation, often measured using CVS, is closely tied to science content knowledge (Koerber & Osterhaus, 2019) because knowledge of relations between variables allows for a more accurate knowledge of what might matter or might not matter in an experiment (Schwchow et al., 2020). In sum, knowledge of the logic of good experimental design is an important contributing factor in children's emerging scientific reasoning.

Knowledge of good experimentation provides a powerful tool through which students generate, test, and evaluate their emerging scientific knowledge (Klahr & Nigam, 2004). The logic of experimentation is illustrated in the foundational idea in which only one thing is varied at a time. The significance of the control of variables in a scientific experiment is highlighted in the Next Generation Science Standards (NGSS, Lead States, 2013). Planning and carrying out sound investigations is one of the essential eight practices identified in the NGSS science and engineering curriculum (National Research Council, 2012), which includes identifying the dependent and independent variables and the control of variables as a key element.

As reviewed above, children spontaneously exhibit the rudiments of CVS, particularly under conditions in which the number of choices are constrained and feedback about outcomes are provided (Köksal-Tuncer & Sodian, 2018). Next, we detail two key factors in CVS interventions that have been effective in helping students acquire understanding of the process. CVS interventions have demonstrated that young children can learn to recognise and set up unconfounded experiments (Klahr & Nigam, 2004; Martella, Klahr, & Li, 2020; Schwichow et al., 2016). These interventions have produced gains in learning through direct instruction (Klahr & Nigam, 2004), inquiry-based learning (Schalk et al., 2019), or a combination of both (Lorch et al., 2010, 2014; Martella et al., 2020; Matlen & Klahr, 2013). An important question is which instructional factors produce the largest improvements? Two key factors were identified in a recent meta-analysis: the use of demonstrations and creating cognitive conflict (Schwichow et al., 2016). Cognitive conflict is theorised to produce its effect through the acquisition of new knowledge, establishing connections to existing knowledge, and by increasing metacognitive awareness of both (Vosniadou, 2019). This is consistent with the framework theory account (Vosniadou & Skopeliti, 2014) that explains how early intuitions about experiments are integrated with the cultural tools of science through cognitive mechanisms (e.g. metacognitive awareness) that are themselves developing (see also Morris et al., 2012; Zohar & Peled, 2008). Demonstrations include visual illustrations of the experimental materials and their relations (e.g. Padilla et al., 1984) often using either hands-on manipulatives (e.g. ramps, Strand-Cary & Klahr, 2008) or computer-based icons (Lazonder & Egberink, 2014). Cognitive conflict is created when students are asked to evaluate, identify, or correct the error in learning materials (Schwichow et al., 2016). For example, in one CVS intervention, children were shown a confounded setup between two ramps and were asked to identify why it would not lead to reliable results (e.g. Strand-Cary & Klahr, 2008).

CVS interventions that engage an entire classroom of students at the same time are critical for scaling instructional approaches for use in both informal settings such as museums and in formal settings such as schools because instructional time, attention, and material resources are limited. For example, individualised instruction or specialised equipment might produce learning gains but would unlikely be scalable, particularly in settings with many learners. So, a key question becomes, can a relatively brief instruction presented to a group, such as an entire classroom, produce CVS learning? Consider outcomes from a few experiments that have used a single instructional session for CVS training. In one investigation, eleven-year-old students who received a single, 20 min of direct instruction session on CVS improved their performance but did not transfer this knowledge to novel contexts (Lazonder & Egberink, 2014). In another investigation, sixth graders with hands-on experiences with ramps improved at posttest performance when instruction was direct but not when students had to discover relations on their own (Wagensveld et al., 2015). Although this intervention required only 20 min, each child was individually trained, which may not be feasible in many classrooms. In a recent comparison of different instantiations of active learning, students were assigned to small groups for 40 min instructional session on CVS (Martella et al., 2020). While all conditions showed some improvement, students who first saw CVS modelled and then were given direct instruction showed the largest gains in performance at posttest. Finally, during 50 min of training, eleven-year-olds who were given direct CVS instruction and promoted to reflect on their own thinking outperformed children given only

direct instruction (Dejonckheere et al., 2011). These and other investigations (for a review, see Schwichow et al., 2016) demonstrate the potential power of CVS interventions, but it may be challenging to scale up their use when instructors do not have larger blocks of time to focus on CVS training or cannot deliver one-on-one instruction to their students.

A related question is the extent to which training on CVS would transfer to a different context. Transfer, or the use of knowledge or skill in a novel context, is a notoriously difficult problem in education because it requires learners to identify relational similarities between two objects, situations, or contexts (Barnett & Ceci, 2002; Jee et al., 2022). We focused on near transfer characterised by identifying similarities in structure across similar contexts rather than far transfer or transfer across dissimilar contexts (Barnett & Ceci, 2002). One way to help students transfer information is to provide multiple learning opportunities that support structural alignment that highlights similarities between examples (Kurtz et al., 2013). Previous research has demonstrated that even young children could learn and transfer the CVS strategy to novel domains (Chen & Klahr, 1999); when transfer was facilitated by presenting multiple examples that highlighted the task and contextual similarity (Klahr & Chen, 2011). Using multiple examples and providing multiple opportunities for learning is the essence of spaced practice, a powerful, evidence-based learning strategy (Rawson & Dunlosky, 2022). Thus, transfer may be difficult for students, particularly young students, who receive only one, brief instructional session, however, including multiple examples in this session might provide some support.

Embedding brief CVS training into an informal-science presentation

Although formal science lessons in schools are a critically important part of a strong science education, informal science activities provide an opportunity to enhance and broaden children's learning and scientific reasoning (Behrendt & Franklin, 2014; Martin et al., 2016). Informal activities provide slightly different learning spaces than formal activities because informal activities are not graded, are guided by student interests, and often connect content more meaningfully to their everyday lives (Staus et al., 2021). In a comparison of sixth graders in formal and informal science settings, students were more likely to report intrinsic motivation for learning as their autonomy increased (Salmi & Thuneberg, 2019). Thus, informal activities provide important learning opportunities and can also provide a powerful complement to formal learning within school settings (Asghar, 2012). What follows is a brief preview of our approach.

Accordingly, in the present research, we evaluated the impact of a brief CVS intervention that demonstrated the CVS while promoting cognitive conflict that was resolved through discussion. This intervention was developed to engage an entire class and was embedded in a science (of popcorn) demonstration that could be adapted and used in the context of other demonstrations. For each class, the demonstration lasted about 30 min, with the CVS intervention requiring about 5–10 min. For the demonstration, one Subject-Matter-Expert (SME) leader (Author 2) described the origins of popcorn and asked students why popcorn popped. He then discussed the science of why popcorn pops and had students yell out their favourite toppings. Then, the second leader (Author 7) showed students how to make popcorn on a stove top and asked them to

explain why oil was used. Then, while the popcorn was popping, the CVS intervention began (for classes assigned to the CVS group) with the first two experimental questions. We provide details of the procedure in the Method section below.

Importantly, note that despite being brief, the classroom discussion about conducting good experiments (to answer questions about cooking popcorn) involved the key components of effective CVS interventions (Schwichow et al., 2016). In particular, the discussions involved interacting with experimental problems to engage students in evaluating each solution (which would produce conflict when students answered incorrectly) followed by feedback about the correct answer and the CVS principle.

Research questions

We propose four research questions. One, will a single, classroom-based CVS intervention that includes demonstrations and cognitive conflict increase correct identification of good or bad experiments at post-test compared to scores at pre-test? Two, will this intervention also increase metacognitive awareness of CVS, measured by self-reported confidence. Three, will this intervention promote effective transfer of CVS knowledge from the context of instruction (i.e. cooking) to a novel context? Four, will the effects of the intervention be similar across ages, specifically US 3rd, 4th, and 5th graders (roughly 7–10 year-olds). We note that the data for the fourth question is underpowered to provide reliable results due to our sampling limitations, however, we can provide preliminary results.

These questions and our approach are novel for four reasons. One, the activity builds in three evidence-based principles for learning and transferring CVS knowledge: demonstrations, cognitive conflict, and spaced practice (Rawson & Dunlosky, 2022). As reviewed above, these approaches provide high levels of guidance that help students understand the nature of the problem itself, potential solutions, and feedback on their efficacy. Two, most interventions have either focused on individual-level instruction, thus creating and testing effective group-level interventions are more closely aligned with authentic instruction in classrooms or public settings such as museums. Three, we used an everyday, familiar cooking activity, popping popcorn, to provide an engaging learning context. Finally, our informal activity is scalable in that it is low-cost, requires little training to implement, can be used in group settings, and is brief in duration.

Secondary questions

Besides evaluating the main prediction discussed above, we also conducted several exploratory analyses to address some questions that the research was not specifically designed to address. In particular, we recruited from local elementary schools with the aim of including as many classrooms as possible; in doing so, classes comprising a range of grade levels (from third to fifth grade) participated. Although the participating classes were over-represented by the fourth grade, we had enough classes in each grade level to conduct exploratory analyses of grade level on CVS performance. Also, we present exploratory analyses of confidence judgments to evaluate whether the CVS intervention also boosts students' confidence in their answers.

Method

Design and participants

The study was approved by the Kent State University IRB and by the administrators and teachers in participating schools. Following 45 CFR 46.101(b)(2), parental consent was waived because data collection occurred in existing classrooms, with teachers present, and no identifying information was collected. Parents were notified about the project in advance and were invited to contact the first author with any questions or concerns and were given the opportunity to opt their child out of the project. Children were given a unique code that was used to identify their data at pre- and post-test to ensure anonymity.

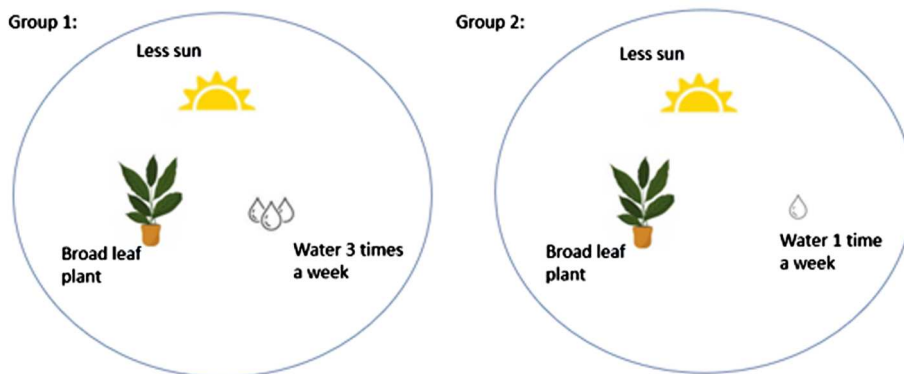
Seventeen (17) classes from the third (5 classes), fourth (9) and fifth (3) grade were recruited from local elementary schools. Classes were randomly assigned to the training groups. One group saw a popcorn activity alone (hereafter control group) or an activity with CVS intervention (hereafter experimental group). Given that there were an odd number of classes per grade, we assigned the extra class for each grade to receive the CVS intervention. Three hundred and seven (307) students completed the pre-activity test, and three-hundred and fifteen (315) completed the post-activity test. Note that the pre- and post-activity responses for each student were matched based on a unique code that each student generated – in some cases, it was evident that a student did not generate the same code, so their pre- and post-activity tests could not be matched. Given that some students were absent for one of the two sessions, we could not estimate how many cases where matches did not occur were due to an absenteeism or due to providing different codes in the two sessions. For analyses, all responses were used to estimate mean (descriptive) values, but non-matched data were automatically dropped from repeated-measures inferential analyses. Even so, given the overall number of participants, the exclusion of some data while conducting inferential analyses still led to powerful tests relevant to the focal issues addressed in this research.

Of the participating students, reported gender included 137 males, 146 females, and 1 student identified as neither male or female (the remainder indicated they did not wish to respond). One hundred (100) students identified as Black, 20 as Asian, 3 as Middle Eastern, and 183 as White. The reported mean age of students was 10.2 years ($SD = 1.0$). Finally, the number of students completing the pre-activity test per grade (and number of classes) are as follows – third: 90 (5); fourth: 155 (9); and fifth: 61 (3). As discussed above, the main analyses involved collapsing across grades, although we do report exploratory analyses involving grade level.

Materials

An example question from the pre-activity test is presented in [Figure 1](#) (adapted from Chen & Klahr, 1999). Each test included 12 CVS questions. Six involved trained content (popcorn popping), and six involved near transfer content (plant growing). We used only near transfer questions because as noted above far transfer is quite difficult and unlikely to occur from a brief intervention (Barnett & Ceci, 2002). For each kind of content, two represented good experiments (vary only one factor relevant to the question prompt), three represented bad experiments in which two factors were

Do plants grow faster when watered 3 times a week or 1 time a week? To answer this question, I run an experiment by comparing the following two groups:



Is this a GOOD experiment or a BAD experiment?

Circle One: Good experiment / Bad experiment

How confident are you in the answer? Circle a number below:

1	2	3	4
Not so sure	Kind of sure	Pretty sure	Totally sure

Figure 1. An example question from the pre-activity test involving the transfer content and a good experiment.

manipulated, and one represented a bad experiment in which an incorrect factor was manipulated. No questions across the tests were repeated.

To conduct the CVS intervention, a magnetic white board was used to display question prompts for the four practice questions (discussed in Procedure below) as well as a pictorial representation of the values for each factor (e.g. red and yellow for the two kinds of popcorn) as illustrated in Figure 2. For the popcorn demonstration, two hot plates were carted to each participating classroom, along with two pots, oil, popcorn kernels, salt, and plastic cups for serving popcorn. Pre-popped popcorn was also distributed to children during the activity and Goldfish crackers were provided for children with food allergies or who did not have caregiver permission for popcorn.

Procedure

The first and final author attended each class to administer the pre-activity test. For a given class, each student first responded to four questions that were meant to uniquely identify the student and would allow us to match their pre-activity questions with their post-activity questions (e.g. What is the second letter of your first name?) and answer three questions about gender, birthdate, and race/ethnicity. Eleven (11) students did not consistently answer the code questions, so their pre- and post-activity data could not be matched for (within-participant) inferential analyses. Next, one of the investigators read the task

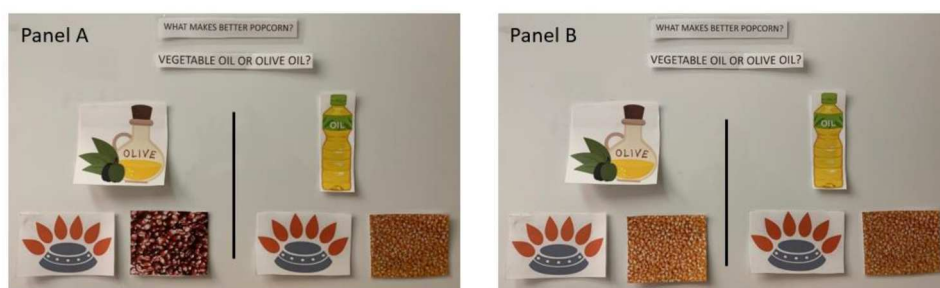


Figure 2. Examples of two experiments aimed at evaluating what makes better popcorn, olive oil or vegetable oil. For each experiment, the components of the two recipes were placed on the board one at a time, and students were engaged in discussion about whether comparing the recipes would illustrate either a bad experiment (Panel A, with two factors being varied) or a good experiment (Panel B, only the correct factor-oil-is varied). Components for these experiments included oil (olive or vegetable), level of heat source (high or low- the latter is not shown above), and type of popcorn (yellow or red, the latter of which is shown in the left recipe in Panel A).

instructions, and then each student completed the pencil-and-paper task by answering each of the twelve questions on their own. In particular, for each question, students were to first respond either ‘yes’ the experiment described is a good one or ‘no’ it is a bad one and then rate their confidence in the answer on the following scale: 1 (not sure), 2 (kind of sure), 3 (pretty sure), and 4 (totally sure). Confidence ratings were included to measure metacognitive changes to determine whether increases in performance were associated with increased awareness of CVS knowledge. Test administration required approximately 10–15 min to complete. In twenty-five cases that were distributed relatively evenly across classes, a student had difficulties reading, and either the teacher or a teachers’ assistant helped them read the questions. All completed tests were included in the analyses. The post-test is provided in the Supplemental Materials.

Approximately 2 weeks after administering the pre-activity test, a team of four to five researchers conducted the popcorn demonstrations. The 20-minute demonstration involved engaging students in the science of popcorn, demonstrating how to make popcorn on a stove top, and then eating popcorn. To begin the demonstration, the activity leader (Author 2) warmed up the students by asking them whether they enjoyed eating popcorn, whether they made it at home, and also asked them about their favourite toppings. He then discussed why popcorn pops and different types of popcorn and then directed students to attend to the second leader (Author 7) who demonstrated how to make popcorn on a stove top. During this part of the demonstration, Author 7 engaged students in a discussion of why oil is used in the process, and then Author 2 invited students to use an infrared camera to explore heat sources (for the pots, hot plates, friends, and anything else in the classroom). The students then were served popcorn and could pose any questions they like to the team leader.

For classes assigned to receive the CVS intervention, a magnetic white board and stand were also brought into the classroom (see Figure 2) for displaying experimental question prompts as well as for setting up the two comparison recipes. For each prompt, the team leader (Author 1) would build each recipe one component of the time. He would choose the components (e.g. olive oil for one recipe and corn oil for the others) and ask students

if his choice was a good one (i.e. if the question involved addressing oil type, this choice would be good, but if the question prompt were about pots, then this choice would involve manipulating another variable). He began with the factor that was presented in the question prompt (e.g. for ‘Does popcorn pop better in olive oil or corn oil?’) and began by displaying the oils and then moved on to the other factors. Occasionally, he would also put up one value of a factor (e.g. olive oil) and ask the class to yell out which value (i.e. either olive or corn oil) should be included in the other recipe. For each component, when students yelled out incorrect answers, he would emphasise the correct one and why it is correct (e.g. in case of two factors being manipulated, he’d remind them that only one should be varied at a time). The intervention script is provided in the Supplemental Materials.

To go through a single question took about 1–2 min, with the entire CVS intervention lasting between about 5–10 min (because each one required less time, given the students began to get more fluent with the task). Across the four questions, one involved a good experiment, two involved bad experiments that manipulated the correct factor along with a second factor, and one involved a bad experiment in which only one factor – but the incorrect one – was manipulated. In this way, all four questions provided demonstrations and the last three provided cognitive conflict. For the latter, the leader would point out that when conducting an experiment, the students need to make sure they are answering the main question. Several times through the CVS intervention, the leader would yell out, ‘Now how many factors should I vary during an experiment’ with students being expected to yell out ‘only one’. If they did not, he would provide a brief reminder about the importance of varying only a single factor. Immediately after the popcorn demonstration, the post-activity test was administered.

Results

Our analyses involve estimating the impact of the CVS intervention on students’ performance on questions about making popcorn (the content used in the intervention) and on questions about growing plants (the transfer content). Our estimates of descriptive values are based on all responses from the pre- and post-intervention tests, whereas the inferential analyses are based on responses from students who completed both the pre- and post-interventions scores and that could be matched. For the latter, some students did not provide the same pre- and post-activity codes and hence we could not match their scores (as noted in the method), so the degrees of freedom for the inferential tests are smaller than would be expected from the overall sample sizes. Most important, the effects that are apparent from inspecting the main descriptive analyses correspond to the outcomes from inferential analyses. After presenting the planned analyses based on collapsing across grade level, we then report secondary analyses pertaining to overall performance as a function of grade and students’ confidence in their answers.

Performance on the CVS questions

Overall performance

For each participant, we computed the proportion of correctly answered CVS questions, separately for the trained and transfer content. As evident from inspecting mean

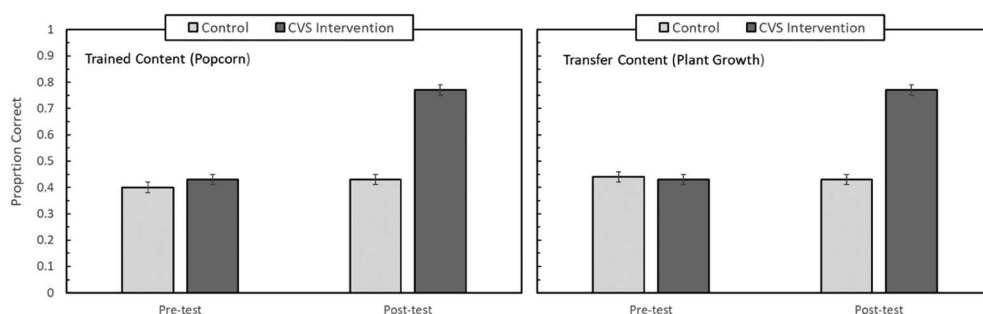


Figure 3. Performance on the CVS questions for those framed with trained content (left panel) and those questions framed with different (transfer) content (right panel). Error bars are standard errors of the mean.

performance in Figure 3, the CVS intervention boosted performance from the pre- to post-activity questions both for the trained content as well as for the transfer content. This observation was supported by outcomes from a 2 (training group: CVS vs. none) $\times$ 2 (test time: Pre- vs. Post activity) $\times$ 2 (kind of question: trained vs. transfer) mixed analysis of variance, with training group being a between-participants variable and the latter two factors being within-participant variables. Namely, the interaction between the training group and test timing was significant, $F(1,261) = 114.3$, $MSE = .06$, $p < .001$, partial $\eta^2 = .30$, indicating that the increase in performance from before to after the popcorn activity occurred when it involved the CVS intervention. Of less interest, the main effects for training group, $F(1,261) = 59.5$, $MSE = .12$, $p < .001$, partial $\eta^2 = .19$, for test timing, $F(1,261) = 151.0$, $MSE = .06$, $p < .01$, partial $\eta^2 = .37$, and for kind of question, $F(1,261) = 4.0$, $MSE = .02$, $p = .047$, partial $\eta^2 = .02$, were significant, and all other interactions were not significant, $ps > .05$.

We provide more detail on two research questions. One, did the CVS intervention improve students' performance on both trained material (i.e. popcorn contexts) and on near transfer material (i.e. plant contexts)? As noted in the mixed ANOVA above, performance in both was improved as a result of the CVS training. Separate t-tests compared the number of correct responses on trained vs. near transfer test items. The results demonstrated that participants in the experimental group scored higher on both trained ($M = .77$, $SD = .05$) and transfer test items ($M = .75$, $SD = .06$) compared to participants in the control group (trained $M = .43$, $SD = .03$; near transfer ($M = .47$, $SD = .05$), ($t(261) = 3.92$, $p = .0001$ and $t(261) = 3.88$, $p = .0001$, respectively).

Two, did the CVS intervention improve students' understanding that the correct factor being varied in an experiment should be informed by the experimental question? Recall that one potential limitation to CVS instruction is that students might interpret any situations in which only one factor is varied constitutes a good test. It is possible that the variable that is isolated is not the target variable. An example of a bad test would be one in which a student wants to test whether the colour of popcorn influences how well it pops, and this student varies only the type of oil. We wanted to measure how well students learned that they must not only vary one thing at a time, but that they must vary the correct target variable. To answer this question, we conducted an analysis on the subset of questions in which an incorrect factor (but only one) was varied, such as

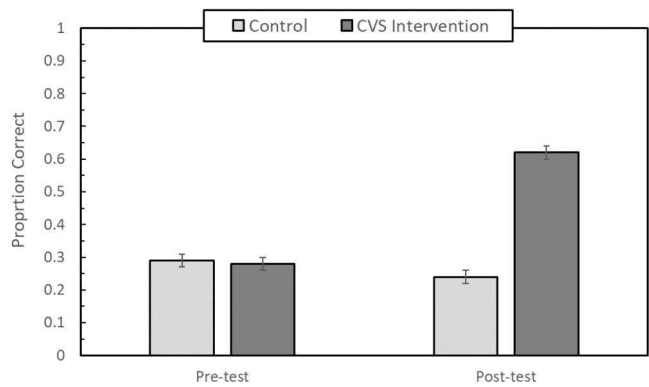


Figure 4. Proportion correct on questions that involve experiments that only vary a single factor but where the factor being varied is not relevant to the experimental question (see text for details). Error bars represent standard errors of the mean.

varying the amount of water to evaluate plant growth when the question pertained to whether the amount of light influenced growth. Given that only two questions of this type were included on each test (one with trained content and one with transfer content), we collapsed across question type to conduct this analysis. As shown in Figure 4, the CVS intervention did improve students’ performance on these questions, which was confirmed by a significant interaction between training group and time of test, $F(1,261) = 47.6$, $MSE = .11$, $p < .001$, partial $\eta^2 = .15$. A separate t-test demonstrates that students in the CVS training group identified comparisons with the correct factor being manipulated more frequently than those in the control group (CVS group $M = .62$, $SD = .05$; control group ($M = .28$, $SD = .04$), ($t(261) = 3.11$, $p = .001$).

The current investigation was not designed to evaluate the degree to which grade-level moderates the impact of the intervention, but during recruitment of classes for participation, we did have a range of grades (from third to fifth). And, class assignments to group per grade were as follows: For the control groups, 2, 4 and 1 classes participated in the 3rd, 4th and 5th grades, respectively; and, for the CVS intervention groups, 3, 3, 5, and 2 classes participated in the 3rd, 4th and 5th grades, respectively. Performance as a function of grade level is presented in Table 1, and outcomes from a 2 (training group: CVS vs. none) x 2 (test time: Pre- vs. Post activity) x 3 (grade: 3rd, 4th, 5th) mixed ANOVA are consistent with the apparent trends. Namely, performance overall increased with grade level – main effect of grade, $F(2,257) = 15.2$, $MSE = .06$, $p < .001$, partial $\eta^2 = .11$; most important, the interaction between test

Table 1. Performance on CVS Questions on the pre-activity test (pre-test) and the post-activity test (post-test).

	No CVS		CVS Intervention	
	Pre-test	Post-test	Pre-test	Post-test
Grade level				
Third	.33 (.02)	.38 (.02)	.39 (.02)	.68 (.04)
Fourth	.43 (.02)	.45 (.02)	.41 (.02)	.76 (.03)
Fifth	.54 (.05)	.45 (.05)	.53 (.04)	.89 (.03)

Note: CVS = control-of-variables strategy. Values in parentheses are standard errors of each mean.

Table 2. Confidence ratings on CVS questions as a function of pre-activity test (pre vs. post) and kind of questions (trained vs. transfer)

	No CVS		CVS Intervention	
	Pre-test	Post-test	Pre-test	Post-test
Trained content	2.8 (.06)	3.1 (.06)	2.9 (.05)	3.5 (.04)
Transfer content	2.9 (.07)	3.1 (.06)	3.0 (.05)	3.4 (.05)

Note: CVS = control-of-variables strategy. Trained content refers to questions framed as a popcorn experiment, and transfer content refers to questions framed as a plant-growing experiment. Higher values indicate greater confidence in answers. Values in parentheses are standard errors of each mean.

time (pre vs. post) and training was significant, $F(1,257) = 104.3$, $MSE = .03$, $p < .001$, partial $\eta^2 = .29$, and no other main effects or interactions qualified this critical interaction. Namely, regardless of grade, students benefited from the CVS intervention.

Students' confidence for answers to CVS questions

After each question on the pre- and post-activity test, students rated their confidence in their answer on a scale from 1 (not so sure) to 4 (totally sure). Means across student ratings are presented in Table 2 as a function of test timing (pre vs. post) and kind of question (trained or transfer), as per our analyses of overall test performance (Figure 2). Several outcomes are notable. First, prior to the intervention, students tended to show relatively high levels of confidence, with the values indicating that on average students were 'pretty sure' (value of 3), despite their overall chance level of performance on those questions (Figure 2). Second, students' confidence tended to increase from pre-activity to post-activity test, with the main effect for test timing being significant, $F(1,259) = 122.6$, $MSE = 90.8$, $p < .001$, partial $\eta^2 = .32$. And, finally, students' confidence increased most after they had received the CVS intervention, indicated by a significant interaction between group and test timing, $F(1,259) = 10.7$, $MSE = 90.8$, $p < .001$, partial $\eta^2 = .04$. Thus, not only did students' performance increase after receiving the CVS intervention, their overall confidence in their answers also increased.

Discussion

The control-of-variables strategy is arguably foundational for thinking like a scientist – it provides a crucial element for making strong conclusions about the effect of a given variable on an outcome and for understanding why causal conclusions cannot be confidently drawn from correlational data. As demonstrated in prior research (for a review, see Schwichow et al., 2016; Schwichow, Brandenburger, & Wilbers, 2022), even younger students can learn CVS when given one-on-one instruction that includes cognitive conflict and resolution. Our approach was inspired by these prior successes at training CVS, with our main aim being to evaluate whether 3rd-5th grade students could learn CVS using a brief intervention that was targeted at a group, rather than individuals (i.e. not one-on-one training for each student). We created a brief intervention that included demonstrations and cognitive conflict, two elements associated with successful interventions in previous research (Schwichow et al., 2016). In the current case, we embedded the brief (fewer than 10 min) intervention into an engaging demonstration on the science of popcorn.

The current CVS intervention did have the expected impact. Regardless of the grade level, students who received the intervention improved their performance on identifying experiments that either did conform to the CVS rule (i.e. a good experiment) or did not (i.e. a bad one). Concerning questions that involved bad experiments, after training students were more likely to correctly identify the experiments as bad if they involved varying (a) two factors or (b) a single, but incorrect factor (Figures 3 and 4). Moreover, the students' confidence in their answers was boosted by the CVS intervention (Table 2), indicating that they realised that they had a better grasp on what makes an experiment good. By contrast, students who did not receive the intervention enjoyed no improvements in performance on CVS questions yet continued to be relatively confident they were answering correctly (Table 2). This change in confidence is consistent with the predictions of the science framework theory in that the acquisition of new knowledge requires a modification of existing knowledge and its relation to existing knowledge, in part, through emerging metacognitive awareness (Vosniadou, 2019; Vosniadou & Skopeliti, 2014).

The benefits of the CVS intervention occurred not only on post-activity questions that involved the trained content (making popcorn, left-panel, Figure 3) but also occurred for the untrained, transfer content (growing plants, right panel). In the present case, the transfer would be considered relatively near, because only a single dimension – the cover story of the experiment – was changed from the questions that included trained content to those that involved transfer content (for an in-depth discussion of transfer, see Barnett & Ceci, 2002). Demonstrating near transfer is an important step toward demonstrating the potential efficacy of an intervention, and future research should be aimed at estimating the degree to which relatively far transfer could be obtained after a brief intervention focusing on a single content domain. Given that transfer will likely be improved through structural alignment between training and evaluation tasks (e.g. Kurtz et al., 2013), will students perform as well when the training materials (Figure 2) do not have the same structural alignment as the test questions? We suspect that a single CVS intervention focused on a single content domain will be limited in the degree to which it does promote relatively far transfer. However, as we discuss below, the brevity of the CVS intervention will afford its implementation across domains and experimental designs, and doing so would be expected to promote far transfer.

What is apparent from inspection of Figure 3, however, is that although the CVS intervention did boost students' performance on the CVS post-activity test, many students did not master the strategy. The fact that students did not fully master the CVS rule is apparent from their performance on the questions that involved manipulating a single, but incorrect factor: The average performance was only about 60% (Figure 3), consistent with previous interventions (e.g. Klahr & Nigam, 2004; Schwichow et al., 2022). Such a low-level of performance could arise from multiple factors: Some students may not have fully comprehended CVS after training, or they may not carefully read the experimental questions (with the bias to call any experiment with a single factor being varied as 'good'). Further instruction and practice solving the problems (with encouragement to read the experimental questions carefully) may be needed to further boost students' understanding and performance. The need for extra training is not surprising, because just as in prior research that involved CVS training, a single intervention (especially a briefer one) would not be expected to lead to mastery, especially over the

long retention intervals in which forgetting would occur (for further rationale, see Rawson & Dunlosky, 2022). This evidence supports the idea that despite young children's impressive naive scientific reasoning skills, children need learning opportunities to acquire more formal scientific reasoning skills.

Given these observations, one promising aspect of the present approach to training CVS is that it can be administered repeatedly across multiple sessions, because many such demonstrations could easily be supplemented with a complementary CVS intervention. For instance, if instructors are discussing the role of light in plant growth, they can have a brief interlude where the entire class is asked to evaluate the quality of two or three experiments that are set up to investigate what makes plants grow better. This would be the case for almost any science demonstration, but as highlighted in the present research, instructors can develop experimental questions for just about any class subject – even how to best cook popcorn.

The tactic of embedding CVS interventions across multiple demonstrations (that themselves are distributed across an entire school year) takes advantage of *spaced practice*, which is practicing the *same* activity (in this case, using CVS knowledge to evaluate experiment quality) on two or more occasions across time. Spaced practice is one of the most effective ways to obtain long-term retention (for reviews, see Dunlosky et al., 2013; Wiseheart et al., 2019), so repeating the brief intervention promises to boost students' long-term retention of the CVS. As important, by embedding the CVS intervention into *different contexts and experimental designs* (e.g. involving more than 3 factors), students would be expected to demonstrate even greater transfer using the CVS across new topics, because variability in training often leads to a larger breadth of transfer (for discussion of the role of variability in transfer, see Gentner & Hoyos, 2017; Schmidt & Bjork, 1992). Of course, these are *possible* benefits arising from a brief, scalable approach for training CVS, and importantly, they can be empirically evaluated in future classroom research.

Although the current research was not designed to estimate grade-level relations, inspection of Table 1 does reflect standard grade-related increases in performance (e.g. see also, Chen & Klahr, 1999, Figure 4). Perhaps most notable, although the third graders did improve their performance after receiving the CVS intervention, they still performed relatively poorly on the post-test (about 68%) and did not obtain the same level of performance as their older peers. Such grade-related differences may arise from multiple sources, such as the younger students (compared to their older counterparts) have less a priori knowledge, demonstrate less engagement during the demonstration, and so forth. One testable hypothesis is that repetition of the CVS intervention across time and contexts – the spaced practice approach discussed above – would by necessity reduce the grade-level difference as all students gained mastery through repeated, spaced practice.

Limitations

Although the results are promising, several potential limitations should be addressed in future research. One limitation is the small samples that limited comparisons by grade. A larger, balanced sample would allow proper comparisons that would help to inform the degree to which the intervention was effective for each grade. A second limitation is the presence of a highly skilled subject matter expert who led the activities. The SME in the present study has extensive experience presenting

highly engaging programming to young children and this factor may have potentially increased the potency of the intervention beyond what might occur when implemented by a less experienced SME. A third, potentially related limitation, is that presenting a novel activity in a traditional school setting might be inherently engaging. In both cases, increasing student arousal, attention, and engagement might have increased the potency of the intervention.

Implications for practice

As noted in the introduction, children begin school curious, interested, and engaged in science (Jirout, 2020). By middle school, this curiosity and engagement begin to decline, with an ever sharper decline for students from underrepresented groups (Riegle-Crumb et al., 2011). Engaging in informal science learning can attenuate this decline (Bonnette et al., 2019) by helping children see that they can ‘do’ science (Lei et al., 2019). Such experiences are particularly effective when they show children how science is relevant to their everyday lives (Rogoff et al., 2018). Cooking activities provide an excellent opportunity to link science with an everyday, lived experience that connects to science learning (Morris et al., 2023), math learning (Leyva et al., 2022), family, culture, and other important life skills (Morris et al., 2021).

This informal science activity was performed in intact classrooms in public schools. This is a more structured environment than a setting such as a science museum in which implementing this activity may be more challenging. This activity might be best implemented as a stage show in which participants (e.g. children and families) are made aware of the activity in advance and plan to stay in the space for the duration of the activity. The assessment of learning used in this experiment is unlikely to be implemented in such spaces and brief, embedded assessments (McManimon, 2021) might be a good option for implementation in public learning spaces such as museums.

In conclusion, a brief CVS intervention presented to intact 3rd through 5th grade classes boosted students’ ability to identify good and bad experiments. Given its scalability across different content, this evidence-based intervention holds promise for providing students with spaced practice of this foundational concept in experimental methods. Our intervention can be used in any setting in which groups are engaged in science activities, is low-cost, and requires little training to implement. This food-related intervention is an example of a context for engaging children in activities that makes science relevant to their lives, notably through their everyday, lived experience. These kinds of informal science activities are fun, engaging learning experiences that help children acquire the formal tools of scientific reasoning.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This material is based upon work supported by the National Science Foundation under Grant Number 1906706. Any opinions, findings, and conclusions or recommendations expressed in

this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Ethics statement

The study reported in this manuscript was reviewed and approved by the XXX Institutional Review Board (Protocol #2022-140) as an **Exempt** request for Approval to Use Human Research Participants. It was classified as Exempt because the research involved only normal educational practices and did not include any potentially identifiable information from participants.

ORCID

Bradley J. Morris  <http://orcid.org/0000-0002-1094-2558>

Katie Asaro  <http://orcid.org/0009-0001-1373-3479>

Yin Zhang  <http://orcid.org/0009-0008-9274-6761>

Michelle Rivers  <http://orcid.org/0000-0002-4931-2895>

John Dunlosky  <http://orcid.org/0000-0002-7367-7958>

References

- Asghar, A. (2012). Informal science contexts: Implications for formal science learning. *LEARNing Landscapes*, 5(2), 55–72. <https://doi.org/10.36510/learnland.v5i2.551>
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612–637. <https://doi.org/10.1037/0033-2909.128.4.612>
- Behrendt, M., & Franklin, T. (2014). A review of research on school field trips and their value in education. *International Journal of Environmental and Science Education*, 9, 235–245.
- Bonnette, R. N., Crowley, K., & Schunn, C. D. (2019). Falling in love and staying in love with science: ongoing informal science experiences support fascination for all children. *International Journal of Science Education*, 41(12), 1626–1643. <https://doi.org/10.1080/09500693.2019.1623431>
- Bramley, N. R., Jones, A., Gureckis, T. M., & Ruggeri, A. (2022). Children’s failure to control variables may reflect adaptive decision-making. *Psychonomic Bulletin & Review*, 29(6), 2314–2324. <https://doi.org/10.3758/s13423-022-02120-1>
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70(5), 1098–1120.
- Dejonckheere, P. J., Van de Keere, K., & Tallir, I. (2011). Are fourth and fifth grade children better scientists through metacognitive learning? *Electronic Journal of Research in Educational Psychology*, 9, 133–156. <https://doi.org/10.25115/ejrep.v9i23.1431>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students’ learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14, 4–58. <https://doi.org/10.1177/1529100612453266>
- Fernbach, P. M., & Erb, C. D. (2013). A quantitative causal model theory of conditional reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(5), 1327.
- Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science*, 9, 672–693. <https://doi.org/10.1111/tops.12278>
- Goff, E. E., Mulvey, K. L., Irvin, M. J., & Hartstone-Rose, A. (2020). The effects of prior informal science and math experiences on undergraduate STEM identity. *Research in Science & Technological Education*, 38(3), 272–288. <https://doi.org/10.1080/02635143.2019.1627307>
- Howell, E. L., & Brossard, D. (2021). (Mis) informed about what? What it means to be a science-literate citizen in a digital world. *Proceedings of the National Academy of Sciences*, 118(15), e1912436117. <https://doi.org/10.1073/pnas.1912436117>

- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures* (Vol. 22). Psychology Press.
- Jee, B. D., Matlen, B. J., Greenlaw, M., Simms, N., & Gentner, D. (2022). Spatial supports for comparison in educational science images. *Instructional Science*, 50(6), 807–827. <https://doi.org/10.1007/s11251-022-09599-0>
- Jirout, J. J. (2020). Supporting early scientific thinking through curiosity. *Frontiers in Psychology*, 11, 1717. <https://doi.org/10.3389/fpsyg.2020.01717>
- Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. MIT Press.
- Klahr, D., & Chen, Z. (2011). Finding one's place in transfer space. *Child Development Perspectives*, 5(3), 196–204. <https://doi.org/10.1111/j.1750-8606.2011.00171.x>
- Klahr, D., & Nigam, M. (2004). The Equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, 15(10), 661–667. <https://doi.org/10.1111/j.0956-7976.2004.00737.x>
- Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational interventions to advance children's scientific thinking. *Science*, 333(6045), 971–975. <https://doi.org/10.1126/science.1204528>
- Koerber, S., & Osterhaus, C. (2019). Individual differences in early scientific thinking: Assessment, cognitive influences, and their relevance for science learning. *Journal of Cognition and Development*, 20(4), 510–533. <https://doi.org/10.1080/15248372.2019.1620232>
- Köksal-Tuncer, Ö., & Sodian, B. (2018). The development of scientific reasoning: Hypothesis testing and argumentation from evidence in young children. *Cognitive Development*, 48, 135–145.
- Kurtz, K. J., Boukrina, O., & Gentner, D. (2013). Comparison promotes learning and transfer of relational categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 1303–1310. <https://doi.org/10.1037/a0031847>
- Lazonder, A. W., & Egberink, A. (2014). Children's acquisition and use of the control-of-variables strategy: effects of explicit and implicit instructional guidance. *Instructional Science*, 42, 291–304. <https://doi.org/10.1007/s11251-013-9284-3>
- Legare, C. H. (2014). The contributions of explanation and exploration to children's scientific reasoning. *Child Development Perspectives*, 8(2), 101–106.
- Lei, R. F., Green, E. R., Leslie, S. J., & Rhodes, M. (2019). Children lose confidence in their potential to “be scientists,” but not in their capacity to “do science”. *Developmental Science*, 22(6), e12837. <https://doi.org/10.1111/desc.12837>
- Leyva, D., Yeomans-Maldonado, G., Weiland, C., & Shapiro, A. (2022). Latino kindergarteners' math growth, approaches to learning, and home numeracy practices. *Journal of Applied Developmental Psychology*, 80, 101417. <https://doi.org/10.1016/j.appdev.2022.101417>
- Lorch, R. F., Lorch, E. P., Calderheard, W. J., Dunlap, E. E., Hodell, E. C., & Freer, B. D. (2010). Learning the control of variables strategy in higher and lower achieving classrooms: Contributions of explicit instruction and experimentation. *Journal of Educational Psychology*, 102, 90–101. <https://doi.org/10.1037/a0017972>
- Lorch, Jr., R. F., Lorch, E. P., Freer, B. D., Dunlap, E. E., Hodell, E. C., & Calderhead, W. J. (2014). Using valid and invalid experimental designs to teach the control of variables strategy in higher and lower achieving classrooms. *Journal of Educational Psychology*, 106(1), 18. <https://doi.org/10.1037/a0034375>
- Martella, A. M., Klahr, D., & Li, W. (2020). The relative effectiveness of different active learning implementations in teaching elementary school students how to design simple experiments. *Journal of Educational Psychology*, 112(8), 1582.
- Martin, A. J., Durksen, T. L., Williamson, D., Kiss, J., & Ginns, P. (2016). The role of a museum-based science education program in promoting content knowledge and science motivation. *Journal of Research in Science Teaching*, 53(9), 1364–1384. <https://doi.org/10.1002/tea.21332>
- Matlen, B. J., & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science*, 41(3), 621–634. <https://doi.org/10.1007/s11251-012-9248-z>

- McManimon, S. K. (2021). Embedded research practices: Practice as process, participatory method, and product in informal learning research. *Journal of Museum Education*, 46(2), 245–254. <https://doi.org/10.1080/10598650.2021.1891757>
- Morris, B. J., Croker, S., Masnick, A. M., & Zimmerman, C. (2012). The emergence of scientific reasoning. In H. Kloos, B. J. Morris, & J. L. Amaral (Eds.), *Current topics in children's learning and cognition* (pp. 61–82). InTech.
- Morris, B. J., Zentall, S. R., *Murray, G., & Owens, W. (2021). Enhancing informal STEM learning through family engagement in cooking. *Proceedings of the Singapore National Academy of Science*, 15(2), 119–133.
- Morris, B. J., Zhang, Y., Asaro, K., Cason, J., Pollock, B., St. Clair, K., & Owens, W. (2023). Cooking up STEM: Adding wh-questions to a recipe increases family STEM talk. *Journal of Applied Developmental Psychology*, 88, 01581.
- National Academies of Sciences Engineering and Medicine. (2016). In C. E. Snow & K. A. Dibner (Eds.), *Science literacy*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/23595>
- National Academies of Sciences, Engineering, and Medicine. (2018). *Graduate STEM education for the 21st century*. The National Academies Press. <https://doi.org/10.17226/25038>
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press. <https://doi.org/10.17226/13165>
- NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.
- Padilla, M. J., Okey, J. R., & Garrard, K. (1984). The effects of instruction on integrated science process skill achievement. *Journal of Research in Science Teaching*, 21(3), 277–287. <https://doi.org/10.1002/tea.3660210305>
- Rawson, K. A., & Dunlosky, J. (2022). Successive relearning: An underexplored but potent technique for obtaining and maintaining knowledge. *Current Directions in Psychological Science*, 31, 362–368. <https://doi.org/10.1177/09637214221100484>
- Riegle-Crumb, C., Moore, C., & Ramos-Wada, A. (2011). Who wants to have a career in science or math? Exploring adolescents' future aspirations by gender and race/ethnicity. *Science Education*, 95(3), 458–476. <https://doi.org/10.1002/sce.20431>
- Rogoff, B., Dahl, A., & Callanan, M. (2018). The importance of understanding children's lived experience. *Developmental Review*, 50, 5–15.
- Salmi, H., & Thuneberg, H. (2019). The role of self-determination in informal and formal science learning contexts. *Learning Environments Research*, 22, 43–63. <https://doi.org/10.1007/s10984-018-9266-0>
- Schalk, L., Edelsbrunner, P. A., Deiglmayr, A., Schumacher, R., & Stern, E. (2019). Improved application of the control-of-variables strategy as a collateral benefit of inquiry-based physics education in elementary school. *Learning and Instruction*, 59, 34–45. <https://doi.org/10.1016/j.learninstruc.2018.09.006>
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3, 207–218. <https://doi.org/10.1111/j.1467-9280.1992.tb00029.x>
- Schwichow, M., Brandenburger, M., & Wilbers, J. (2022). Analysis of experimental design errors in elementary school: how do students identify, interpret, and justify controlled and confounded experiments?. *International Journal of Science Education*, 44(1), 91–114.
- Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., & Härtig, H. (2016). Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review*, 39, 37–63. <https://doi.org/10.1016/j.dr.2015.12.001>
- Schwichow, M., Osterhaus, C., & Edelsbrunner, P. A. (2020). The relation between the control-of-variables strategy and content knowledge in physics in secondary school. *Contemporary Educational Psychology*, 63, 101923.

- Sharon, A. J., & Baram-Tsabari, A. (2020). Can science literacy help individuals identify \misinformation in everyday life? *Science Education*, 104(5), 873–894. <https://doi.org/10.1002/sce.21581>
- Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology*, 42(6).
- Staus, N. L., Falk, J. H., Price, A., Tai, R. H., & Dierking, L. D. (2021). Measuring the long-term effects of informal science education experiences: challenges and potential solutions. *Disciplinary and Interdisciplinary Science Education Research*, 3(1), 1–15. <https://doi.org/10.1186/s43031-021-00031-0>
- Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path independence. *Cognitive Development*, 23(4), 488–511. <https://doi.org/10.1016/j.cogdev.2008.09.005>
- Teig, N., & Scherer, R. (2016). Bringing formal and informal reasoning together—A new era of assessment? *Frontiers in Psychology*, 7, 1097. <https://doi.org/10.3389/fpsyg.2016.01097>
- van der Graaf, J., Segers, E., & Verhoeven, L. (2015). Scientific reasoning abilities in kindergarten: Dynamic assessment of the control of variables strategy. *Instructional Science*, 43, 381–400. <https://doi.org/10.1007/s11251-015-9344-y>
- Vincent-Ruz, P., & Schunn, C. D. (2021). Identity complexes and science identity in early secondary: Mono-topical or in combination with other topical identities. *Research in Science Education*, 51, 369–390. <https://doi.org/10.1007/s11165-019-09882-0>
- Vosniadou, S. (2019). The development of students' understanding of science. *Frontiers in Education*, 4, 32. <https://doi.org/10.3389/feduc.2019.00032>
- Vosniadou, S., & Skopeliti, I. (2014). Conceptual change from the framework theory side of the fence. *Science & Education*, 23, 1427–1445. <https://doi.org/10.1007/s11191-013-9640-3>
- Wagensveld, B., Segers, E., Kleemans, T., & Verhoeven, L. (2015). Child predictors of learning to control variables via instruction or self-discovery. *Instructional Science*, 43(3), 365–379. <https://doi.org/10.1007/s11251-014-9334-5>
- Weisberg, D. S., & Sobel, D. M. (2022). *Constructing science: Connecting causal reasoning to scientific thinking in young children*. MIT Press.
- Wiseheart, M., Küpper-Tetzel, C. E., Weston, T., Kim, A. S. N., Kapler, I. V., & Foot-Seymour, V. (2019). Enhancing the quality of student learning using distributed practice. In J. Dunlosky, & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 550–583). Cambridge University Press.
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27(2), 172–223.
- Zohar, A., & Peled, B. (2008). The effects of explicit teaching of metastrategic knowledge on low- and high-achieving students. *Learning and Instruction*, 18(4), 337–353. <https://doi.org/10.1016/j.learninstruc.2007.07.001>