



Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



Right or wrong? How feedback content and source influence children's mathematics performance and persistence



Megan Merrick, Emily R. Fyfe^{*}

Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA

ARTICLE INFO

Article history:

Received 5 May 2023

Revised 6 January 2024

Accepted 8 January 2024

Available online 5 February 2024

Keywords:

Feedback

Mathematical equivalence

Motivation

Persistence

Strategy use

ABSTRACT

The current study examined how different features of corrective feedback influenced children's performance and motivational outcomes on a mathematics task. Elementary school-aged children from the United States ($N = 130$; $M_{\text{age}} = 7.61$ years; 35% female; 60% White) participated in a Zoom session with a trained researcher. During the learning activity, children solved a series of mathematical equivalence problems and were assigned to different feedback conditions that varied in feedback content (correct answer alone vs. correct answer with verification) and feedback source (computer alone vs. computer with person). In terms of content, feedback with verification cues led to decreased persistence, decreased strategy variability, and higher reliance on entrenched strategies relative to feedback that contained the correct answer alone. In terms of source, feedback from the computer alone enhanced children's accuracy; however, the most resilient children received feedback from the computer and a person. Findings are discussed in light of existing feedback theories.

© 2024 Elsevier Inc. All rights reserved.

Introduction

Students constantly receive feedback in learning settings—feedback on their homework, their writing, and their behavior, to name a few examples. And this feedback is often provided with good

^{*} Corresponding author.

E-mail address: efyfe@indiana.edu (E.R. Fyfe).

intentions—to supply information about a person's performance that can be used for reaching a goal (e.g., learning a correct strategy, remembering target content). Furthermore, research in psychology and education indicates that this type of corrective feedback can be a powerful learning tool; meta-analyses consistently demonstrate positive effects of feedback on learning and performance relative to no-feedback control conditions (e.g., Van der Kleij et al., 2015; Wisniewski et al., 2020). However, feedback can have unintended consequences; there is considerable variability in the effects of feedback, and it may even hinder learning in some cases (Hattie & Gan, 2011, Kluger & DeNisi, 1996). This begs the question: Which features make feedback effective in triggering positive cognitive change? The goal of the current research was to experimentally evaluate how two features—the *content of feedback* and the *source of feedback*—influence children's in-the-moment performance and persistence during a target math task.

Theoretical foundations

Multiple theoretical frameworks suggest that the *evaluative* nature of feedback may have unintended consequences that prevent optimal performance. For example, *feedback intervention theory* posits that feedback has the potential to direct the learner's attention to task-level processes (e.g., examining problem structure) or to meta-task processes, which include attention to the self (Kluger & DeNisi, 1996; see also Hattie & Timperley, 2007). According to this theory, feedback is more likely to hinder performance when it directs attention toward the self because fewer cognitive resources are available to solve the task at hand (Kluger & DeNisi, 1998). For example, if a teacher tells students that they solved a problem incorrectly, the students' cognitive resources may be directed toward worrying about their abilities or the teacher's impression of them rather than toward learning from their error. In fact, empirical work suggests that adults learn less from this type of negative feedback (i.e., you are incorrect) relative to positive feedback (i.e., you are correct), partly because negative feedback produces threats to learners' self-esteem that lead the learners to tune out (Eskreis-Winkler & Fishbach, 2019).

Similarly, in the *feedback as an individual resource* framework, Ashford and Cummings (1983) discussed both the costs and benefits of seeking out feedback in various contexts and how it relates to reputation management. They argued that “though in many situations it may be very useful to obtain feedback for its error corrective and uncertainty reducing properties, individuals may be reluctant to actively pursue it in an attempt to protect their self-esteem” (p. 377). More recently, Grundmann et al. (2021) proposed the *model of motivated feedback disengagement*, which suggests that these risks associated with feedback can induce negative affect (e.g., disappointment, shame, anger) that sometimes leads people to actively disengage from the feedback. Together, these theoretical frameworks suggest that features that escalate the evaluative nature of feedback may have unintended consequences for performance. Two of these features (among others) are the *content* of the feedback message and the *source* of the feedback message, which we now consider in turn.

The content of feedback

The content of the feedback message can vary in the amount and type of information presented (Shute, 2008). For example, feedback can include a verification cue, which signals whether the response was correct or incorrect (i.e., a right–wrong judgment), and feedback can also include the correct answer or elaborated explanations of the correct response (Dempsey, 1993). Previous research has often contrasted verification-only information with conditions that include additional information (e.g., Pashler et al., 2005; Roper, 1977) because verification-only is often viewed as representing minimal information load for the learner (Kulhavy & Stock, 1989). Based on these types of comparisons, researchers tend to argue that “effective feedback should include elements of both verification and elaboration” (Shute, 2008, p. 158) and that “the efficacy of feedback increases substantially when the correct answer is added to the verification component” (Butler & Woodward, 2018, p. 11). The idea is that verification-only often provides insufficient information to correct any errors, and this should be supplemented with additional relevant information.

However, it remains an open question as to whether the verification component is needed. In the current study, we took a different approach and used correct-answer feedback as the baseline and either did or did not supplement it with verification cues. On the one hand, the verification component may be necessary to remove any ambiguity; learners certainly benefit from clear markers of what is correct information versus incorrect information (e.g., McGinn et al., 2015). On the other hand, including this type of explicit right–wrong judgment may exacerbate the evaluative nature of feedback and have unintended consequences. In line with feedback intervention theory, explicit verification cues may be perceived as threatening or lead to normative comparisons (e.g., “I bet other children my age got this correct”), which in turn may detract from task-relevant processing (Kluger & DeNisi, 1996). Similarly, in line with the feedback as an individual resource framework, verification statements may include inherent trade-offs between obtaining the most detailed and direct information and risking one’s reputation (Ashford & Cummings, 1983). It may be less risky to just seek a model of the desired response (e.g., the correct answer alone).

Thus, in general, the inclusion of verification cues may heighten the evaluative nature of feedback as it relates to cognitive resources, reputation management, and emotion regulation. Especially for instances of failure, initiating the feedback message with a stark indication of one’s error may be more likely to heighten attention to the self or threaten one’s reputation compared with relaying a feedback message that merely contains the correct response. We hypothesize that adding verification feedback to the correct answer will decrease performance and persistence during children’s mathematics problem solving.

The source of feedback

In addition to content, the *source* of the feedback message—in terms of who or where it comes from—may also influence learners’ outcomes. Learning from feedback is often a social exchange, but advances in educational technology provide a range of tools for providing feedback in the absence of a person (e.g., e-learning apps, online worksheets). There are intuitive reasons to expect feedback from a person to be valuable. For example, self-report data suggest that students appreciate the individualized nature of face-to-face verbal feedback because it allows for additional clarification (Mulliner & Tucker, 2017), and other research attests to the benefits of socially contingent input that is afforded by interacting with a person (Roseberry et al., 2014).

However, verbal feedback from a person may also have unintended consequences in certain situations. For example, Kluger and DeNisi (1996) specifically identified “person-mediated” feedback as having a high probability of directing attention to meta-task processes such as worrying about self-image or the teacher’s impressions. In a similar way, the feedback as an individual resource framework suggests that seeking and accepting feedback *from a person* may be particularly risky given the potential “loss of face” in this interpersonal interaction (Ashford & Cummings, 1983, p. 377). Finally, the model of motivated feedback disengagement includes situations in which learners actively avoid feedback, and although it does not specifically distinguish between person-mediated and computer-mediated feedback (e.g., presented verbally or in an e-mail), it is specifically about feedback *from other people*. These theoretical considerations of person-mediated feedback have also influenced recommendations: Shute (2008) advised to “avoid delivering feedback orally” and instead to use written or computer-delivered feedback to maintain a neutral and non-evaluative learning environment (p. 178).

A large number of studies have implemented computer-mediated feedback (see Van der Kleij et al., 2015), but it is exceedingly rare to directly and experimentally contrast computer-based feedback with or without person-based feedback, and the existing research is mixed. A small set of studies finds benefits of feedback that is exclusively from a computer. For example, Kluger and Adler (1993) had undergraduate engineering students complete a series of mathematics problems, and 20 of 21 participants (95%) who had the option to receive computer-based feedback requested it, but only 1 of 14 participants (7%) who had the option to receive person-based feedback requested it. Comer (2007) reported a similar finding among undergraduate students completing a computer simulation task. In particular, participants who received negative feedback (e.g., “You are at 59%; that is below the goal”) from the computer had higher intrinsic motivation and fewer negative emotions compared with participants who received negative feedback from a person. Finally, Đorić and colleagues (2021)

examined middle school students learning physics in a computer simulation task. Feedback regarding students' performance was provided within the simulation (computer-based) or provided traditionally by a teacher (person-based). The condition with computer-based feedback outperformed the person-based feedback condition on a posttest measure.

However, these benefits of computer-based feedback alone do not always translate to increased performance relative to person-based feedback (e.g., Comer, 2007; Kluger & Adler, 1993). One study with middle school students found the reverse; person-based feedback led to better text comprehension than computer-based feedback (Golke et al., 2015). The researchers speculated that person-based feedback was effective "because the students were committed to actively engage in processing the feedback in the presence of the experimenter" (p. 133). Another study with university students showed that perceived differences in feedback source had no effect on improving essay scores (Lipnevich & Smith, 2009a,b). There were no differences in scores depending on whether the feedback was perceived to be from a computer or from the instructor.

These findings suggest that the benefits of computer-based feedback alone—in the absence of a person—may be context specific, and research is needed to compare computer-based and person-based feedback on a range of learning outcomes. Given the theoretical frameworks, we hypothesized that adding verbal feedback from a person to computer-based feedback will increase attention to the self in a way that decreases performance and persistence during children's mathematics problem solving.

The current study

The goal of the current study was to experimentally evaluate how two features of the learning environment—the content of feedback and the source of feedback—influence children's mathematics problem solving. This research contributes in novel ways to the literature by (a) experimentally isolating how verification feedback influences performance, (b) comparing computer-mediated input with and without feedback from a person, (c) focusing on a sample of children learning critical STEM (science, technology, engineering, and math) content, and (d) incorporating multiple in-the-moment outcomes, including children's accuracy on the task and their persistence to keep going.

In the current study, elementary school children practiced solving math problems during an online video-call session with different forms of feedback. For some children the feedback was provided by the computer alone, and for others it was provided by the computer and verbally by a person who was present on the video call. Within each of those groups, some children received only the correct answer in the feedback message and others received verification feedback (i.e., an explicit correct/incorrect judgment) plus the correct answer. The problems the children solved were math equivalence problems, which often have operations on both sides of the equal sign (e.g., $5 + 4 = 2 + \underline{\quad}$). Math equivalence is a foundational standards-based math topic that predicts later achievement (Hornburg et al., 2022; Matthews & Fuchs, 2020; McNeil et al., 2019). Unfortunately, children in the United States often exhibit misconceptions about math equivalence and struggle to solve these problems correctly (see McNeil, 2014), making it a critical domain in which to examine how feedback can influence performance and persistence. We created a fairly non-evaluative context to support children's progress on a difficult topic and to examine differing feedback effects during a low-stakes learning-based activity.

One outcome of interest for this study was children's accuracy on the math problems to see how the different types of feedback influenced their success on the task. We also assessed children's exploratory behaviors—including whether feedback motivated children to explore the space more widely (e.g., choosing to solve more items, trying new strategies) or to refrain from exploring (e.g., choosing to solve fewer items, relying on common and familiar strategies). We opted to study ongoing performance (as opposed to post-training learning or retention) to see how feedback influenced in-the-moment decisions (e.g., try this answer, stop the task) because these have implications for how rich or how long the learning experience is for any given child.

Broadly, we expected feedback that was more evaluative in nature to be less effective because it can draw children's attention away from the task and toward their own self-image and abilities. In contrast, less evaluative feedback may support children in feeling free to explore the learning space without as much concern about their self-image or reputation. With regard to feedback content, we

hypothesized that adding verification cues would make the feedback more evaluative. With regard to feedback source, we hypothesized that adding person-based feedback would make the feedback more evaluative. Based on these broad hypotheses, we formed three specific predictions. First, we expected correct-answer feedback *with* verification to reduce accuracy, strategy exploration, and persistence relative to correct-answer feedback alone. Second, we expected computer-based feedback *with* verbal feedback from a person to reduce accuracy, strategy exploration, and persistence relative to computer-based feedback alone. Third, we expected the combination of verification cues and person-based feedback to produce the lowest outcomes given that this includes both components that are more evaluative in nature.

Method

The study materials, de-identified data, and analytic code are available on the Open Science Framework (<https://osf.io/jhqvb>). The study and analysis plan were not preregistered.

Participants

Original participants were 142 children aged 6 to 8 years, which was in line with our original goal to recruit a minimum of 100 children. The participants were primarily recruited using a database of contact information for local families hosted by the authors' university in a midwestern community in the United States. Some families signed up via a flyer posted in the community or an advertisement posted on ChildrenHelpingScience.com, a website where parents and children can view available research studies and opt in to participate. Parent consent and child assent were obtained for all participants, and the full study was approved by the institutional review board at Indiana University.

Twelve children were excluded from analyses due to experimenter or technology error ($n = 8$) or due to a failed manipulation check ($n = 4$). The final analytic sample contained 130 children ($M_{\text{age}} = 7.61$ years, $SD = 0.86$, range = 6.13–8.99). A power analysis in G*Power 3.1 indicated that a sample size of 130 provides 80% power to detect a medium-sized effect ($f = .25$) with four groups in a factorial analysis of variance (ANOVA) where the effect has a numerator degrees of freedom of 1 and alpha is set to .05. Optional demographic information was provided by a subset of parents; based on this information, the sample was 36% female and 64% male ($n = 124$ choosing to report) and 66% White, 18% multiracial, 12% Asian, 2% Black or African American, and 1% Hispanic ($n = 118$ choosing to report). Each participant completed a single online session via Zoom with a trained researcher. Data from a portion of this sample are reported in Merrick & Fyfe (2023) to answer separate research questions focused on children's emotional experiences during mathematics problem solving.¹

Design

The study had a 2×2 between-participants design with children randomly assigned to condition. The condition manipulation occurred in the context of a training game in which children progressed through a series of levels by solving math problems with feedback. After solving each problem, children received feedback that varied on two dimensions: feedback content (correct answer alone vs. correct answer with verification) and feedback source (computer alone vs. computer with person). Thus, children were assigned to one of four feedback conditions: computer alone with verification ($n = 33$), computer alone without verification ($n = 32$), computer and person with verification ($n = 32$), or computer and person without verification ($n = 33$).

Conditions were well-matched on background characteristics given that there were no significant differences between conditions in terms of age, gender (i.e., proportion female), or race (i.e., proportion White) ($ps > .05$). We were primarily interested in children's performance and persistence during

¹ The conclusions reported in this article and the conclusions reported in Merrick & Fyfe (2023) are based on distinct data points. The only overlap is that Merrick & Fyfe (2023) contains descriptive statistics (means and standard deviations) for training accuracy and training persistence for a subset of the sample ($n = 87$); it does not report condition differences on these two variables.

the training game, but the study design also included a brief baseline measure prior to the game as well as several post-game survey items.

Materials

All materials were presented on a computer screen via the video-call platform Zoom due to COVID-19 pandemic data collection protocols. The experimenter prepared all the measures using Keynote presentation software. During the session, the experimenter shared their screen so that children could view all the materials and measures on their own digital device.

Baseline items

Four baseline items were administered to assess children's prior knowledge of math equivalence (adapted from McNeil et al., 2011). For the *equal-sign definition* item, children saw an image of the equal sign presented on the screen and were asked to provide a verbal definition of the symbol. For the three *problem-solving* items, children solved three math equivalence problems (see Appendix A). All three items were of the form $a + b = c + _$. The first two items were presented using pictures of objects (e.g., two seashells plus seven seashells on the left and six seashells plus a blank on the right). The final item was presented using written numbers (e.g., $1 + 5 = 2 + _$) but was accompanied by a concrete narrative about sharing stickers.

Training game items

The training game included five sets of 4 problems each for a total of 20 symbolic math equivalence items (see Appendix A). Within each set, the 4 items were related in that they were each equivalent to the same value. For example, the 4 items in Level 1 were each equivalent to 10. Across sets, the items increased in difficulty. For example, Level 1 included items with operations on just one side of the equal sign (e.g., $3 + _ = 10$), and items with operations on both sides of the equal sign (e.g., $3 + 7 = 10 + _$). In contrast, Level 5 included only items with operations on both sides of the equal sign, some with four addends (e.g., $8 + 10 = 7 + _$) and some with five addends (e.g., $6 + 4 + 8 = 3 + _$).

Manipulation check items

Two brief survey items were administered after the training game. To check whether children found the feedback useful, the experimenter asked the children how helpful it was to be shown the correct answer after each question. Children were presented with three options: "a little," "a medium amount," and "a lot." Three children refused to select an option because they thought it was not helpful at all, and two children refused to select an option because they did not know. Of the 125 children who selected a response, 42% selected "a lot," 30% selected "a medium amount," and 28% selected "a little." Selections did not vary by condition, $\chi^2(6, N = 125) = 7.42, p = .284$.

To check whether children detected the different sources of feedback, the experimenter presented the children with two options and asked them to verbally say who provided the correct answer after each math problem; "me" corresponded to a stick figure of a person, and "the computer" corresponded to a cartoon image of a computer. Four children were excluded from analysis because they were in the computer alone condition but indicated that the experimenter provided feedback during the game. All other children correctly identified the feedback source.

Procedure

Each child completed a single one-on-one experimental session conducted over Zoom. The experimenter logged on to the video call from a lab-based computer, and the child participants joined the video call from their own device, typically in their home, using a link that was shared via e-mail. Each session lasted approximately 30 min.

Prior to training game

The experimenter began the video call by ensuring that parental consent and child assent were provided. Then, the experimenter provided brief instructions on how to set up the screen so that

the children could view all the necessary information. Once the initial setup was complete, children completed the four baseline measure items. These were introduced as warm-up activities to the learning game. For the equal-sign item, the equal-sign symbol was presented on the screen along with three other symbols. Children pointed to the equal sign and then verbally defined it. For the three problem-solving items, the experimenter introduced children to bunny's side of the math problem and doggy's side of the math problem and then asked the children to help bunny and doggy have the same amount. Each item was presented one at a time on the screen. Children gave their answers as a verbal response, and no feedback was given.

Training game

The experimenter then introduced the main activity as a "new computer game" that we were trying out because "it is supposed to help kids practice math problems." The experimenter first showed children an example screen to familiarize them with the rules of the game. Children were told that a problem would appear on the screen, and they were trained on how to type and submit their answers via the chat function on the video call. They were also told that the game had many levels and that after each set of 4 problems they could choose to go on to the next level by typing "go" or they could choose to end the game by typing "stop." Children then proceeded to play the game by solving math problems one at a time. A math problem appeared, children typed their answer via the chat box, and then they received corrective feedback before proceeding to the next math problem. The game ended when children typed "stop" in the chat or when they completed all 20 problems. Figs. 1 and 2 depict the feedback conditions.

Computer alone, correct answer only. In this condition, the experimenter "left the game" (i.e., muted her microphone and turned off her video) and let children play on their own. In reality, the experimenter was still on the call and advancing the slides. The experimenter said, "I have to work on something while you play the computer game. So you can play the game with the computer all by yourself, and the computer will give you the answers to the questions that you complete. I will come back as soon as you are done playing. Type 'go' to start the game." Children then solved the problems and viewed the feedback on their own. Whether the children answered correctly or incorrectly, the feedback displayed the correct answer in blue on the computer screen and it was accompanied by a neutral auditory tone.

Computer alone, correct answer with verification. In this condition, the experimenter also "left the game" and explained that children would play the game on the computer "all by yourself." For the feedback, children saw the correct answer in blue on the computer screen and also received verification cues. If children's answer was correct, the blue answer was accompanied by a green check mark on the screen and an ascending tone. If children's answer was incorrect, the blue answer was accompanied by a red X on the screen and a descending tone.

Computer with person, correct answer only. In this condition, the experimenter stayed (i.e., kept her microphone and video on) and said, "I will stay while you play the game by yourself. I will be here to give you the answers to the questions that you complete. Type 'go' to start the game." Children then solved the problems. Whether the children answered correctly or incorrectly, the feedback displayed the correct answer in blue on the computer screen and it was accompanied by a neutral auditory tone. The correct was also spoken aloud by the experimenter (e.g., "The correct answer to this problem is 4").

Computer with person, correct answer with verification. In this condition, the experimenter stayed and explained that she would provide the answers as feedback. For the feedback, children saw the correct answer in blue on the computer screen and also received verification cues. If children's answer was correct, the blue answer was accompanied by a green check mark on the screen and an ascending tone, and the experimenter also spoke aloud to provide verification and the answer (e.g., "That's right! The correct answer is 4."). If children's answer was incorrect, the blue answer was accompanied by a red X

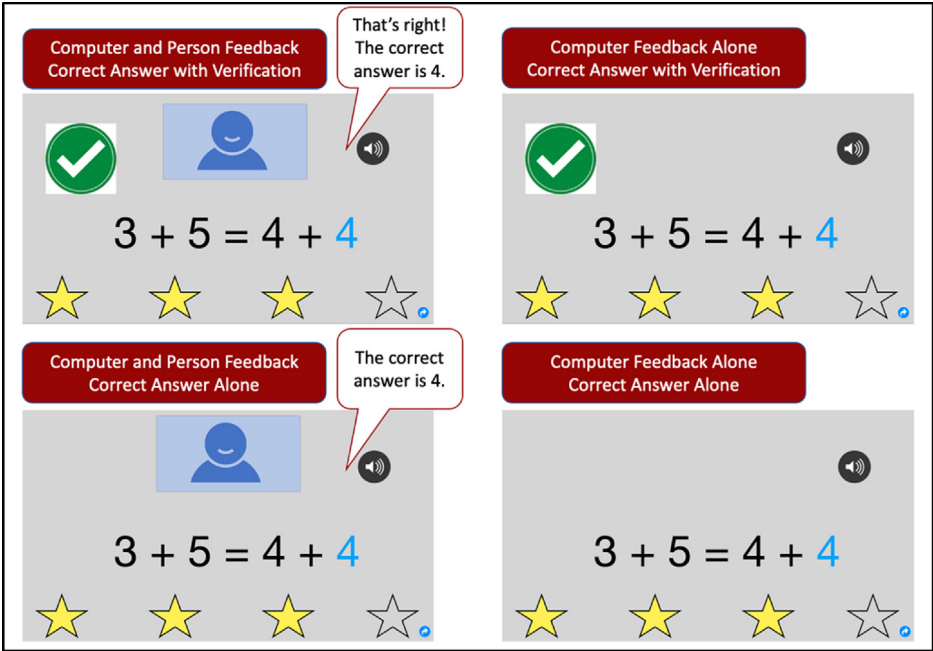


Fig. 1. Visual schematic of the feedback on correct answers across conditions.

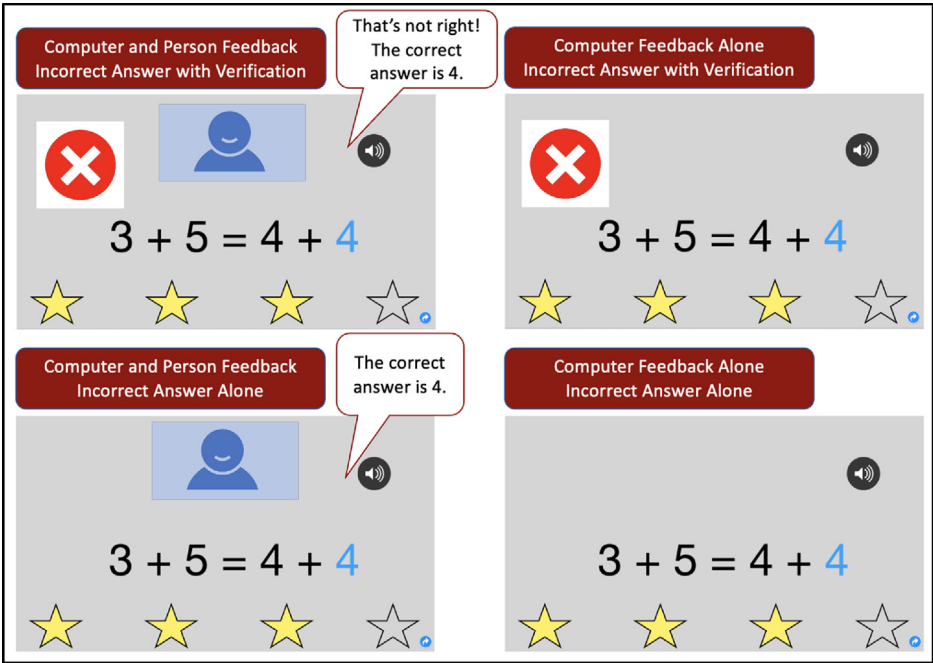


Fig. 2. Visual schematic of the feedback on incorrect answers across conditions.

on the screen and a descending tone, and the experimenter also spoke aloud to provide verification and the answer (e.g., “Uh oh, that’s not right. The correct answer is 4.”).

After the training game

Once the game ended, the experimenter asked two questions about children’s perceptions of the feedback—whether it was helpful and who was providing it. The children and parents answered a few additional questions (e.g., sound and video quality, re-asked about consent for sharing video). Finally, the experimenter thanked the children and ended the Zoom call.

Coding

Baseline measure

We coded children’s accuracy on the baseline items to determine their prior knowledge. For the equal-sign definition, responses were awarded 1 point if they included a relational definition (e.g., “it means the same as,” “same amount on both sides”). Given the subjective nature of this scoring, a second researcher independently scored 40% of the definitions, and agreement was high (94% agreement, $\kappa = .74$). For the three problem-solving items, responses were awarded 1 point if the exactly correct numerical solution was provided.

Training game

Based on children’s behaviors during the game, we coded their accuracy, strategy use, persistence, and resilience. For *accuracy*, children were assigned an accuracy score based on the proportion of items they solved using an exactly correct numerical solution. We then coded children’s *strategy use* based on their numerical solutions using a system from prior research (e.g., McNeil & Alibali, 2005). Table 1 presents examples of each strategy. We were primarily interested in the frequency of using two specific incorrect strategies: the “add all” and “add to equal” strategies. These strategies are common incorrect strategies, and they are thought to reflect entrenched knowledge from traditional arithmetic practice (e.g., for $4 + 3 = __$, adding all the numbers produces a correct response). We assessed whether the different feedback conditions prevented children’s strategy exploration, represented by the over-reliance on these two entrenched strategies. A second researcher independently coded the strategies for 40% of the sample, and agreement was high (97% agreement, $\kappa = .94$).

For *persistence*, children were assigned a persistence score based on how many questions they chose to complete before deciding to end the game (minimum of 4, maximum of 20). We also calculated a measure of *resilience* to reflect that some children chose to keep playing despite solving many items wrong. We multiplied the number of questions completed by the inverse of their accuracy score [i.e., $\text{Resilience} = \text{Questions Completed} * (1 - \text{Accuracy})$]. This means that children who solved all the questions correctly necessarily scored a zero on resilience given that they never experienced failure, so they were excluded from analyses on this measure. The minimum possible value was zero, and the maximum possible value was 20 (e.g., a child who solved all the items wrong yet continued to answer all 20 questions).

Results

Our goal was to examine the effects of feedback condition on children’s training outcomes.² Below we include three main subsections of results: (a) preliminary analyses on the baseline measure to ensure that training conditions were balanced at the outset of the study; (b) primary analyses focused on how feedback influenced children’s problem-solving solutions, including their overall accuracy and the types of strategies they employed; and (c) primary analyses focused on how feedback influenced children’s motivation, including their persistence to keep playing the game and their resilience in the face of failure.

² The first item during training occurred prior to any feedback. Exploratory analyses indicated no condition differences on that item, and they also indicated that the inclusion versus exclusion of that first item did not change any conclusions. We included data from the first item in the outcome measures to maximize the data used.

Table 1
Strategies children used on the mathematical equivalence problems

Strategy	Description	Example problem: 6 + 4 + 8 = <u> </u> + 3	Computer and person: Yes verification	Computer and person: No verification	Computer alone: Yes verification	Computer alone: No verification	Total
Correct	Provides the exactly correct numerical solution that makes both sides the same amount	6 + 4 + 8 = <u>15</u> + 3	54 (33)	67 (35)	67 (30)	69 (32)	64 (33)
Add all	Adds all the numbers in the problem	6 + 4 + 8 = <u>21</u> + 3	16 (22)	8 (13)	12 (15)	7 (20)	11 (18)
Add to equal	Adds all the numbers prior to the equal sign	6 + 4 + 8 = <u>18</u> + 3	7 (12)	5 (6)	9 (12)	8 (12)	7 (10)
Carry	Copies a number from the left side of the equal sign to the blank on the right side	6 + 4 + 8 = <u>6</u> + 3	9 (11)	10 (16)	6 (11)	6 (11)	8 (13)
Repeat	Repeats a number from the right side of the equal sign to the blank on the right side	6 + 4 + 8 = <u>3</u> + 3	1 (3)	1 (3)	1 (3)	1 (1)	1 (3)
Add two	Adds two random numbers in the problem	6 + 4 + 8 = <u>10</u> + 3	3 (8)	2 (6)	2 (8)	3 (6)	2 (6)
Arithmetic error	Numerical response is within one of the exactly correct numerical solution	6 + 4 + 8 = <u>14</u> + 8	2 (2)	3 (7)	1 (2)	1 (3)	2 (5)
Other incorrect	Incorrect numerical solution that does not fall into one of the above categories	6 + 4 + 8 = <u>45</u> + 3	8 (8)	3 (6)	3 (8)	4 (5)	5 (10)

Note. Values represent the means (and standard deviations) for the use of each strategy by condition.

Our primary analyses relied on a series of $2 \times 2 \times 2$ ANOVAs with feedback content (correct answer alone vs. correct answer with verification), feedback source (computer alone vs. computer and person), and prior knowledge (high vs. low) entered as between-participants factors along with their interactions.³ Detailed statistics for each model (including main effects and interaction effects) are provided in the only [supplementary material](#). In the main text, we report only detailed statistics for effects that are identified as statistically significant at the $\alpha = .05$ level.

Baseline performance

On average, children solved 2.69 baseline problems correctly out of 4 ($SD = 0.85$). As expected, performance on the two nonsymbolic equations (i.e., with pictures of objects) was higher (95% and 88% correct) than on the symbolic equation (72% correct), and performance defining the equal sign was lowest (15% correct). The conditions were well-matched in their baseline performance (see [Table 2](#)). To examine this statistically, we ran a 2×2 ANOVA with feedback content and feedback type entered as between-participants factors predicting baseline scores out of 4. This model revealed no main effects or an interaction ($ps > .05$). We also ran a chi-square analysis and confirmed that the proportion of high-knowledge children (i.e., scoring a 3 or 4 out of 4) was not significantly different across conditions, $\chi^2(3, N = 130) = 3.75, p = .289$.

We included baseline performance in subsequent models to explore whether the effects of condition depended on prior knowledge. Given the distribution of baseline scores (median = 3, mode = 3), and given that two items were substantially easier than the others, we categorized children who scored 0, 1, or 2 on the baseline as the low-knowledge group ($n = 43$) and children who scored 3 or 4 on the baseline as the high-knowledge group ($n = 87$).

Training performance: Problem-solving solutions

Accuracy

We first examined accuracy on Level 1 (the first four items) to get a measure of performance based on all children solving the same number of items at the same level of difficulty. Overall, children's performance on Level 1 was modest, with an average of 65% correct ($SD = 33$, range = 0–100). A $2 \times 2 \times 2$ ANOVA with feedback content, feedback source, and prior knowledge revealed no significant main effects or interactions ($ps > .05$).

We then examined accuracy across all five levels of the training game by calculating a percentage score for each child based on the number of items children solved correctly out of the number of items they attempted. The average score was 64% ($SD = 33$, range = 0–100). The $2 \times 2 \times 2$ ANOVA revealed a large main effect of prior knowledge, $F(1, 122) = 29.23, p < .001, \eta_p^2 = .19$, with children in the high-knowledge group having higher accuracy ($EM = 74\%, SE = 3$) than children in the low-knowledge group ($EM = 44\%, SE = 5$). There was also a small main effect of feedback source, $F(1, 122) = 4.49, p = .036, \eta_p^2 = .04$, with children who received feedback from the computer alone having higher accuracy ($EM = 65\%, SE = 4$) than children who received feedback from the computer and person ($EM = 53\%, SE = 4$). There was not a main effect of feedback content ($p = .316$), and none of the interactions was statistically significant.

Because the training items got more difficult as the levels progressed, it is possible that accuracy scores were influenced by persistence. Children who persisted longer in the game might have lower scores because they were solving harder items, not necessarily because of the type of feedback they received. However, the data suggest that this was not the case. In fact, accuracy scores and persistence scores were actually positively correlated, $r(128) = .416, p < .001$, suggesting that children who chose to solve more items tended to also have higher accuracy scores. In addition, we re-ran the $2 \times 2 \times 2$ ANOVA predicting accuracy scores but including persistence (i.e., the number of items completed) as a covariate. The results remained unchanged. There was still a significant main effect of feedback source

³ The online supplementary material contains information about simpler 2×2 models (content and source) that do not include prior knowledge.

Table 2
Descriptive statistics by condition

Measure	Computer and person: Yes verification <i>M (SD)</i>	Computer and person: No verification <i>M (SD)</i>	Computer alone: Yes verification <i>M (SD)</i>	Computer alone: No verification <i>M (SD)</i>	Total <i>M (SD)</i>
Baseline score (out of 4)	2.78 (0.71)	2.70 (0.81)	2.61 (0.97)	2.69 (0.93)	2.69 (0.85)
Level 1 accuracy (% correct)	57.03 (34.33)	68.94 (31.89)	64.39 (33.09)	71.09 (33.67)	65.38 (33.29)
Total accuracy (% correct)	54.24 (33.47)	66.93 (34.74)	66.55 (29.68)	69.23 (31.82)	64.28 (32.63)
Number of different strategies (out of 8)	3.47 (1.97)	3.15 (1.94)	2.97 (1.33)	2.91 (1.87)	3.12 (1.79)
Arithmetic strategy (% of all trials)	22.40 (22.67)	12.71 (15.39)	19.82 (20.82)	15.55 (18.78)	17.60 (19.71)
Persistence (out of 20)	13.63 (6.08)	15.52 (5.27)	13.21 (6.74)	13.87 (6.50)	14.06 (6.17)
Resilience (out of 20)	5.63 (4.91)	5.81 (4.93)	3.93 (2.99)	4.79 (4.25)	5.05 (4.35)

Note. For resilience, scores were based on the 108 children who solved at least one training item incorrectly. For the remaining measures, scores were based on the full sample of 130 children. A small portion of the data from this table overlaps with data in Merrick & Fyfe (2023), which contains descriptive statistics for a portion of this sample ($n = 87$) focused on children's total accuracy and their persistence.

on children's accuracy scores, $F(1, 121) = 5.59, p = .020, \eta_p^2 = .04$, suggesting a robust effect that was not explained away by differences in persistence.

Strategy use

In addition to accuracy, we also examined children's strategy use to gain insight into their exploratory behaviors during the math game—whether they tended to rely on common incorrect strategies or to try out different strategies. Children's numerical solutions reflected either a correct strategy or one of seven different types of errors (Table 1). On average, children used 3.12 different strategies as they played the game (median = 3, $SD = 1.79$, range = 1–8), which typically reflected children using a correct strategy on a simpler item and using two different incorrect strategies across the remaining items. But this varied by feedback content. A $2 \times 2 \times 2$ ANOVA with number of different strategies used as the outcome measure revealed a small but significant feedback content by prior knowledge interaction, $F(1, 122) = 5.11, p = .026, \eta_p^2 = .04$. Low-knowledge children used a wider variety of different strategies when they received the correct answer alone ($EM = 3.87, SE = 0.40$) than when they also received verification feedback ($EM = 2.97, SE = 0.39$), but the same was not true for high-knowledge children. No other effects in the model were significant ($ps > .05$).

We also examined the frequency of using the two arithmetic-specific incorrect strategies as a way to assess children's reliance on common entrenched knowledge. Across all attempted trials, children used these two arithmetic-specific strategies 18% of the time ($SD = 20$), and these two strategies made up 51% of all the errors. The $2 \times 2 \times 2$ ANOVA predicting the frequency of using these two strategies revealed two significant effects. There was a medium main effect of prior knowledge, $F(1, 122) = 12.04, p < .001, \eta_p^2 = .09$, with high-knowledge children using these strategies less often ($EM = 13\%, SE = 2$) than low-knowledge children ($EM = 26\%, SE = 3$). There was also a small main effect of feedback content, $F(1, 122) = 5.54, p = .020, \eta_p^2 = .04$. Children who received feedback with the correct answer alone used these strategies less often ($EM = 15\%, SE = 3$) than children who received the correct answer with explicit verification ($EM = 24\%, SE = 3$). These results suggest that explicit verification cues in the

feedback message tended to reduce children's exploratory behavior—lowering some children's strategy variability and promoting the use of entrenched arithmetic-based strategies.⁴

Training performance: Motivation to continue

Persistence

Children were generally persistent and opted to keep playing the game. On average, children completed 14.1 questions ($SD = 6.2$); the median was 16 questions (four of the five levels), and the mode was all 20 questions. The $2 \times 2 \times 2$ ANOVA predicting persistence revealed a medium main effect of prior knowledge, $F(1, 122) = 11.13, p = .001, \eta_p^2 = .80$, with high-knowledge children completing more items ($EM = 15.3, SE = 0.6$) than low-knowledge children ($EM = 11.7, SE = 0.9$). There was also a small main effect of feedback content, $F(1, 122) = 4.69, p = .032, \eta_p^2 = .04$, with children who received the correct answer alone completing more items ($EM = 14.7, SE = 0.8$) than children who received the correct answer with verification ($EM = 12.3, SE = 0.8$). These main effects were also qualified by a content by prior knowledge interaction, $F(1, 122) = 11.03, p = .001, \eta_p^2 = .08$. The influence of feedback content was driven by children in the low-knowledge group, who completed more items with the correct answer alone ($EM = 14.7, SE = 1.3$) than with the correct answer plus verification ($EM = 8.7, SE = 1.3$). No other effects in the model were statistically significant ($ps > .05$).

Resilience

Our measure of resilience was based on a combination of accuracy and persistence, and high scores represented children who kept going in the game despite low levels of success. These analyses were based on the 108 children who solved at least one item incorrectly given that those who solved all the items correctly did not have an opportunity to exhibit resilience. The average resilience score was 5.0 ($SD = 4.4$). The $2 \times 2 \times 2$ ANOVA revealed no significant effects ($ps > .05$). However, there were interesting qualitative trends. The vast majority of children (70%) had resilience scores between 1 and 5 (median = 4, mode = 4), and these scores tended to represent two types of children: those with low accuracy who quit after one or two levels and those with high accuracy who kept going. But we were interested in the subset of children who had remarkably high resilience scores, and 18 children had resilience scores of 10 or higher (2.5 times the modal resilience). What was unique about these children? It did not appear to be age or baseline measure performance, nor did it appear to be the feedback content they were exposed to (e.g., 8 had verification and 10 did not). However, 14 of these 18 children (78%) received computer with person feedback, and all 18 used a variety of strategy types. Their average strategy variability was 5.83, whereas the average for the rest of the sample was 3.10 different strategies. Although qualitative, these descriptive results suggest that our most resilient children received feedback from a person and tended to feel comfortable in exploring the problem space.

Discussion

The current study investigated elementary school children's performance while completing an online learning activity and receiving different forms of feedback. The feedback was provided either by the computer alone or with additional verbal feedback from the researcher, and it either contained the correct answer alone or contained additional verification cues. Both features of feedback influenced children's in-the-moment behaviors (see Table 3 for a summary of the results). Feedback *content* influenced strategy use and persistence. Those who received the correct answer alone used a wider variety of strategies, relied less on entrenched strategies, and opted to solve more problems relative to those who received the correct answer with explicit verification cues. These effects were more pronounced for children with low prior knowledge. Feedback *source* primarily influenced accuracy. Those who received feedback from a computer alone solved more problems correctly than those

⁴ We qualitatively explored children's "discovery" of new strategies during the game (e.g., moments of using a novel strategy with continued use afterward), but we were unable to detect any noticeable condition differences.

Table 3
Summary of the statistical findings

Outcome measure	Feedback content	Feedback source
Level 1 accuracy	X	X
Total accuracy	X	✓ (computer alone increases)
Strategy variability	✓ (verification decreases for low-knowledge children)	X
Arithmetic-based strategies	✓ (verification increases)	X
Persistence	✓ (verification decreases for low-knowledge children)	X
Resilience	X	X

Note. A check mark indicates that the variable had a statistically significant influence on that outcome, and an X indicates that the variable had a null effect on that outcome. All statistically significant condition effects had a small effect size.

who received feedback from a computer and a person. We outline the implications of these results for theory and practice.

Feedback content

We originally hypothesized that the addition of verification cues to the correct answer would negatively influence children's outcomes during mathematics problem solving because the cues heighten the evaluative nature of the feedback. This hypothesis was partly confirmed, with the inclusion of verification feedback resulting in detrimental effects across various outcome measures, primarily for children with low prior knowledge. The addition of verification cues reduced children's strategy variability, increased their reliance on entrenched arithmetic strategies, and decreased their persistence during the game. These results suggest that there may be unintended consequences when feedback includes explicit right–wrong cues (e.g., check marks, Xs), especially in relation to children's willingness to explore the problem space.

One potential explanation for the hindering effects of verification cues relates to the *quantity* of information in the feedback. Some children just saw the correct answer filled into the equation (along with neutral auditory cues), whereas others saw that same information plus a visual and auditory cue indicating whether their answer was correct or incorrect. This additional information may have taken more time and attention to process (e.g., Dunlosky et al., 2020; Sweller, 1988) given that children in this condition needed to encode the extra cues, identify the correct answer, and integrate this information to decide whether or not to adjust their strategies on future trials. These tasks may have required more time, for example, to notice and think about the meaning of the check mark. In addition, these tasks may have required more attention, for example, more eye movements between different types of information. And given the limited time between trials, these additional constraints may have reduced children's capacity to explore the space—for example, causing children to try fewer new strategies and instead to rely more on entrenched arithmetic-specific strategies (McNeil, 2014) or to stop the game earlier. This finding might be especially relevant for children given that past research using eye tracking with 11-year-olds found that it can be challenging for them to detect, encode, and process the full contents of feedback (Tärning et al., 2020).

Another potential explanation for the hindering effects of verification cues relates to the *quality* of information. That is, the type of information provided—the explicit indication of a correct or incorrect response—may have directed children's attention toward their self-image rather than the task, reducing their exportation of the problem space. Children who just saw the correct answer may have been more likely to attend solely to the correct answer and how it fit into the equation, thereby allowing them to try new (and potentially correct) strategies on the next trial. This explanation is consistent with the fact that children who received the answer alone used a wider variety of strategies and were less likely to rely on the default entrenched incorrect strategies relative to children who received the verification cues. If the verification cues drew children's attention to the self, it could have produced

emotional responses (see Lipnevich & Smith, 2009a,b; Rowe & Fitness, 2018) that threatened their safety to explore new strategies or their motivation to carry on with the task at hand.

These findings were especially pronounced for children with low prior knowledge, which suggests that a high frequency of *negative* feedback (e.g., indicating an incorrect response) may be likely to reduce strategy variability and persistence. In this way, verification cues may deprive children of an opportunity to problem solve without fear of explicit judgment of their mistakes. These findings align with Kuklick and Lindner's (2021) study with fifth and sixth graders in which different types of verification cues (e.g., in text, sound, color) resulted in lower feelings of pride relative to no feedback, but only for children with low prior knowledge. Previous research attests to some benefits of providing learners with at least some exploratory and low-stakes opportunities in a way that allows failure to be productive and allows children to identify critical problem features (e.g., Bonawitz et al., 2011; Schwartz et al., 2011). Potentially, children just starting to explore a specific task may benefit from an open environment with sufficient information to make progress (i.e., the correct answer) but void of judgment cues.

Regardless of the mechanism, these results have potential implications for administering feedback in practice. In certain contexts, it may be beneficial to eliminate the explicit verification cues and just provide the model response. This recommendation contrasts with existing claims that "effective feedback should include elements of both verification and elaboration" (Shute, 2008, p. 158). These claims are well-intentioned and based on empirical findings; it is often better to provide additional information than to provide *only* verification (e.g., Marsh et al., 2012). But there may be situations where the verification can be removed. Future research is needed to determine the boundary conditions around this effect given that there are likely times when learners need the explicit cues to better identify when they are right versus wrong.

Feedback source

We also hypothesized that the addition of verbal feedback from a person would negatively influence children's outcomes during mathematics problem solving because it could heighten the evaluative nature of feedback. This hypothesis was also partly confirmed, with the results showing that the addition of person feedback to the computer feedback decreased children's accuracy during the task relative to computer feedback alone. We also expected that person feedback would be particularly detrimental when verification cues were present, but that was not confirmed given that we did not detect any reliable interactions between these factors.

Although few studies have directly and experimentally contrasted person-mediated and computer-mediated feedback, the result of improved accuracy with computer-based feedback is consistent with some previous research reporting benefits of computer-based feedback on learning (e.g., Đorić et al., 2021). There are several potential explanations. It seems likely that the addition of person-based input to the computer feedback drew children's attention to their self and reputation, as opposed to the task, and detracted from task-relevant processing (e.g., Kluger & Adler, 1993). Even at young ages, children are attuned to issues related to impression management and the desire to please others (Silver & Shaw, 2018). Drawing attention to one's self may have consequences, especially for activities like the ones in the current study that require full attention to detailed features of the task in order to perform well (e.g., the location of the equal sign amid multiple addends in a math equivalence problem).

However, it is also important to consider that our person-based condition differed in several ways from the computer-based condition: it included (a) the presence of a person's face on the screen during problem solving, (b) verbal feedback provided by the person, and (c) redundant feedback information provided on the screen. This means that it is hard to tease apart effects related to the *presence* of a person versus *feedback* from a person. At least one study suggested that these results may be specific to the feedback; undergraduate students completing math problems performed better when a person was present but not interacting than when a person was present and also providing feedback (Kluger & Adler, 1993). Another explanation is related to redundancy; the person-based feedback could have resulted in lower accuracy because it included the same information presented visually on the screen and verbally by the experimenter, which may have taken unnecessary cognitive effort to process both simultaneously (Dunlosky et al., 2020). Future research is needed to tease apart these explanations

and identify the circumstances under which person-based feedback has consequences for performance.

Certainly, these results should not be interpreted to mean that there are no benefits to having a person deliver feedback. In other contexts, person-based feedback has had empirical advantages (Golke et al., 2015). In fact, even in this study, qualitative results suggested a potential benefit; the majority of the most resilient children in the sample received feedback with a person present. That is, children who received feedback from the person were more likely to keep persisting in the game despite high failure rates. Perhaps the presence of a person made them want to keep going, especially in the non-evaluative learning environment we created. Furthermore, the researcher generally delivered feedback in a calm and encouraging voice, and an extensive literature suggests that humans mirror others' emotional states (Navarretta, 2016). Thus, the presence of a person delivering feedback in a non-evaluative context may have fortified the low-performing children's resilience to keep playing the game. Alternatively, the proximity of the person may have simply increased obedience, with children feeling more obligated to keep playing the game (see Milgram, 1965). Or perhaps the instructions in the computer-alone condition may have led children to devalue the task (e.g., "If the experimenter decided to leave, why should I stay?") and reduced resilience.

Limitations and future directions

Several limitations of the current study suggest directions for future research. One limitation of this study was the lack of environmental control given that the study was performed via a video call in children's home. The researcher had limited control over who or what was distracting the children during the study session. Anecdotally, the most pervasive form of distraction was social in nature (e.g., parents talking to their children), which could potentially influence key achievement outcomes during the session given our focus on manipulating the presence of person-based feedback. On the other hand, this study provides a unique snapshot of children's naturalistic environment as they completed an academic activity in front of family members—in a similar way to how they might complete homework assigned from school.

Future work should also continue to compare the effects of different types of feedback. For example, some research suggests that providing elaborated feedback, such as explanations of the correct response, can be even more beneficial than just providing the correct answer (e.g., Butler et al., 2013). It is an open question whether verification cues would have the same detrimental effects if more elaborate information was available. Furthermore, our person-based feedback was provided virtually as opposed to live, and future research could make the contrast between the computer and person conditions more evident by including feedback from a physically present person. In addition, although we incorporated multiple outcomes including both accuracy and persistence, they were limited to in-the-moment behaviors and did not assess *learning* from the feedback experience on an independent posttest. Previous work suggests that some features of the environment can influence in-the-moment performance without influencing learning or retention and vice versa (Bjork & Bjork, 2011), so future research should include assessments that are administered after the training experience.

Finally, our sample size was fairly limited for detecting interaction effects and also represented a convenience sample. It would be valuable to collect data from a larger, more heterogeneous sample to understand how these features of feedback interact with each other and with a variety of individual differences (e.g., gender, race). For example, one might imagine that children who are underrepresented in math domains may be more sensitive to ego-threatening feedback cues (Spencer et al., 1999; Steele & Aronson, 1995). Future studies could take a more person-centered approach to understand how and for whom these effects of feedback occur.

Conclusion

Feedback is an important educational tool that can be used to alter key achievement outcomes. This study worked to acknowledge that the use of feedback sits in a larger social context of the learning environment. The current findings supported portions of our hypotheses in ways that were aligned

with feedback intervention theory. Feedback with verification cues led to lower persistence and higher reliance on entrenched strategies relative to feedback that contained the correct answer alone, especially for low-knowledge children. Computer feedback alone led to increased accuracy relative to computer feedback supplemented with verbal feedback from a person. Together, these results suggest that there may be benefits to giving children computer-mediated experiences that allow them to receive correct-answer feedback in meaningful and motivating ways without explicit judgment.

CRedit authorship contribution statement

Megan Merrick: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Emily R. Fyfe:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing.

Data availability

The anonymized data and code is available on OSF. The link is provided in the manuscript.

Acknowledgments

Megan Merrick was supported by a training grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (T32 HD007475). The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health. The authors thank Alex Bondi, Collin Byers, Elsie Gasaway, Jessica Ousterhout, Summer Smith, and Olivia Weed for their help with data collection and coding.

Appendix A

List of 4 items provided on the screening measure:

1. Define the equal sign
2. $4 + 3 = 4 + __$, presented as pictures of cookies being distributed to Bunny vs. Doggy
3. $2 + 7 = 6 + __$, presented as pictures of seashells being distributed to Bunny vs. Doggy
4. $1 + 5 = 2 + __$, presented with written numbers but with a narrative about stickers being distributed to Bunny vs. Doggy (i.e., Bunny has one sticker and five more stickers and Doggy has only two stickers. How many more stickers does Doggy need to have an equal amount of stickers as Bunny?).

List of 20 possible items provided during the training game:

Level 1

1. $3 + __ = 10$
2. $10 = 7 + __$
3. $3 + 7 = 3 + __$
4. $3 + 7 = 10 + __$

Level 2

1. $8 = 5 + __$
2. $3 + 5 = 5 + __$
3. $3 + 5 = 4 + __$
4. $7 + 1 = __ + 4$

Level 3

1. $4 + 7 = 4 + \underline{\quad}$
2. $4 + 7 = 3 + \underline{\quad}$
3. $6 + 2 + 3 = \underline{\quad} + 8$
4. $4 + 2 + 5 = 5 + \underline{\quad}$

Level 4

1. $9 + 6 = \underline{\quad} + 5$
2. $5 + 5 + 5 = \underline{\quad} + 6$
3. $5 + 6 + 4 = 4 + \underline{\quad}$
4. $8 + 2 + 5 = \underline{\quad} + 5$

Level 5

1. $8 + 10 = 7 + \underline{\quad}$
2. $5 + 8 + 5 = \underline{\quad} + 6$
3. $6 + 4 + 8 = 3 + \underline{\quad}$
4. $9 + 2 + 7 = \underline{\quad} + 7$

Appendix B. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jecp.2024.105865>.

References

- Ashford, S. J., & Cummings, L. L. (1983). Feedback as an individual resource: Personal strategies of creating information. *Organizational Behavior and Human Performance*, 32(3), 370–398. [https://doi.org/10.1016/0030-5073\(83\)90156-3](https://doi.org/10.1016/0030-5073(83)90156-3).
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In FABBBS Foundation, M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (Vol. 2, pp. 59–68). Worth.
- Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, 120(3), 322–330. <https://doi.org/10.1016/j.cognition.2010.10.001>.
- Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, 105(2), 290. <https://doi.org/10.1037/a0031026>.
- Butler, A. C., & Woodward, N. R. (2018). Toward consilience in the use of task-level feedback to promote learning. *Psychology of Learning and Motivation*, 69, 1–38. <https://doi.org/10.1016/bs.plm.2018.09.001>.
- Comer, C. L. (2007). *Benefits of the task for the delivery of negative feedback* Unpublished doctoral dissertation. Kansas State University.
- Dempsey, J. V. (1993). Interactive instruction and feedback. *Educational Technology*.
- Đorić, B., Lambić, D., & Jovanović, Ž. (2021). The use of different simulations and different types of feedback and students' academic performance in physics. *Research in Science Education*, 51, 1437–1457. <https://doi.org/10.1007/s11165-019-9858-4>.
- Dunlosky, J., Badali, S., Rivers, M. L., & Rawson, K. A. (2020). The role of effort in understanding educational achievement: Objective effort as an explanatory construct versus effort as a student perception. *Educational Psychology Review*, 32, 1163–1175. <https://doi.org/10.1007/s10648-020-09577-3>.
- Eskreis-Winkler, L., & Fishbach, A. (2019). Not learning from failure—The greatest failure of all. *Psychological Science*, 30(12), 1733–1744. <https://doi.org/10.1177/095679761988113>.
- Golke, S., Dörfler, T., & Artelt, C. (2015). The impact of elaborated feedback on text comprehension within a computer-based assessment. *Learning and Instruction*, 39, 123–136. <https://doi.org/10.1016/j.learninstruc.2015.05.009>.
- Grundmann, F., Scheibe, S., & Epstude, K. (2021). When ignoring negative feedback is functional: Presenting a model of motivated feedback disengagement. *Current Directions in Psychological Science*, 30(1), 3–10. <https://doi.org/10.1177/0963721420969386>.
- Hattie, J., & Gan, M. (2011). Instruction based on feedback. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 249–271). Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.

- Hornburg, C. B., Devlin, B. L., & McNeil, N. M. (2022). Earlier understanding of mathematical equivalence in elementary school predicts greater algebra readiness in middle school. *Journal of Educational Psychology*, 114(3), 540–559. <https://doi.org/10.1037/edu0000683>.
- Kluger, A. N., & Adler, S. (1993). Person- versus computer-mediated feedback. *Computers in Human Behavior*, 9(1), 1–16. [https://doi.org/10.1016/0747-5632\(93\)90017-M](https://doi.org/10.1016/0747-5632(93)90017-M).
- Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037/0033-2909.119.2.254>.
- Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science*, 7(3), 67–72. <http://www.jstor.org/stable/20182507>.
- Kuklick, L., & Lindner, M. A. (2021). Computer-based knowledge of results feedback in different delivery modes: Effects on performance, motivation, and achievement emotions. *Contemporary Educational Psychology*, 67, 102001. <https://doi.org/10.1016/j.cedpsych.2021.102001>.
- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1, 279–308. <https://doi.org/10.1007/BF01320096>.
- Lipnevich, A. A., & Smith, J. K. (2009a). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied*, 15(4), 319–333. <https://doi.org/10.1037/a0017841>.
- Lipnevich, A. A., & Smith, J. K. (2009b). "I really need feedback to learn": Students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation and Accountability*, 21, 347–367. <https://doi.org/10.1007/s11092-009-9082-2>.
- Marsh, E. J., Loizito, J. P., Umanath, S., Bjork, E. L., & Bjork, R. A. (2012). Using verification feedback to correct errors made on a multiple-choice test. *Memory*, 20(6), 645–653. <https://doi.org/10.1080/09658211.2012.684882>.
- Matthews, P. G., & Fuchs, L. S. (2020). Keys to the gate? Equal sign knowledge at second grade predicts fourth-grade algebra competence. *Child Development*, 91(1), e14–e28. <https://doi.org/10.1111/cdev.13144>.
- McGinn, K. M., Lange, K. E., & Booth, J. L. (2015). A worked example for creating worked examples. *Mathematics Teaching in the Middle School*, 21(1), 26–33. <https://doi.org/10.5951/mathteacmiddscho.21.1.0026>.
- McNeil, N. M. (2014). A change-resistance account of children's difficulties understanding mathematics equivalence. *Child Development Perspectives*, 8(1), 42–47. <https://doi.org/10.1111/cdep.12062>.
- McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development*, 76(4), 883–899. <https://doi.org/10.1111/j.1467-8624.2005.00884.x>.
- McNeil, N. M., Fyfe, E. R., Petersen, L. A., Dunwiddie, A. E., & Brletic-Shipley, H. (2011). Benefits of practicing $4 = 2 + 2$: Nontraditional problem formats facilitate children's understanding of mathematics equations. *Child Development Perspectives*, 82(5), 1620–1633. <https://doi.org/10.1111/j.1467-8624.2011.01622.x>.
- McNeil, N. M., Hornburg, C. B., Brletic-Shipley, H., & Matthews, J. M. (2019). Improving children's understanding of mathematical equivalence via an intervention that goes beyond nontraditional arithmetic practice. *Journal of Educational Psychology*, 111(6), 1023–1044. <https://doi.org/10.1037/edu0000337>.
- Merrick, M., & Fyfe, E. R. (2023). Feelings on feedback: Children's emotional responses during mathematics problem solving. *Contemporary Educational Psychology*, 74, 102209. <https://doi.org/10.1016/j.cedpsych.2023.102209>.
- Milgram, S. (1965). Some conditions of obedience and disobedience to authority. *Human Relations*, 18(1), 57–76. <https://doi.org/10.1177/001872676501800105>.
- Mulliner, E., & Tucker, M. (2017). Feedback on feedback practice: Perceptions of students and academics. *Assessment & Evaluation in Higher Education*, 42(2), 266–288. <https://doi.org/10.1080/02602938.2015.1103365>.
- Navaretta, C. (2016). Mirroring facial expressions and emotions in dyadic conversations. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 469–474). European Language Resources Association <https://aclanthology.org/L16-1075>.
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 3–8. <https://doi.org/10.1037/02787393.31.1.3>.
- Roper, W. J. (1977). Feedback in computer assisted instruction. *Programmed Learning and Educational Technology*, 14, 43–49.
- Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. M. (2014). Skype me! Socially contingent interactions help toddlers learn language. *Child Development*, 85(3), 956–970. <https://doi.org/10.1111/cdev.12166>.
- Rowe, A. D., & Fitness, J. (2018). Understanding the role of negative emotions in adult learning and achievement: A social functional perspective. *Behavioral Sciences*, 8(2), Article 27. <https://doi.org/10.3390/bs8020027>.
- Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103(4), 759–775. <https://doi.org/10.1037/a0025140>.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>.
- Silver, I. M., & Shaw, A. (2018). Pint-sized public relations: The development of reputation management. *Trends in Cognitive Sciences*, 22(4), 277–279. <https://doi.org/10.1016/j.tics.2018.01.006>.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4–28. <https://doi.org/10.1006/jesp.1998.1373>.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811. <https://doi.org/10.1037/0022-3514.69.5.797>.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7).

- Tärning, B., Lee, Y. J., Andersson, R., Månsson, K., Gulz, A., & Haake, M. (2020). Assessing the black box of feedback neglect in a digital educational game for elementary school. *Journal of the Learning Sciences*, 29(4–5), 511–549. <https://doi.org/10.1080/10508406.2020.1770092>.
- Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, 85(4), 475–511. <https://doi.org/10.3102/0034654314564881>.
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, 3087. <https://doi.org/10.3389/fpsyg.2019.03087>.