



Confidence in COVID problem solving: What factors predict adults' item-level metacognitive judgments on health-related math problems before and after an educational intervention?

Daniel A. Scheibe¹ · Charles J. Fitzsimmons¹ · Marta K. Mielicki¹ · Jennifer M. Taber¹ · Pooja G. Sidney² · Karin Coifman³ · Clarissa A. Thompson¹

Received: 30 September 2021 / Accepted: 12 April 2022 / Published online: 24 May 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The advent of COVID-19 highlighted widespread misconceptions regarding people's accuracy in interpreting quantitative health information. How do people judge whether they accurately answered health-related math problems? Which individual differences predict these item-by-item metacognitive monitoring judgments? How does a brief intervention targeting math skills—which increased problem-solving accuracy—affect people's monitoring judgments? We investigated these pre-registered questions in a secondary analysis of data from a large Qualtrics panel of adults ($N=1,297$). Pretest performance accuracy, math self-efficacy, gender, and math anxiety were associated with pretest item-level monitoring judgments. Participants randomly assigned to the intervention condition, relative to the control condition, made higher monitoring judgments post intervention. That is, these participants *believed* they were more accurate when answering problems. Regardless of experimental condition, those who actually were correct on health-related math problems made higher monitoring judgments than those who answered incorrectly. Finally, consistent with prior research, math anxiety explained additional variance in monitoring judgments beyond trait anxiety. Together, findings indicated the importance of considering both objective (e.g., problem accuracy) and subjective factors (e.g., math self-efficacy, math anxiety) to better understand adults' metacognitive monitoring.

Keywords Math cognition · Metacognition · Monitoring judgments · Math self-efficacy · COVID-19 · Math anxiety

When people solve problems, they often have a sense of how confident they are about their accuracy (Ackerman & Thompson, 2017; Desender et al., 2018; Koriat & Adiv, 2016;

✉ Daniel A. Scheibe
dscheib2@kent.edu

¹ Kent State University (Psychological Sciences), Kent, OH, USA

² University of Kentucky (Psychological Sciences), Lexington, KY, USA

³ Kent State University (Clinical Psychology), Kent, OH, USA

Peters et al., 2019). A primary research goal in the study of metacognition is to decipher the bases of metacognitive monitoring judgments¹ (i.e., confidence judgments; Dunlosky & Thiede, 2013; Koriat & Adiv, 2016). In general, people tend to report greater confidence when they are correct compared to when they are incorrect (Nelson & Fyfe, 2019; Rinne & Mazzocco, 2014; Rivers et al., 2020; Wall et al., 2016). This alignment between performance and confidence reflects accurate metacognitive monitoring. The accuracy of monitoring judgments is important because people can, and often do, use these judgments to make decisions such as whether to share their opinions, seek help, check their work for errors, or submit their work for evaluation (e.g., Dunlosky & Metcalfe, 2009; Dunlosky & Rawson, 2012; Nelson & Fyfe, 2019; Wall et al., 2016). However, people's judgments sometimes misalign with their performance (Dunning et al., 2003; Kruger & Dunning, 1999). This misalignment could be especially problematic in applied contexts such as health-decision making. If people are unaware of their own lack of understanding of health metrics, including probabilities relating to risk, they may make decisions based on faulty interpretations of numerical health information. For example, if people mistakenly believe with high confidence that they have an insignificant risk of disease, they might be more willing to engage in risky behaviors or less likely to engage in preventive health behaviors. Thus, it is important to evaluate what factors influence monitoring judgments during problem solving involving health statistics. The current study examined predictors of adults' item-level monitoring judgments in the context of health-related math problems about COVID-19 (hereafter referred to as *problems*; see Fig. 2).

Why are judgments sometimes misaligned with performance?

Misalignment between people's metacognitive judgements and accuracy occurs because a wide variety of cues inform monitoring judgments. According to cue-utilization theory (Koriat, 1997; Koriat & Levy-Sadot, 1999), monitoring judgments can be informed by information- and experience-based cues. Information-based cues include beliefs and knowledge about the self (e.g., Händel et al., 2020), about cognition in general (e.g., Rivers et al. 2020), or about the task at hand (e.g., Mueller & Dunlosky, 2017). Experience-based cues arise from people's experiences during the task, such as how quickly they solved a problem (e.g., Desender & Sasanguie, 2021; Koriat et al., 2008; Leonesio & Nelson, 1990; Sanchez & Dunning, 2020), their familiarity with the problem features (Fitzsimmons & Thompson, 2021; Fitzsimmons et al., 2020a; Reder & Ritter, 1992), or whether they received feedback or instruction on the problem (Labuhn et al., 2010). The accuracy of judgments is dependent upon the validity of the cues that are used (Koriat, 1997; see also Ackerman & Thompson, 2017).

In the current study (pre-registered on OSF: <https://osf.io/vxm8d>), we examined which information-based and experience-based cues influenced monitoring judgments using

¹ There are many different types of metacognitive judgments (see Dunlosky & Metcalfe, 2009; Rhodes, 2019 for reviews). In the current study, participants made monitoring judgments on a scale from 0–100% confidence about the accuracy of a problem they just completed. These types of metacognitive judgments are considered monitoring judgments. We use the term *monitoring judgments* to refer to participants' judgments (see Fig. 2 for the exact questions).

data from a publicly available data set (Thompson et al., 2021). In Thompson and colleagues' study, adults solved a series of problems and indicated, "How confident are you in your decision about which disease is the most fatal from 0%=I am not confident at all, to 100%=I am totally confident?" after solving each of the problems. Our interdisciplinary team examined which information- and experience-based cues influenced these monitoring judgments. Specifically, we: (a) explored whether information-based cues, such as math self-efficacy and math anxiety impacted monitoring judgments, (b) examined how experience-based cues, such as worked-example training impacted monitoring judgments, and (c) assessed whether the magnitude of monitoring judgments was higher when people answered questions correctly versus incorrectly. Next, we discuss relevant literature pertaining to each of these information-based and experience-based cues.

Information-based cues

Math self-efficacy

One type of information-based cue that can influence monitoring judgments during math tasks is an individual's math self-efficacy (Efklides, 2006; Händel et al., 2020; Stankov et al., 2012). Math self-efficacy is defined in the current study as one's beliefs about one's competence for completing specific tasks (Bandura, 1982; Lee, 2009; Pajares, 1996). In general, people tend to be less confident in their math ability compared to other domains, such as reading (Dowker et al., 2016; Punaro & Reeve, 2012; Wigfield & Meece, 1988), and it is common for people to endorse the idea: "I am not a math person" (Dowker et al., 2016; Miller-Cotto & Lewis, 2020; Nolen et al., 2014; Peters, 2020; Peters et al., 2019). Note that math self-efficacy is generally lower among women (Ashcraft & Ridley, 2005; Dowker et al., 2016; Hembree, 1990; Morony et al., 2013; Pajares & Miller, 1994), and may contribute to the underrepresentation of women in math-heavy careers (Huang et al., 2019), thus we also examined gender in our models.

Problem-solvers' beliefs about their own math ability are related to both task-specific performance (Pajares & Miller, 1994) and academic math performance (Ahmed et al., 2012). Because these constructs are related to math performance, people may draw on these math beliefs when making monitoring judgments. That is, information (e.g., attitudes and prior knowledge) and experiences (e.g., feelings elicited in the moment) are sources of inferential information that influence judgments (cf. cue-utilization theory, Koriat, 1997). In the current study, we considered math self-efficacy to be an information-based cue which was guided primarily by prior experiences with math and could influence item-by-item monitoring judgments. Given that the role of math perceptions' influence on monitoring judgments may be underestimated in the literature (Händel et al., 2020), and that many people report low confidence in their math ability (Ashcraft & Ridley, 2005; Barroso et al., 2020; Dowker et al., 2016; Gough, 1954; Hembree, 1990), we tested whether math self-efficacy accounted for unique variance in monitoring judgments, even after accounting for performance accuracy.

In the current study, we adopted Peters et al. (2019) approach to evaluating math self-efficacy by analyzing the first four items of the Subjective Numeracy Scale (SNS: Fagerlin et al., 2007; Peters, 2020). These items from the SNS assess participants' confidence in their math skills pertaining to fractions and percentages (e.g., "How good are you at working with fractions?"). In our view, these SNS questions are better characterized as

a measure of participants' perceptions about their math ability (i.e., math self-efficacy), specifically for rational numbers (i.e., ratios such as fractions and percentages), as opposed to math more generally. This view of these SNS items is supported by strong correlations between the items and objective measures of numeracy—how good people are at working with probability and math concepts (Lipkus et al., 2001; Peters, 2020; Schwartz et al., 1997). Moreover, these SNS items are also strongly correlated with health decision-making measures (Låg et al., 2014; Peters, 2020; Peters et al., 2019; Waters et al., 2018), because health statistics are commonly presented as rational numbers (e.g., the ratio of people who experience side effects of a medication relative to all the people who take the medication). We expected that math self-efficacy (operationalized as responses to the first four SNS items) would contribute to, but not completely overlap with, monitoring judgments. That is, higher math self-efficacy would be positively related to monitoring judgments. Other individual differences, such as math and trait anxiety, may also contribute towards monitoring judgments.

Math anxiety and trait anxiety

An information-based cue, math anxiety, is defined as a fear or apprehension about mathematics (Ashcraft, 2002). When individuals who are math anxious find themselves in math-intensive situations, their anxiety may influence their monitoring judgments. That is, their apprehension about math in general may lead them to indicate they are less confident in their answer regardless of whether they are accurate or whether other cues are available to indicate they should have higher confidence. Because math anxiety is an information-based cue of which people have at least some awareness, higher levels of math anxiety could be a cue that might influence participants to report lower monitoring judgments (Desender and Sasanguie, 2021; Jain & Dowson, 2009).

In the current study, we tested the unique relations between monitoring judgments, math anxiety, and trait anxiety. There is an ongoing debate whether math anxiety is a distinct construct from other forms of anxiety (e.g., trait anxiety) given the parallels between math anxiety and generalized anxiety. For example, similar to generalized anxiety (Cresswell et al., 2010; Ooi et al., 2015), math anxiety may be modeled by teachers and parents who exhibit high math anxiety (Beilock et al., 2010; Ramirez & Beilock, 2011). Meta-analytic reviews (Barroso et al., 2020; Hembree, 1990; Ma, 1999; Namkung et al., 2019; Zhang et al., 2019) support the idea that mathematics anxiety is distinct from general anxiety. Indeed, some researchers consider mathematics to be a particularly common context for heightened anxiety, which can have functional implications similar to specific phobias (Ashcraft, 2019; Ashcraft & Ridley, 2005; Cipora et al., 2019; Dreger & Aiken, 1957; Gough, 1954; Núñez-Peña et al., 2014).

We anticipated that participants' math anxiety would be more closely (and negatively) related to participants' monitoring judgments about their math performance than trait anxiety. In Thompson et al. (2021), math anxiety predicted problem accuracy even when trait anxiety was also included in the model. This indicated that participants who were less likely to be accurate on their problem-solving performance were not just more anxious in general. Thompson and colleagues collected trait anxiety data as a critical variable for predicting downstream effects of COVID-19 worry and risk perceptions; trait anxiety was included in the current study primarily to contribute to the literature regarding whether math anxiety is distinct from other forms of anxiety (e.g., Ashcraft, 2019; Ashcraft & Ridley, 2005; Hembree, 1990; Ma, 1999; Namkung et al., 2019).

Experience-based cues: worked-example training and problem-solving accuracy

In addition to the information-based cues described above, people may also rely on experience-based cues – cues based on the actual experience of completing the task at hand – when making monitoring judgments. Thus, we also tested whether experience-based cues (problem-solving accuracy and the worked-example content of the educational intervention) related to adults' item-level monitoring judgments.

Worked-example training

To address participants' tendency towards whole number bias (Thompson et al., [in press](#)), Thompson et al. (2021) randomly assigned participants to an educational intervention or control condition. In the intervention condition, participants saw a worked example² demonstrating, step-by-step, how to solve the target problems. This training was effective: post-training problem solving was more accurate in the experimental than control condition. Given that the intervention improved problem-solving accuracy, and problem-solving accuracy often influences monitoring judgments, we anticipated that the intervention would *also* result in higher monitoring judgments compared to the control condition (see Fig. 1). We examined whether the worked example influenced participants' monitoring judgments while controlling for their problem-solving accuracy. In this way, we tested whether the *experience* of training, regardless of its effect on problem-solving accuracy, influences monitoring judgments. Without feedback or practice solving problems, instruction may artificially increase the magnitude of monitoring judgments in problem-solving. That is, it might lead to misalignment between confidence and performance. For instance, Baars et al. (2017) found that solving practice problems *after* a worked example led to greater monitoring accuracy compared to when worked examples were not followed by practice problems. However, evidence is mounting which indicates that even when performance accuracy on a math task is improved with training, monitoring judgments may not improve in tandem (Fitzsimmons & Thompson, 2021; Fitzsimmons et al., 2021).

Problem-solving accuracy

Problem-solving accuracy (i.e., whether a participant answered a problem correctly or not) often has a strong influence on monitoring judgments (e.g., Händel et al., 2020; Wall et al., 2016). As discussed, confidence tends to be greater when people are accurate as compared to inaccurate on the task at hand. However, in the current study, participants solved problems in which people often demonstrate misconceptions. One such misconception is whole number bias, in which individuals misapply whole-number knowledge when reasoning about ratios, such as fractions (Ni & Zhou, 2005). For example, someone may say that COVID is more dangerous than the flu because 22,000 people died from the flu and only 9,318 had died from COVID (as of March 2020 when data were collected). Even though 22,000 is greater than 9,318, when one takes the ratio of the numerator to denominator into consideration, the case-fatality rate of COVID-19 ($9,318/227,743 = 0.041 = 4.1\%$) was

² The intervention also included other components, such as drawing an analogy from a less- to more-familiar context. Details about the training can be found in the original study.

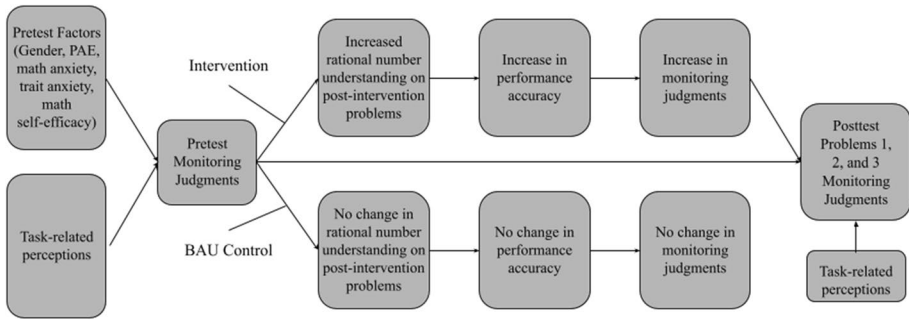


Fig. 1 Proposed Theory of Change

higher than that of the flu ($22,000/36,000,000 = 0.00061 = 0.06\%$). Given that people often focus on just the number of deaths and not the number of deaths relative to the number of infections, people may believe that the flu is more fatal than COVID-19 with high confidence. That is, they may be *both* ‘unskilled and unaware’ (Dunning et al., 2003; Kruger & Dunning, 1999; Pennycook et al., 2017). Problem accuracy was not a direct cue in the current study because participants did not receive feedback on their accuracy; however, it is possible that participants’ accuracy affected their task-related metacognitive experiences, such as feelings of ease and difficulty. Thus, we evaluated whether those participants who answered problems accurately also reported higher monitoring judgments than participants who answered inaccurately.

Other individual differences included in statistical models

In addition to the information- and experience-based cues described above, we also included several relevant individual difference variables in our analyses. Specifically, we wanted to account for individual differences in objective math skills to control for pretest math ability, explore how individual differences in math skills are related to monitoring judgments, and account for potential gender differences in monitoring judgments.

Objective math skills: Number-line estimation and Berlin numeracy We focused on measures from both the math cognition (i.e., number-line estimation performance) and health decision-making literature (i.e., Berlin numeracy) because of the interdisciplinary nature of the research questions (Thompson et al., [in press](#)).

The number-line estimation task taps numerical-magnitude knowledge (e.g., Siegler, 2016; Siegler & Booth, 2004; Siegler et al., 2011), which is a critical component of correctly solving the problems in the current study. Magnitude understanding is at the core of math achievement (Schneider et al., 2018; Siegler, 2016), and more precise estimates are predictive of greater math skills (e.g., Fazio et al., 2014; Siegler & Thompson, 2014; Siegler et al., 2011). Thus, number-line estimation precision served as a proxy for objective math ability in the current study. We also included a measure of objective numeracy (as measured by the 1-item Berlin numeracy task in the current study; Cokely et al., 2012), commonly used in the decision-making literature. This measure assesses people’s ability to reason with and compute numerical information. Thus, our two measures of objective math

skills accounted for numerical-magnitude estimation and calculation skills, both of which are critical to complete the target problems.

Gender It was necessary to include gender in our statistical models as a covariate because gender was not equally distributed across experimental conditions in Thompson et al. (2021). However, we were also interested in investigating the effect of gender on monitoring judgments in the domain of COVID-19-related math problem solving given that previous studies have reported gender differences in metacognitive judgments in non-health math contexts (Devine et al., 2012; Dowker et al., 2016; Else-Quest et al., 2010; Hembree, 1990; Rivers et al., 2020; Spencer et al., 2016). Specifically, women and girls tend to judge their performance with less confidence than men and boys, even when accounting for their objective performance on such tasks (Morony et al., 2013; Rivers et al., 2020). Given these gender differences in math and monitoring judgments, combined with findings that women report being more math anxious than men (e.g., Devine et al., 2012), and have more negative math attitudes than men (Sidney et al., 2021), we expected that women would make lower monitoring judgments than men in their problem solving, even when accounting for performance accuracy on the item.

Current study

The current study expands the literature in several important and novel ways. First, the analyses we conducted, which assessed whether math self-efficacy was associated with monitoring judgments, was a novel approach to unite distinct literatures in health decision making and metacognition. Second, Rivers et al. (2020) found that men were more precise and confident in their number-line estimates than were women; the current study tests whether that finding extends to the applied domain of health decision-making, specifically with problems pertaining to COVID-19. Third, it was unclear to our team how monitoring judgments would be affected by the worked-example educational intervention. Participants who engaged with the worked example, which was designed to improve relational understanding of rational numbers, might increase their monitoring judgments if the training helped them to gain confidence for the task at hand. However, it is also possible that the worked example highlighted for participants just how incorrect they were when trying to solve the problems, thus decreasing the magnitude of their subsequent monitoring judgments. The current study tracks how monitoring judgments evolved after participants engaged in a worked-example intervention, thus extending the literature on the relations between cues used to make monitoring judgments and interventions in an applied context.

Research questions and hypotheses

Our pre-registered analyses are broken up into directional hypotheses (H1 and H2) and exploratory, non-directional research questions (RQ1 and RQ2). We:

(H1) Predicted that, because the intervention led to higher performance accuracy, monitoring judgments would also be higher post intervention in the intervention compared to the control group.

(H2) Predicted that math anxiety would explain more variance in posttest monitoring judgments than trait anxiety, controlling for gender, pretest monitoring judgment, and experimental condition. Because participants completed a variety of math tasks in the current study, we expected a domain-specific construct, math anxiety, to be more strongly related to monitoring judgments than a domain-general measure of trait anxiety.

(RQ1) Non-directionally explored which individual differences were associated with monitoring judgments. We pre-registered plans to explore several individual differences (e.g., gender, math anxiety, pretest monitoring judgment, experimental condition), and we subsequently added math self-efficacy (measured with a selection of items from the SNS) to our analytic plan for the reasons indicated above.

(RQ2) Non-directionally explored whether participants who correctly answered the problems also reported higher monitoring judgments than participants who answered incorrectly, regardless of experimental condition.

Method

The ³dataset we used for the present study involved several measures related to general cognitive abilities, affect, math cognition, perception of COVID-19 susceptibility and severity, health literacy, and more. In this paper we only discuss the measures relevant to the current hypotheses (for a full analysis of the measures used in this study and the full study flow, see Thompson et al., 2021).

Participants

The parent study, from which these data were collected, was approved by the Kent State University IRB; all participants provided online consent for their participation, and their participation was voluntary. Data were collected from March 24 to April 9, 2020 through Qualtrics panels. As indicated in the pre-registration of the original project (<https://osf.io/9hc7d>), the authors planned to sample 1,200 people in the baseline survey to get 10 days of daily diaries from at least 625 people. The secondary data analyses reported in this paper were drawn from the baseline survey only, as there were no math-related or monitoring questions answered by participants in the daily diary portion of the study.

After exclusions, the final sample for our secondary data analyses included 1,177 participants. We excluded 120 participants for having incomplete data so that we could compare included participants on all analyses. Seventy-five percent of participants self-identified as White (10.33% identified as Black or African American, 3.93% identified as Hispanic or Latino, 3.86% identified as Asian, 0.62% identified as American Indian or Alaska Native, 0.23% identified as Native Hawaiian or Pacific Islander, and the remaining 6.32% identified as multiple races or ethnicities, “other,” or did not report), 46% identified as male, 41% reported being employed for wages, and 70% reported having between some college experience and a graduate degree. The average reported age of participants was 46.9 years ($SD = 17.34$ years; range: 18–85 years). As noted in the parent study, there were some differences between those participants who were included and excluded from analyses: those

³ We pre-registered that we would *explore* whether problem accuracy affected judgments at pretest and posttest. Thus, this is a non-directional research question.

excluded were younger and more likely to identify as White or female or to be students or self-employed. They were also less likely to be retired or employed for wages, report lower income, have taken fewer math courses, and be incorrect on an objective numeracy (Cokely et al., 2012) and baseline problem-solving question. See the original paper (Thompson et al., 2021) and the original project's pre-registration (<https://osf.io/9hc7d>) for full demographic information and data cleaning procedures.

Experimental design and procedure

Participants completed a pretest problem in a generic Disease A vs. Disease B context (see Fig. 2). Then, participants were randomly assigned to one of two intervention conditions. In the educational intervention condition, participants completed a step-by-step worked example of the correct procedures to complete the problems. The correct procedures involve accurately identifying the problem as a comparison of two ratios, and then correctly transforming the ratios to make them manageable to compare. Participants in the educational

Pretest Problem			Posttest Problem 1			Posttest Problem 2			Posttest Problem 3
	Total # deaths	Total # infected		Total # deaths	Total # infected		Total # deaths	Total # infected	<p>According to the WHO Situation Report, 81,174 people have been infected by COVID-19 in China, whereas 35,713 people have been infected by COVID-19 in Italy.</p> <p>According to the UN, China has a total population of 1,439,323,774, and Italy has a total population of 60,461,828.</p> <p>Does China have a higher, lower, or the same infection rate as Italy?</p> <p>Please describe in as much detail as possible how you made your decision about the number of people who have been infected.</p> <p>How confident are you in your decision regarding the infection rates? 0% = I am not confident at all, to 100% = I am totally confident?</p>
A	2,125	55,924	Flu	22,000	36,000,000	Jan. 30	170	7,818	
B	16,777	1,677,720	COVID-19	9,318	227,743	Feb. 6	565	28,276	
<p>Which disease is more fatal: Disease A, Disease B, or are they equally fatal?</p> <p>Please describe in as much detail as possible how you made your decision about which disease is more fatal.</p> <p>How confident are you in your decision about which disease is the most fatal from 0% = I am not confident at all, to 100% = I am totally confident?</p>			<p>Which disease is more fatal: the flu, COVID-19, or are they equally fatal?</p> <p>Please describe in as much detail as possible how you made your decision about which disease is more fatal.</p> <p>How confident are you in your decision about which disease is the most fatal from 0% = I am not confident at all, to 100% = I am totally confident?</p>			<p>Did the COVID-19 fatality rate: increase, decrease, or stay the same from January 30th to February 6th?</p> <p>Please describe in as much detail as possible how you made your decision about the possible change in fatality rates.</p> <p>How confident are you in your decision regarding the fatality rates? 0% = I am not confident at all, to 100% = I am totally confident?</p>			

Fig. 2 Overview of the four problems. *Note:* All problems were forced choice. We have bolded the correct responses here for readers. Participants made their monitoring judgments on a slider with endpoints of 0% = I am not confident at all to 100% = I am totally confident

intervention condition received the worked example designed to eliminate common mathematical errors; the business-as-usual condition saw relevant statistics – number of deaths and number of infected individuals – but were not shown how to calculate case-fatality rates or to consider the relation between these numbers.

After the intervention, all participants completed three more problems (see Fig. 2) and reported the strategy that they used to solve the problem. Strategy reports are useful because they reveal unique, convergent evidence into how participants solved the problems (see Alibali & Sidney, 2015; Sidney et al., 2018). Each problem only had three possible answers, so random responding would result in an average of 33% accuracy. Because of the high possibility for getting a problem right for the wrong reason—or getting a problem wrong for the right reason—strategy reports provide a more complete picture of the participants' cognitive processes (Fazio et al., 2017; Fitzsimmons et al., 2020b; Reder, 1987; Sidney et al., 2018, 2021; Siegler & Thompson, 2014; Siegler et al., 2011).

Immediately after each problem, participants reported their strategy use, and then provided a monitoring judgment by rating their confidence using a slider between 0 and 100% to answer the question “How confident are you in your decision regarding the fatality rates? 0% = I am not confident at all, to 100% = I am totally confident?” Each of these monitoring judgments reflected participants' confidence in the accuracy of their answer to the problem they had just completed.

Participants completed the pretest problem, the math anxiety scale, and the number line estimation tasks prior to the intervention. After the intervention, participants completed the remaining problems and the trait anxiety scale. There were a number of other measures included in the parent project (e.g., risk perceptions related to COVID-19) that are not central to the current hypotheses (see Appendix 3 in Thompson et al., 2021).

Materials

Measures relevant to the current analyses are described in the order they were completed by participants.

Pretest problem

As shown in Fig. 2, participants compared health statistics for two hypothetical diseases and chose which of the two was more fatal. Disease A (analogous to flu statistics at the time) included a bigger numerator and a bigger denominator as compared to Disease B (analogous to COVID-19 statistics at the time), even though the magnitude of the risk was larger for Disease B.

Pretest subjective and objective tasks

Participants rated their math attitudes and math anxiety prior to completing measures of their objective math skills because prior research (e.g., Sidney et al., 2021) suggested that when participants completed difficult fraction tasks first, it resulted in more negative attitudes about math in general and attitudes about fractions, whole numbers, and percentages, specifically. The order of the measures within these two blocks was randomized for all participants.

Math anxiety Participants rated their overall math anxiety on the Single-Item Math Anxiety Scale (SIMA; Ashcraft, 2002; Núñez-Peña et al., 2014) and their math anxiety about specific types of numbers (e.g., fractions) across four items on Likert-like scales ranging from 1 = “Not anxious” to 10 = “Very anxious.” We calculated a math anxiety index by averaging scores across the five items.

Trait anxiety We included the 20-item validated state-trait anxiety scale from Spielberger et al. (1970). A sample item includes: “I worry too much over something that really doesn’t matter.” Participants made their ratings on four-point Likert-like scales that ranged from “Almost never” to “Almost always.” We aggregated the ratings by taking a sum across all 20 items (leaving a possible range of 20–80, with higher scores representing higher reported anxiety).

Math attitudes Participants completed the 20-item Math Attitudes Questionnaire (MAQ; adapted from Sidney et al., 2021) regarding their attitudes about math in general, as well as specific math attitudes toward whole numbers, fractions, and percentages. We calculated a math-attitudes index by averaging scores across the 20 items. We planned to include the MAQ in our regression models, but this variable was strongly related to the SNS scale. Thus, including both predictors in our models would lead to issues of multicollinearity. The SNS subscale accounted for more variance in metacognitive monitoring judgments, thus only the SNS was retained in the reported model. Additionally, given that Peters et al. (2019) considered the first four items of the SNS a measure of “confidence,” we explored whether it accounted for unique variance in item-level monitoring judgments. We also included the MAQ in the correlation matrix to indicate how the task correlated with other tasks in the current study.

Math self-efficacy Peters et al. (2019) conceptualized “confidence” in math as the first four items on the Subjective Numeracy Scale (SNS), which involves eight questions pertaining to people’s subjective preferences for and comfort with math (Fagerlin et al., 2007). In the introduction, we argued that this conceptualization of confidence is not consistent with the way this construct is handled in the domain of metacognition. Rather, we argued that the first four items of the SNS should be considered a proxy for participants’ math self-efficacy with rational numbers because the questions specifically address perceived ability with fractions and percentages: (a) How good are you at working with fractions, (b) How good are you at working with percentages, (c), How good are you at calculating a 15% tip, and (d), How good are you at figuring out how much a shirt will cost if it is 25% off? Similar to the self-perceived ability subscale of the MAQ, we also adopted this subscale of the SNS to examine individual differences in participants’ pretest monitoring judgments.

We did not preregister plans to include the SNS in our statistical models. However, as our project developed, we realized that this subscale may more accurately reflect math self-efficacy and preferences pertaining *specifically* to rational numbers. Thus, we deviated from our pre-registered analysis plan to include this additional, theoretically valuable variable in our models (see results section for more details).

Magnitude knowledge We measured number-magnitude knowledge with number-line estimation tasks for fractions, whole number frequencies, and percentages. Number-line estimation is an ideal proxy for overall mathematics skill because it is quick (Fazio et al., 2017) and easy to administer. Performance on each trial was measured as percentage of

absolute error (i.e., PAE; Siegler & Booth, 2004). PAE is calculated by taking the absolute value of the difference between the person’s estimate and the to-be-estimated number and dividing by the scale of the number line ([person’s estimate - to-be-estimated number] / scale of estimates). We averaged PAE across trials within each range and calculated an average across ranges such that higher PAE scores indicated greater error of estimation (worse performance).

Objective numeracy We operationalized objective numeracy—or the ability to “run the numbers” correctly (Lipkus et al., 2001; Peters, 2020; Peters et al., 2019; Schwartz et al., 1997)—by adopting Cokely et al.’s (2012) 1-item, free-response version of the Berlin Numeracy Test. This measure asks, “Imagine we are throwing a 5-sided die 50 times. On average, out of these 50 throws how many times would this 5-sided die show an odd number (1, 3, or 5)?” Participants’ answers were coded as either correct (i.e., 30 times) or incorrect. Including this measure, commonly used in the health decision-making literature, in our models allowed us to account for possible differences in general numerical ability.

Educational intervention vs. business-as-usual control

Participants were randomly assigned to one of two experimental conditions: The educational intervention to combat whole number bias errors (problem-solving errors where people attend only to the individual components of the ratio and ignore the relational nature; Ni & Zhou, 2005) or the business-as-usual control condition (Fig. 3). The intervention included an analogy to a familiar context—an apple orchard—designed to help participants understand the problems conceptually. That is, the intervention included a worked example in which a familiar context (apples rotting at different rates in two apple orchards) illustrated the procedural steps that could be followed to calculate a case-fatality rate, emphasizing the use of number lines and relational reasoning. Then, participants were asked to draw an analogy from the apple orchard worked example to a worked example in a health context which showed how to calculate and compare the case-fatality rates for COVID-19 and the flu. See Thompson et al. (2021) for details. For the purposes of the secondary data analyses in the current study, we

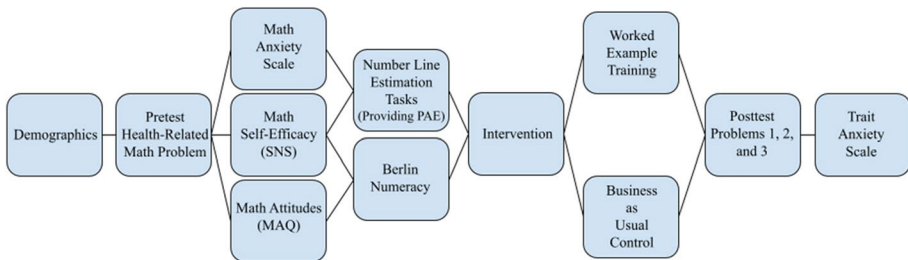


Fig. 3 Order in Which Participants Completed Tasks for Both Experimental and Control Groups. *Note.* Order of measures was randomized for the block including math anxiety, math self-efficacy, and math attitudes. Order was also randomized for the objective math measures including number line estimation and Berlin Numeracy

ensured that pretest item-level monitoring judgments did not significantly differ across the experimental versus control groups ($p = 0.592$).

Results

Data analytic plan

We preregistered hypotheses on OSF (https://osf.io/vxm8d/?view_only=ca422f16682947d98a801e42f4d94b90). First, we computed zero-order correlations to assess the relations between pretest and post-intervention item-level metacognitive judgments and other relevant variables (e.g., math anxiety and self-perceived math ability). We ran linear regressions to identify the factors associated with item-level metacognitive judgments (RQ1), whether performance accuracy, regardless of condition, was associated with item-level metacognitive judgments (RQ2), and whether math anxiety was more strongly associated with item-level metacognitive judgments than trait anxiety (H2). Finally, to test hypothesis 1, in which we predicted higher item-level metacognitive judgments in the intervention vs. control group, we conducted ANCOVAs.

Correlations among study variables

As ⁴anticipated, many of the variables in this study were correlated because they tapped similar underlying constructs (see Table 1 for descriptive statistics and correlations). For example, number-line estimation PAE and objective numeracy were strongly correlated (i.e., objective math skills), as were subjective numeracy, math anxiety, and math attitudes (i.e., affective factors), item-level metacognitive judgments before and after the intervention (i.e., awareness of performance on the problems), and math anxiety and trait anxiety (i.e., anxiety more broadly). Measures of magnitude understanding (PAE) and affective reactions to math (math attitudes and math anxiety) were moderately related (absolute value of r 's ≥ 0.32 , p 's < 0.001). As seen in Table 1, trait anxiety and math anxiety were also moderately correlated ($r = 0.31$, $p < 0.001$).

Monitoring judgments from pretest and posttest were all strongly related to one another (r 's = .48 to .70, p 's < 0.001). That is, those who rated their item-level metacognition high at pretest also rated their item-level metacognition high at posttest across all items. To foreshadow our subsequent results, self-perceived math ability (operationalized as the first four items on the SNS) was moderately associated with monitoring judgments across problems, even when accounting for our other predictors.

Primary models

Variables associated with monitoring judgments

To explore the factors that predicted item-level metacognitive judgments (RQ1), we conducted linear regressions in which we predicted item-level judgments separately for each

⁴ For interested readers, we report our full pre-registered models in Appendix A, Appendix B, Appendix C, and a supplemental file.

Table 1 Descriptive Statistics and Correlations for Study Variables

	Mean (SD)	1	2	3	4	5	6	7	8	9	10	11
1. Pretest Monitoring Judgment	83.99 (21.49)	–										
2. Posttest Problem 1 Monitoring Judgment	82.41 (23.76)	.49*	–									
3. Posttest Problem 2 Monitoring Judgment	80.91 (23.71)	.51*	.68*	–								
4. Posttest Problem 3 Monitoring Judgment	78.95 (24.99)	.48*	.66*	.70*	–							
5. Percent Absolute Error	0.16 (0.09)	-.13*	-.18*	-.13*	-.13*	–						
6. Math Anxiety	4.54 (2.51)	-.16*	-.19*	-.18*	-.19*	.37*	–					
7. Trait Anxiety	43.00 (11.74)	-.10*	-.11*	-.12*	-.16*	.06*	.31*	–				
8. Subjective Numeracy Scale (SNS)	33.84 (8.62)	.30*	.34*	.30*	.34*	-.41*	-.38*	-.17*	–			
9. Math Self-Efficacy (first four SNS items)	16.72 (5.39)	.28*	.31*	.28*	.33*	-.38*	-.36*	-.17*	.92*	–		
10. Math Attitudes Questionnaire	4.11 (1.04)	.29*	.30*	.28*	.35*	-.32*	-.34*	-.15*	.79*	.81*	–	
11. MAQ Self-Perceived Ability Subscale	30.21 (7.89)	.29*	.31*	.27*	.33*	-.38*	-.41*	-.17*	.79*	.82*	0.89*	–

Correlations between SNS, math self-efficacy, MAQ, and MAQ self-perceived math ability subscale were all ≥ 0.79 , suggesting redundancy. As mentioned above, due to this high level of overlap, only the SNS subscale was used in the reported regression model for testing which factors were significantly associated with item-level metacognitive judgments (RQ1).
* p -value < 0.001 .

of the four problems. We included condition (except at pretest), pretest monitoring judgments (except at pretest), participants' pretest problem-solving accuracy (except at pretest), gender, PAE, problem-solving accuracy for the current problem, math self-efficacy, objective numeracy, math anxiety, and trait anxiety in our models. As indicated in the note for Table 1, we did not include the full MAQ or the MAQ self-perceived math ability subscale in our models because math self-efficacy (i.e., the SNS self-perceived math ability subscale), the full MAQ, and the MAQ self-perceived math ability subscale were highly correlated (all r 's > 0.78). We were particularly interested in how the SNS subscale was associated with item-level metacognitive judgments; thus, we retained this variable in our models instead of MAQ.

As seen in Table 2, when accounting for individual differences in pretest monitoring judgments, the only variable which was a significant predictor of monitoring judgments on all four problems was the 4-item SNS self-perceived math ability subscale (β 's ≥ 0.12 , r 's ≥ 4.10 , p 's < 0.001). Problem-solving on the current problem (except at pretest) was significantly associated with monitoring judgments, as was gender. See Appendices A and B for the full models.

Differences in monitoring judgments by experimental condition

We hypothesized that post-intervention monitoring judgments would be higher in the intervention group as compared to the control group, controlling only for pretest monitoring judgments and gender (H1). Our hypothesis stemmed from the fact that participants were

Table 2 Linear Regression Standardized Beta Coefficients for Predictors of Monitoring Judgments

	Pretest Problem	Posttest Problem 1	Posttest Problem 2	Posttest Problem 3
Condition		0.04	0.04	0.05*
Pretest Monitoring Judgments		0.43***	0.45***	0.40***
Pretest Accuracy		0.01	-0.08**	-0.03
Gender	0.11***	0.05*	0.01	0.05*
PAE	0.01	-0.03	<0.01	0.07*
Accuracy (on the current problem)	-0.01	0.07*	0.06*	0.14***
Math Self-Efficacy	0.21***	0.12***	0.13***	0.17***
Objective Numeracy Scale	0.07*	0.03	0.03	0.06*
Math Anxiety	-0.05	-0.03	-0.04	-0.02
Trait Anxiety	-0.04	-0.03	-0.04	-0.07**
Model Fit				
R^2	0.10	0.29	0.29	0.31
Adjusted R^2	0.10	0.28	0.29	0.30
F	18.60***	47.38***	48.60***	51.61***

Condition was dummy-coded as control vs. intervention, gender was dummy-coded as not-male vs. male, accuracy was dummy-coded as incorrect vs. correct, and objective numeracy was dummy-coded as incorrect vs. correct. Math anxiety was associated with monitoring judgments on the pretest and post-intervention problems when math self-efficacy was not in the model.

* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

more accurate at solving problems (both posttest problem 1 and 2) in the intervention compared to the control group (Thompson et al., 2021), thus we anticipated increased monitoring judgments as well.

To test this hypothesis, we conducted ANCOVAs comparing the effects of experimental condition on monitoring judgments (dummy coded as intervention vs. control) while controlling for gender (dummy coded as male vs. not male), and pretest monitoring judgments. Consistent with hypothesis 2, participants' monitoring judgments were higher in the intervention group than the control group on all post-intervention problems (see Table 3). On posttest problem 1, the intervention group reported higher monitoring judgments ($M=83.89, SE=0.84$) than the control group ($M=80.88, SE=0.85$), $F(1, 1173)=6.30, p=0.012$, partial $\eta^2=0.01$. Similarly, on posttest problem 2, the intervention group ($M=82.11, SE=0.83$) reported higher monitoring judgments than the control group ($M=79.67, SE=0.85$), $F(1, 1173)=4.19, p=0.041$, partial $\eta^2=0.004$. Finally, on posttest problem 3, the intervention group ($M=80.40, SE=0.89$) again reported higher monitoring judgments than the control group ($M=77.45, SE=0.91$), $F(1, 1173)=5.37, p=0.021$, partial $\eta^2=0.01$.

Problem accuracy and monitoring judgments

To test our second research question (RQ2), we ran separate hierarchical linear regressions to provide convergent evidence that people who answered the problems correctly also reported higher monitoring judgments, regardless of experimental condition. Participants who were accurate (dummy coded as correct or incorrect) reported higher monitoring judgments on the pretest problem, $\beta=0.08, t=2.63, p=0.009$, but not when gender was added to the model, $\beta=0.05, t=1.71, p=0.087$. However, males reported higher monitoring judgments than females, $\beta=0.19, t=6.43, p<0.001$.

Then, we included accuracy on each respective problem in block 1 and gender and pretest monitoring judgments in block 2 to predict monitoring judgments on each of the post-intervention problems in separate regression models. All six models exhibited good model fit (block 1: all F 's $> 19.90, p$'s $< 0.001, R^2$'s > 0.02 ; block 2: all F 's $> 143.10, p$'s $< 0.001, R^2$'s > 0.26), and most of the individual predictors were related to participants' monitoring judgments; the only exception was that gender did not improve model fit for posttest problem 2, $\beta=0.04, t=1.42, p=0.156$.

Table 3 Observed means, Standard deviations, and Analysis of Covariance for Hypothesis 1 (Comparing Differences in Monitoring Judgments by Experimental Condition)

	Intervention		BAU Control		$F(1, 1173)$	Partial η^2
	M	SD	M	SD		
Pretest Problem	84.32	21.22	83.65	21.78	0.93	<0.01
Posttest Problem 1	83.90	23.38	80.86	24.07	6.30*	0.01
Posttest Problem 2	82.21	22.51	79.56	24.83	4.19*	<0.01
Posttest Problem 3	80.41	23.79	77.44	26.11	5.37*	0.01

Pretest is included in this table to illustrate that there were no significant mean differences between the intervention group and the control group prior to the intervention. The descriptive statistics reported in this table are means and standard deviations; the descriptive statistics reported in the text above are estimated marginal means and standard errors of the estimates.

*= $p < 0.05$.

To summarize, confidence (i.e., the magnitude of monitoring judgments) was higher for those who were accurate on problems relative to those who were inaccurate, suggesting some level of metacognitive awareness at the group level. Means, standard deviations, and independent sample *t*-tests are shown in Table 4 illustrating mean differences in monitoring judgments by problem accuracy. Note that the difference between accurate and inaccurate responders on monitoring judgments was smallest at pretest, $t(1175)=2.81$, $p<0.01$, $d=0.16$, compared to posttest d 's between 0.30 and 0.42 (see Table 4). The larger differences in monitoring judgments between accurate and inaccurate responders appears to be driven by lower confidence in inaccurate responses rather than higher confidence in accurate responses. Given that participants did not receive feedback on their performance, it seems likely that characteristics of the problems, elicited experience-based cues such as perceptions of difficulty, influenced their monitoring judgments. For example, posttest problem three included exceptionally large numbers (e.g., 14 billion), did *not* include a contingency table, and had the smallest difference in magnitude of infection rate between countries (see Fig. 2 for problem features). These features of the problem may have served as cues that influenced participants' monitoring judgments, even though Thompson et al. (2021) reported that accuracy on posttest Problem 3 was higher than expected, potentially because of news sources reporting on Italy's and China's infection rates. In fact, monitoring judgments were lowest on this final problem, regardless of accuracy (Table 4).

Math anxiety, trait anxiety, and monitoring judgments

Finally, we hypothesized that math anxiety would explain more variance in posttest monitoring judgments than trait anxiety, controlling for gender, pretest monitoring judgments, and experimental condition (H2). We based this prediction on prior research that suggested math anxiety might play a unique role above and beyond general trait anxiety when examining negative affective reactions in a specific math domain (Ashcraft & Ridley, 2005; Barroso et al., 2020; Dreger & Aiken, 1957; Gough, 1954; Hembree, 1990; Zhang et al., 2019).

We regressed posttest monitoring judgments (for each individual problem) onto gender, experimental condition, pretest monitoring judgments, and trait anxiety in the first step. We evaluated the variance explained by math and trait anxiety above and beyond all other

Table 4 Observed means, Standard Deviations, and *t*-tests for Research Question 2 (Comparing Monitoring Judgments for Accurate vs. Inaccurate Problem-Solvers)

	Accurate		Inaccurate		<i>t</i> (1175)	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Pretest Problem	85.94	20.06	82.45	22.46	2.81*	0.16
Posttest Problem 1	84.35	22.65	77.31	25.80	4.33**	0.30
Posttest Problem 2	87.54	20.73	78.56	24.25	6.23**	0.38
Posttest Problem 3	83.00	22.33	72.79	27.46	6.71**	0.42

The degrees of freedom listed in this table is 1175; however, Levene's Test for Equality of Variances was violated for all problems (inaccurate responders were more variant in their monitoring judgments than accurate responders). Thus, different degrees of freedom were used for the *t*-tests at pretest (1158.94), posttest problem 1 (525.20), posttest problem 2 (625.04), and posttest problem 3 (853.54).

* = $p < 0.01$, ** = $p < 0.001$.

Table 5 Hierarchical Linear Regression Standardized Beta Coefficients for Hypothesis 2 (Unique Variance in Monitoring Judgments Accounted for by Math Anxiety)

	Pretest Problem		Posttest Problem 1		Posttest Problem 2		Posttest Problem 3	
	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2
Gender	0.18***	0.031	0.10***	0.009	0.04	0.002	0.10***	0.009
Condition			0.06*	0.004	0.05*	0.002	0.06*	0.003
Pretest			0.45***	0.194	0.48***	0.218	0.44***	0.182
Monitoring Judgments								
Trait Anxiety	-0.05	0.002	-0.03	0.001	-0.05	0.002	-0.09**	0.007
Math Anxiety	-0.13***	0.014	-0.09**	0.008	-0.08**	0.006	-0.08**	0.005
	Pretest Problem		Posttest Problem 1		Posttest Problem 2		Posttest Problem 3	
R ²	0.06		0.27		0.28		0.26	
Adjusted R ²	0.06		0.26		0.27		0.26	
F	24.97		85.26		88.78		83.45	

ΔR^2 represents the change in R^2 when the variable is added to the model with all other predictors.

* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

predictors in the model for each of the four problems. The change in R^2 values in Table 5 reflect the additional variance explained by each of the four models when an individual predictor is added to the model with all other variables. Across all three post-intervention problems, adding math anxiety to the model increased the model fit (all ΔR^2 values = 0.01; all ΔF p -values < 0.01).

As shown in Table 5, math anxiety remained a significant predictor of monitoring judgments (β 's between -0.08 and 0.10; all p 's < 0.01) above and beyond trait anxiety. However, trait anxiety was only a significant predictor above and beyond math anxiety for the third posttest problem, $\Delta F(1, 1191) = 10.64$, $\Delta R^2 = 0.007$, $p = 0.001$. It is an open question as to why trait anxiety was a stronger predictor for post intervention problem 3 than the other problems. One possible explanation is that there was something unique about problem 3. The problems in this study were complicated and involved real-world scenarios. For example, problem 3 compared two countries that were discussed extensively in the news at the time, and thus accuracy on the problem may have been high even if participants did not engage deeply with the mathematics involved in the problem. See Table 5 for the standardized beta coefficients and Appendix 3 for the full models.

Discussion

In this secondary data analysis, we investigated the type of information-based (i.e., math self-efficacy, math anxiety, trait anxiety) and experience-based cues (i.e., worked-example intervention) that related to adults' monitoring judgments about their performance on problems (Koriat, 1997; Koriat & Levy-Sadot, 1999).

First, several individual differences accounted for variance in metacognitive monitoring judgments: Males reported higher monitoring judgments than females, accurate responders reported higher monitoring judgments than inaccurate responders,

and participants who made more precise estimates of numerical magnitudes reported higher monitoring judgments than participants who made less precise estimates of numerical magnitudes. Further exploratory analyses (see Supplemental File) revealed that math anxiety mediated the effect of gender on pretest monitoring judgments. Second, consistent with our first hypothesis, participants in the intervention condition reported higher monitoring judgments than participants in the control condition. Third, individual differences accounted for variance in item-level monitoring judgments: (a) Participants with higher math self-efficacy reported higher item-level monitoring judgments than participants with lower math self-efficacy, controlling for a number of other predictors as described in the introduction, (b) participants with lower math anxiety reported higher monitoring judgments than participants with higher math anxiety, and (c) participants with more positive math attitudes reported higher monitoring judgments than participants with less positive math attitudes (see the correlation matrix in Table 1). Finally, consistent with prior research (e.g., Ashcraft, 2019; Ashcraft & Ridley, 2005; Hembree, 1990), participants' math anxiety appeared to be separable from participants' trait anxiety.

Significance of the present study

Our data indicate that when people solve math problems, they bring a variety of individual differences to the experimental context, which can contribute to differences in their monitoring judgments, regardless of their accuracy on the problem (Koriat, 2011). For example, on two of the three post-intervention problems, participants' gender was associated with their monitoring judgments even when the other predictors were in the model (see Table 2). Furthermore, across all problems, ratings of math attitudes and math anxiety in our study were significantly associated with monitoring judgments (all correlations $r \geq .34$, all p -values < 0.001). More subjective cues, such as math attitudes and anxiety may serve as commonly used cues that inform monitoring judgments.

Furthermore, consistent with prior research, our data suggested that women report lower monitoring judgments than men when rating how sure they were that they had correctly answered math problems (Barroso et al., 2020; Devine et al., 2012; Hembree, 1990; Wigfield et al., 1991), and people with higher math anxiety reported lower monitoring judgments (Hembree, 1990; Morsanyi et al., 2014; Rolison et al., 2016). Women—who are on average less precise when estimating the magnitudes of numbers (Geary et al., 2020; Hutchison et al., 2019; Rivers et al., 2020; Thompson & Opfer, 2008) and rate themselves as less confident in their ability to estimate magnitudes (Rivers et al., 2020)—have higher math anxiety (Devine et al., 2012), lower numeracy (Weller et al., 2013), and more negative attitudes about rational numbers than men (Sidney et al., 2021). In the current study, women also provided lower monitoring judgments on their problem-solving performance than men. The reasons for these gender differences are up for debate (Halpern et al., 2007); however, gender differences in spatial abilities (Geary et al., 2020), which may stem from early gender differences in spatial experiences (e.g., spatial language and block play; Pruden et al., 2011), are one widely accepted predictor of gender differences in math.

Relevance for health care

Many people struggle to comprehend numerical health information (Peters et al., 2019; Waters et al., 2016). People may be likely to avoid numerical health information if they believe that they cannot interpret the information (Afifi & Weiner, 2004; Sweeny et al., 2010), if they have high math anxiety (Rolison et al., 2016), or if they have negative attitudes toward the numbers (Sidney et al., 2021). People's level of certainty in their interpretations of numerical health information, their comprehension of health statistics, and their own perceived disease susceptibility can influence their medical decision-making (Desender et al., 2018; Peters et al., 2019; Taber & Klein, 2016). If people are not confident in their ability to accurately interpret quantitative health information, they may turn to other types of information, such as affective factors and personal values, to make health decisions, which could have downstream implications for uptake of preventive health recommendations (e.g., social distancing, wearing a mask, getting vaccinated). Conversely, if people are inappropriately confident, while lacking adequate understanding of numeric health information, then they might avoid healthy behaviors, engage in unhealthy behaviors for loved ones and themselves, and be less likely to listen to opposing data. One real-world application of monitoring judgments is affecting when people offer opinions (Dunlosky & Metcalfe, 2009). Monitoring judgments paired with inaccurate understanding of numerical health information could contribute to the spread of misinformation. Note that COVID-19 was the domain in which we tested our hypotheses due to its international relevance; however, we do not believe that our findings are strictly limited to the context of COVID-19 and should be applicable to adults' metacognitive monitoring of any health statistics. It is also important to note that while the current study offers some data on the relation between monitoring judgments and several other factors, some of the reported effect sizes are small. Small effect sizes can often be indicative of real, important effects (Funder & Ozer, 2019), especially when the effects translate to large numbers of people, as with an international pandemic. Regardless of large or small effects, it is important to thoughtfully consider the appropriate reach of recommendations and implications (see Robinson et al., 2013; Robinson & Levine, 2019).

Future research might assess whether monitoring judgments also play a role in conveying incorrect health information to others. There is a vast amount of misinformation regarding COVID-19, exacerbated by easy access to social media (Frenkel et al., 2020; Pennycook et al., 2020; Russonello, 2020). It is possible that a person who confidently believes that COVID-19 is no worse than the flu (or similar minimizing beliefs) could be more likely (compared to a person who is not confident in this belief) to perpetuate this false belief by sharing the information with friends and family. More than two years after the WHO declared COVID-19 a pandemic, misinformation continues to spread, and accurate information about COVID-19 can be hard to find.

Relevance for education

There are important educational implications regarding the role of gender and math anxiety (Ashcraft & Ridley, 2005; Baloğlu, 2004; Barroso et al., 2020; Beilock et al., 2010; Devine et al., 2012; Dowker et al., 2016; Hembree, 1990; Ma, 1999; Tomasetto, 2019). In the current study, we explored how gender and math anxiety were related to monitoring

judgments in health-related math problem solving. With a broader lens, we also examined how math self-efficacy for rational numbers (i.e., first four items of the SNS) was associated with math anxiety and monitoring judgments. It remains an open question as to whether math self-efficacy is a domain-general trait which could affect participants' perceptions of their ability in a wide variety of contexts including simple math-related decisions, such as choosing an appropriate tip, to more complex math-related decisions, such as evaluating health care options.

Future research could assess whether other math interventions, which have been successful in academic settings, might also lead to high monitoring judgments in health and other domains. Our data suggest that primary targets for future interventions could be negative affective reactions to math, such as poor math attitudes and high math anxiety (cf. Jamieson et al., 2010; Park et al., 2014; Supekar et al., 2015). Decreasing these negative affective reactions may increase people's monitoring judgments and may be beneficial both in educational and health contexts.

Limitations and future directions

One limitation of the present study is inherent to any secondary data analyses: We were only able to investigate the measures included in the publicly-available data set. Future studies might include several additional measures, such as various measures of math self-efficacy, math self-concept, and math anxiety. For example, math anxiety as a construct has been measured in many ways (see Cipora et al., 2019 for a review). Some inventories focus on formal academic learning and testing contexts (Carey et al., 2017; Hopko et al., 2003; Jameson, 2013; Yáñez-Marquina et al., 2017), others focus on affective reactions to math such as worry (Bai, 2011; Harari et al., 2013; Wigfield & Meece, 1988), and still others focus on everyday math contexts (Hunt et al., 2011; Yáñez-Marquina and Villardón-Gallego, 2017). Math anxiety in different contexts was not the focus of the current study; future research could establish whether health-related statistics cause adults to feel more or less math anxious than when they encounter math in academic contexts. Similar to the way math perceptions (e.g., math self-efficacy, math self-concept, and math anxiety) often overlap and are presented in different ways, terminology related to *confidence* is discussed differently in different domains (Stankov et al., 2015). Notably, domains with substantial overlap (e.g., health decision making, math cognition, and metacognition) share terminology—confidence—yet operationalize and interpret findings differently (see Thompson et al., [in press](#) for a more in-depth discussion of this topic).

Another similar limitation, which could guide future research, is the high level of overlap between math anxiety, trait anxiety, math attitudes (especially the efficacy subscale), and math self-efficacy. Correlations among these variables were moderate to extremely high, suggesting that the constructs were somewhat overlapping. It is possible that the significant correlations between these measures is largely due to the nature of self-report surveys: Participants might not be able to differentiate between small differences in stimuli between math anxiety, math attitudes, and their perceptions of their own math ability (i.e., math self-efficacy). While self-report surveys are by far the most commonly used way to assess math anxiety (Cipora et al., 2019), other methodologies could be used in the future, such as physiological measures and administering state, as opposed to trait measures. Additionally regarding math anxiety, it is possible that

participants had reactive effects to completing math anxiety and math attitude items near the beginning of the experiment. Order effects should regularly be a consideration for any experiment that deals with many different measures. However, previous research on math attitudes suggested that completing math tasks first would induce *lower* overall math attitudes, thus, Thompson et al. (2021) did not counterbalance in the current study (see Sidney et al., 2021 for a discussion of order effects in this domain).

Finally, several other interesting open research questions remain. For example, are mathematical monitoring judgments in general strongly related to monitoring judgments across different rational number types (e.g., ratios, fractions, decimals, percentages)? Similar to attitudes about math (e.g., Sidney et al., 2021), are gender differences in **math anxiety** and monitoring judgments differentiated by the type of mathematics being done? Because some less diagnostic subjective cues, such as math self-efficacy, are stronger predictors of pretest monitoring judgments than more diagnostic, direct-access cues, such as problem-solving accuracy, it is a worthwhile endeavor for future research to investigate ways to de-bias people's tendency to focus on less diagnostic cues (Fitzsimmons et al., 2020b). Future research should investigate how aware people are of using specific cues as they make math-related health decisions, given that a common assumption of cue-utilization theory (e.g., Koriat & Adiv, 2016) is that cues shaping subjective confidence judgments (i.e., monitoring judgments) can operate below conscious awareness. Metacognitive experiences (Efklides, 2006) are complicated inferential processes that use a variety of cues to help guide decision-making and problem solving. Follow-up studies should investigate if people's beliefs about what drives their monitoring judgments align with the data presented in the current study.

Conclusions

The current study involved a secondary data analysis of Thompson et al.'s (2021) data. In the original project, a national panel of adults learned how to calculate and compare COVID-19 and flu case-fatality rates. In this secondary data analysis, we examined the factors that predicted people's monitoring judgments (i.e., confidence) pertaining to their performance on problems involving COVID-19. Monitoring judgments are only as useful as the cues from which they are derived (Ackerman & Thompson, 2017; Koriat, 1997). Thus, developing a more rigorous understanding of which cues inform math-related decision making in health contexts deserves future research. Our data suggested that people use both experience-based and information-based cues when making monitoring judgments when solving health-related math problems. We argue that both math self-efficacy and monitoring judgments have relevant implications for downstream effects in both health and education domains.

Appendix 1

Table 6 Full Linear Regression Coefficients for Research Question 1

Variable	Statistic	Pretest	Posttest Problem #1	Posttest Problem #2	Posttest Problem #3
Condition	<i>B</i> (SE)		1.76 (1.23)	1.80 (1.18)	2.44 (1.23)
	β		0.04	0.04	0.05
	<i>t</i>		1.43	1.53	1.99
	<i>p</i>		0.153	0.127	0.047
	95% CI		[-0.65, 4.17]	[-0.51, 4.11]	[0.03, 4.85]
Pretest Monitoring Judgment	<i>B</i> (SE)		0.48 (0.03)	0.50 (0.03)	0.47 (0.03)
	β		0.43	0.45	0.40
	<i>t</i>		16.51	17.36	15.64
	<i>p</i>		<0.001	<0.001	<0.001
	95% CI		[0.42, 0.53]	[0.44, 0.56]	[0.41, 0.53]
Pretest Accuracy	<i>B</i> (SE)		0.45 (1.32)	-3.79 (1.31)	-1.58 (1.39)
	β		0.01	-0.08	-0.03
	<i>t</i>		0.34	-2.90	-1.14
	<i>p</i>		0.732	0.004	0.254
	95% CI		[-2.14, 3.05]	[-6.36, -1.23]	[-4.30, 1.14]
Gender	<i>B</i> (SE)	4.85 (1.27)	2.56 (1.26)	0.66 (1.26)	2.72 (1.31)
	β	0.11	0.05	0.01	0.05
	<i>t</i>	3.82	2.02	0.53	2.08
	<i>p</i>	<0.001	0.043	0.599	0.038
	95% CI	[2.35, 7.34]	[0.08, 5.04]	[-1.80, 3.12]	[0.15, 5.30]
Percent Absolute Error (PAE)	<i>B</i> (SE)	1.73 (8.55)	-7.56 (8.47)	0.73 (8.48)	21.22 (8.87)
	β	0.01	-0.03	0.00	0.07
	<i>t</i>	0.20	-0.89	0.09	2.39
	<i>p</i>	0.840	0.372	0.931	0.017
	95% CI	[-15.05, 18.50]	[-24.17, 9.05]	[-15.91, 17.37]	[3.82, 38.61]

Table 6 (continued)

Variable	Statistic	Pretest	Posttest Problem #1	Posttest Problem #2	Posttest Problem #3
Accuracy on the Current Problem	<i>B</i> (SE)	-0.63 (1.31)	3.46 (1.46)	3.22 (1.48)	7.31 (1.37)
	β	-0.01	0.07	0.06	0.14
	<i>t</i>	-0.48	2.37	2.18	5.34
	<i>p</i>	0.634	0.018	0.030	<0.001
Math Self-Efficacy	95% CI	[-3.20, 1.95]	[0.59, 6.32]	[0.32, 6.12]	[4.62, 10.00]
	<i>B</i> (SE)	0.82 (0.13)	0.53 (0.13)	0.55 (0.13)	0.80 (0.14)
	β	0.21	0.12	0.13	0.17
	<i>t</i>	6.36	4.09	4.28	5.97
Objective Numeracy	<i>p</i>	<0.001	<0.001	<0.001	<0.001
	95% CI	[0.57, 1.08]	[0.28, 0.79]	[0.30, 0.81]	[0.54, 1.07]
	<i>B</i> (SE)	3.53 (1.52)	1.36 (1.50)	1.85 (1.50)	3.19 (1.56)
	β	0.07	0.03	0.03	0.06
Math Anxiety	<i>t</i>	2.33	0.91	1.24	2.04
	<i>p</i>	0.020	0.364	0.217	0.041
	95% CI	[0.55, 6.51]	[-1.58, 4.31]	[-1.09, 4.79]	[0.13, 6.26]
	<i>B</i> (SE)	-0.43 (0.28)	-0.28 (0.27)	-0.40 (0.27)	-0.23 (0.28)
Math Anxiety	β	-0.05	-0.03	-0.04	-0.02
	<i>t</i>	-1.53	-1.04	-1.46	-0.80
	<i>p</i>	0.126	0.300	0.144	0.422
	95% CI	[-0.97, 0.12]	[-0.82, 0.25]	[-0.93, 0.14]	[-0.79, 0.33]

Table 6 (continued)

Variable	Statistic	Pretest	Posttest Problem #1	Posttest Problem #2	Posttest Problem #3
Trait Anxiety	<i>B</i> (<i>SE</i>)	-0.07 (0.05)	-0.05 (0.05)	-0.08 (0.05)	-0.16 (0.06)
	β	-0.04	-0.03	-0.04	-0.07
	<i>t</i>	-1.29	-0.96	-1.42	-2.83
	<i>p</i>	0.196	0.336	0.157	0.005
Pretest	95% CI	[-0.18, 0.04]	[-0.16, 0.05]	[-0.18, 0.03]	[-0.27, -0.05]
Posttest Problem #1		$F(7, 1169) = 18.60, p < 0.001, R^2 = 0.10, \text{Adjusted } R^2 = 0.10$			
Posttest Problem #2		$F(10, 1166) = 47.38, p < 0.001, R^2 = 0.29, \text{Adjusted } R^2 = 0.28$			
Posttest Problem #3		$F(10, 1166) = 48.60, p < 0.001, R^2 = 0.29, \text{Adjusted } R^2 = 0.29$			
		$F(10, 1166) = 51.61, p < 0.001, R^2 = 0.31, \text{Adjusted } R^2 = 0.30$			

Appendix 2

Table 7 Full Linear Regression Coefficients for Research Question 1 (With Math Self-Efficacy Removed From the Model)

Variable	Statistic	Pretest	Posttest Problem #1	Posttest Problem #2	Posttest Problem #3
Condition	<i>B</i> (SE)		1.90 (1.24)	2.02 (1.19)	2.80 (1.24)
	β		0.04	0.04	0.06
	<i>t</i>		1.54	1.70	2.25
	<i>p</i>		0.125	0.089	0.025
	95% CI		[-0.53, 4.33]	[-0.31, 4.34]	[0.36, 5.24]
Pretest Metacognitive Judgment	<i>B</i> (SE)		0.50 (0.03)	0.52 (0.03)	0.50 (0.03)
	β		0.45	0.47	0.43
	<i>t</i>		17.46	18.27	16.75
	<i>p</i>		<0.001	<0.001	<0.001
	95% CI		[0.44, 0.55]	[0.46, 0.58]	[0.44, 0.56]
Pretest Accuracy	<i>B</i> (SE)		0.86 (1.33)	-3.35 (1.31)	-0.92 (1.40)
	β		0.02	-0.07	-0.02
	<i>t</i>		0.64	-2.55	-0.65
	<i>p</i>		0.520	0.011	0.513
	95% CI		[-1.75, 3.46]	[-5.93, -0.78]	[-3.66, 1.83]
Gender	<i>B</i> (SE)	6.86 (1.25)	3.71 (1.24)	1.86 (1.23)	4.50 (1.30)
	β	0.16	0.08	0.04	0.09
	<i>t</i>	5.48	2.99	1.51	3.47
	<i>p</i>	<0.001	0.003	0.131	0.001
	95% CI	[4.40, 9.31]	[1.27, 6.14]	[-0.56, 4.28]	[1.95, 7.04]
Percent Absolute Error (PAE)	<i>B</i> (SE)	-9.36 (8.51)	-14.25 (8.36)	-6.16 (8.39)	10.93 (8.83)
	β	-0.04	-0.05	-0.02	0.04
	<i>t</i>	-1.10	-1.71	-0.74	1.24
	<i>p</i>	0.272	0.089	0.463	0.216
	95% CI	[-26.05, 7.34]	[-30.66, 2.15]	[-22.62, 10.30]	[-6.39, 28.25]

Table 7 (continued)

Variable	Statistic	Pretest	Posttest Problem #1	Posttest Problem #2	Posttest Problem #3
Accuracy on the Current Problem	<i>B</i> (SE)	0.12 (1.33)	3.84 (1.47)	3.55 (1.49)	7.53 (1.39)
	β	0.00	0.07	0.07	0.15
	<i>t</i>	0.09	2.61	2.39	5.42
	<i>p</i>	0.931	0.009	0.017	<0.001
	95% CI	[-2.50, 2.73]	[0.95, 6.72]	[0.63, 6.47]	[4.80, 10.26]
Objective Numeracy	<i>B</i> (SE)	3.99 (1.54)	1.53 (1.51)	2.02 (1.51)	3.46 (1.59)
	β	0.08	0.03	0.04	0.06
	<i>t</i>	2.59	1.02	1.34	2.18
	<i>p</i>	0.010	0.310	0.182	0.029
	95% CI	[0.96, 7.01]	[-1.43, 4.50]	[-0.94, 4.97]	[0.35, 6.56]
Math Anxiety	<i>B</i> (SE)	-0.78 (0.28)	-0.49 (0.27)	-0.61 (0.27)	-0.55 (0.28)
	β	-0.09	-0.05	-0.07	-0.06
	<i>t</i>	-2.80	-1.81	-2.27	-1.93
	<i>p</i>	0.005	0.070	0.023	0.054
	95% CI	[-1.32, -0.23]	[-1.02, 0.04]	[-1.14, -0.08]	[-1.10, 0.01]
Trait Anxiety	<i>B</i> (SE)	-0.10 (0.06)	-0.07 (0.05)	-0.09 (0.05)	-0.18 (0.06)
	β	-0.06	-0.03	-0.05	-0.09
	<i>t</i>	-1.87	-1.30	-1.76	-3.29
	<i>p</i>	0.062	0.194	0.079	0.001
	95% CI	[-0.21, 0.01]	[-0.17, 0.04]	[-0.20, 0.01]	[-0.29, -0.07]
Pretest	<i>F</i> (6, 1170) = 14.48, <i>p</i> < 0.001, <i>R</i> ² = 0.07, Adjusted <i>R</i> ² = 0.06				
Posttest Problem #1	<i>F</i> (9, 1167) = 50.11, <i>p</i> < 0.001, <i>R</i> ² = 0.28, Adjusted <i>R</i> ² = 0.27				
Posttest Problem #2	<i>F</i> (9, 1167) = 51.21, <i>p</i> < 0.001, <i>R</i> ² = 0.28, Adjusted <i>R</i> ² = 0.28				
Posttest Problem #3	<i>F</i> (9, 1167) = 51.85, <i>p</i> < 0.001, <i>R</i> ² = 0.29, Adjusted <i>R</i> ² = 0.28				

Appendix 3

Table 8 Full Hierarchical Linear Regression Coefficients for Hypothesis 2 (Adding Math Anxiety to the Model in Block 2)

Variable	Statistic	Pretest		Posttest Problem #1		Posttest Problem #2		Posttest Problem #3	
		Block 1	Block 2	Block 1	Block 2	Block 1	Block 2	Block 1	Block 2
Condition	<i>B</i> (SE)			2.91 (1.20)	2.99 (1.19)	2.30 (1.19)	2.37 (1.18)	2.72 (1.26)	2.79 (1.26)
	β			0.06	0.06	0.05	0.05	0.06	0.06
	<i>t</i>			2.43	2.51	1.94	2.00	2.16	2.22
	<i>p</i>			0.015	0.012	0.053	0.045	0.031	0.027
	95% CI			[0.56, 5.26]	[0.65, 5.33]	[-0.03, 4.63]	[0.05, 4.69]	[0.25, 5.20]	[0.33, 5.26]
Pretest Monitoring Judgment	<i>B</i> (SE)			0.51 (0.03)	0.50 (0.03)	0.54 (0.03)	0.53 (0.03)	0.52 (0.03)	0.51 (0.03)
	β			0.47	0.45	0.49	0.48	0.45	0.44
	<i>t</i>			18.07	17.59	19.20	18.74	17.44	17.01
	<i>p</i>			<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	95% CI			[0.46, 0.57]	[0.45, 0.56]	[0.49, 0.60]	[0.49, 0.60]	[0.46, 0.58]	[0.45, 0.57]
Gender	<i>B</i> (SE)	8.23 (1.23)	7.65 (1.23)	5.06 (1.22)	4.69 (1.22)	2.42 (1.21)	2.10 (1.21)	5.29 (1.29)	4.98 (1.29)
	β	0.19	0.18	0.11	0.10	0.05	0.04	0.11	0.10
	<i>t</i>	6.70	6.23	4.14	3.84	2.00	1.73	4.11	3.86
	<i>p</i>	<0.001	<0.001	<0.001	<0.001	0.046	0.084	<0.001	<0.001
	95% CI	[5.82, 10.64]	[5.24, 10.06]	[2.66, 7.46]	[2.30, 7.09]	[0.04, 4.80]	[-0.28, 4.47]	[2.76, 7.81]	[2.45, 7.50]
Trait Anxiety	<i>B</i> (SE)	-0.16 (0.05)	-0.09 (0.06)	-0.11 (0.05)	-0.05 (0.05)	-0.15 (0.05)	-0.10 (0.05)	-0.23 (0.05)	-0.19 (0.06)
	β	-0.09	-0.05	-0.06	-0.03	-0.07	-0.05	-0.11	-0.09
	<i>t</i>	-3.09	-1.64	-2.16	-1.01	-2.86	-1.79	-4.34	-3.29
	<i>p</i>	0.002	0.102	0.031	0.311	0.004	0.074	<0.001	0.001
	95% CI	[-0.26, -0.06]	[-0.20, 0.02]	[-0.21, -0.01]	[-0.16, 0.05]	[-0.25, -0.05]	[-0.20, 0.01]	[-0.34, -0.13]	[-0.30, -0.08]

Table 8 (continued)

Variable	Statistic		Pretest		Posttest Problem #1		Posttest Problem #2		Posttest Problem #3	
	Block 1	Block 2	Block 1	Block 2	Block 1	Block 2	Block 1	Block 2	Block 1	Block 2
Math Anxiety										
<i>B</i> (SE)		-1.09 (0.26)		-0.88 (0.25)		-0.78 (0.25)		-0.76 (0.27)		
β		-0.13		-0.09		-0.08		-0.08		
<i>t</i>		-4.23		-3.47		-3.12		-2.83		
<i>p</i>		< 0.001		0.001		0.002		0.005		
95% CI		[-1.59, -0.58]		[-1.37, -0.38]		[-1.28, -0.29]		[-1.28, -0.23]		
Pretest										
		$R^2=0.05$, Adjusted $R^2=0.04$, $F(2, 1174)=28.10$, $p<0.001$								
		$R^2=0.06$, Adjusted $R^2=0.06$, $F(3, 1173)=24.97$, $p<0.001$								
		Change from Model 1 to Model 2: $\Delta R^2=0.01$, $\Delta F(1, 1173)=17.90$, $p<0.001$								
Posttest Problem #1										
		$R^2=0.26$, Adjusted $R^2=0.26$, $F(4, 1172)=102.61$, $p<0.001$								
		$R^2=0.27$, Adjusted $R^2=0.26$, $F(5, 1171)=85.26$, $p<0.001$								
		Change from Model 1 to Model 2: $\Delta R^2=0.01$, $\Delta F(1, 1171)=12.01$, $p=0.001$								
Posttest Problem #2										
		$R^2=0.27$, Adjusted $R^2=0.27$, $F(4, 1172)=107.74$, $p<0.001$								
		$R^2=0.28$, Adjusted $R^2=0.27$, $F(5, 1171)=88.78$, $p<0.001$								
		Change from Model 1 to Model 2: $\Delta R^2=0.01$, $\Delta F(1, 1171)=9.75$, $p=0.002$								
Posttest Problem #3										
		$R^2=0.26$, Adjusted $R^2=0.26$, $F(4, 1172)=101.69$, $p<0.001$								
		$R^2=0.26$, Adjusted $R^2=0.26$, $F(5, 1171)=83.45$, $p<0.001$								
		Change from Model 1 to Model 2: $\Delta R^2=0.01$, $\Delta F(1, 1171)=8.02$, $p=0.005$								

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11409-022-09300-3>.

Authors' contributions DAS: Conceptualization, analyses, first draft writing, revisions.

CAT: Collection of data from parent study, conceptualization, revisions.

CJF: Collection of data from parent study, analyses, conceptualization, revisions.

MKM: Conceptualization, revisions.

JMT: Conceptualization, revisions.

PGS: Conceptualization, revisions.

KC: Conceptualization, revisions.

Funding This research was supported in part by the U.S. Department of Education Institute of Education Sciences Grants R305A160295 and R305U200004 to C. A. Thompson at Kent State University.

Data availability This study was pre-registered on the Open Science Framework (OSF; <https://osf.io/vxm8d>). All statistical output will be available through OSF.

Code availability Syntax available upon request; output available through OSF.

Declarations

Conflicts of interest/competing interests The authors declare they have no conflicts of interest or competing interests.

References

- Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, 21(8), 607–617. <https://doi.org/10.1016/j.tics.2017.05.004>
- Afifi, W. A., & Weiner, J. L. (2004). Toward a theory of motivated information management. *Communication Theory*, 14, 167–190. <https://doi.org/10.1111/j.1468-2885.2004.tb00310.x>
- Ahmed, W., Minnaert, A., Kuyper, H., & van der Werf, G. (2012). Reciprocal relationships between math self-concept and math anxiety. *Learning and Individual Differences*, 22(3), 385–389. <https://doi.org/10.1016/j.lindif.2011.12.004>
- Alibali, M. W., & Sidney, P. G. (2015). Variability in the natural number bias: Who, when, how, and why. *Learning and Instruction*, 37, 56–61. <https://doi.org/10.1016/j.learninstruc.2015.01.003>
- Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Directions in Psychological Science*, 11(50), 181–185. <https://doi.org/10.1111/1467-8721.00196>
- Ashcraft, M. H. (2019). Models of math anxiety. In I. C. Mammarella, S. Caviola, & A. Dowker (Eds.), *Mathematics Anxiety: What is known and what is still to be understood*. Routledge.
- Ashcraft, M. H., & Ridley, K. S. (2005). Math anxiety and its cognitive consequences: A tutorial review. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 315–327). Psychology Press.
- Baars, M., van Gog, T., de Bruin, A., & Paas, F. (2017). Effects of problem solving after worked example study on secondary school children's monitoring accuracy. *Educational Psychology*, 37(7), 810–834. <https://doi.org/10.1080/01443410.2016.1150419>
- Bai, H. (2011). Cross-validating a bidimensional mathematics anxiety scale. *Assessment*, 18(1), 115–122. <https://doi.org/10.1177/1073191110364312>
- Baloğlu, M. (2004). Statistics anxiety and mathematics anxiety: Some interesting differences. *Contemporary Educational Research Quarterly*, 27(3), 38–48.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122–147. <https://doi.org/10.1037/0003-066X.37.2.122>
- Barroso, C., Ganley, C. M., McGraw, A. L., Geer, E. A., Hart, S. A., & Daucourt, M. C. (2020). A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin*. Advance online publication. <https://doi.org/10.1037/bul0000307>

- Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, 107(5), 1860–1863. <https://doi.org/10.1073/pnas.0910967107>
- Carey, E., Hill, F., Devine, A., & Szűcs, D. (2017). The modified abbreviated math anxiety scale: A valid and reliable instrument for use with children. *Frontiers in Psychology*, 8, 1–13. <https://doi.org/10.3389/fpsyg.2017.00011>
- Choi, S. S., Taber, J. M., Thompson, C. A., & Sidney, P. G. (2020). Math anxiety, but not induced stress, is associated with objective numeracy. *Journal of Experimental Psychology: Applied*. Advance online publication. <https://doi.org/10.1037/xap0000268>
- Cipora, K., Artemenko, C., & Nuerk, H.-C. (2019). Different ways to measure math anxiety. In I. C. Mammarella, S. Caviola, & A. Dowker (Eds.), *Mathematics Anxiety: What is known and what is still to be understood*. Routledge.
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making*, 7, 25–47.
- Cresswell, C., Cooper, P. J., & Murray, L. (2010). Intergenerational transmission of anxious information processing biases. In J. Hadwin & A. P. Field (Eds.), *Information processing biases and anxiety: A developmental perspective* (pp. 279–295). Wiley-Blackwell.
- Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, 29(5), 761–778. <https://doi.org/10.1177/0956797617744771>
- Desender, K., & Sasanguie, D. (2021). Math anxiety relates positively to metacognitive insight into mathematical decision making. *Psychological Research*. Advanced online publication. <https://doi.org/10.1007/s00426-021-01511-8>
- Devine, A., Fawcett, K., Szűcs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions*, 8(33), 1–9. <https://doi.org/10.1186/1744-9081-8-33>
- Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, 7(508), 1–16. <https://doi.org/10.3389/fpsyg.2016.00508>
- Dreger, R. M., & Aiken, L. R. (1957). The identification of number anxiety in a college population. *Journal of Educational Psychology*, 48(6), 344–351. <https://doi.org/10.1037/h0045894>
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Sage Publications.
- Dunlosky, J., & Rawson, K. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, 22(4), 271–280. <https://doi.org/10.1016/j.learninstruc.2011.08.003>
- Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction*, 24, 58–61. <https://doi.org/10.1016/j.learninstruc.2012.05.002>
- Dunning, D., Kerri, J., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12, 83–87. <https://doi.org/10.1111/1467-8721.01235>
- Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1, 3–14. <https://doi.org/10.1016/j.edurev.2005.11.001>
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, 136(1), 103–127. <https://doi.org/10.1037/a0018053>
- Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a mathematics test: Development of the subjective numeracy scale. *Medical Decision Making*, 27, 672–680. <https://doi.org/10.1177/0272989X07304449>
- Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, 123, 53–72. <https://doi.org/10.1016/j.jecp.2014.01.013>
- Fazio, L. K., DeWolf, M., & Siegler, R. S. (2017). Strategy use and strategy choice in fraction magnitude comparison. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, 42, 1–16. <https://doi.org/10.1037/xlm0000153>
- Fitzsimmons, C. J., Morehead, K., Thompson, C. A., Buerke, M., & Dunlosky, J. (2021). Can feedback, correct, and incorrect worked examples improve numerical magnitude estimation precision? *Journal of Experimental Education*. Advanced online publication. <https://doi.org/10.1080/00220973.2021.1891009>
- Fitzsimmons, C. J., & Thompson, C. A. (2021). *Is there whole-number bias in children's and adults' monitoring of magnitude estimation? The differential role of familiarity when monitoring whole-number and fraction estimates [Manuscript submitted for publication]*. Kent State University.
- Fitzsimmons, C. J., Thompson, C. A., & Sidney, P. G. (2020). Confident or familiar? The role of familiarity ratings in adults' confidence judgments when estimating fraction magnitudes. *Metacognition and Learning*, 15, 215–231. <https://doi.org/10.1007/s11409-020-09225-9>

- Fitzsimmons, C., Thompson, C. A., & Sidney, P. (2020b). Do adults treat equivalent fractions equally? Adults' strategies and errors during fraction reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(11), 2049–2074. <https://doi.org/10.1037/xlm0000839>
- Frenkel, S., Alba, D., & Zhong, R. (2020, March 8). Surge of virus misinformation stumps Facebook and Twitter. *The New York Times*. Retrieved from <https://www.nytimes.com/2020/03/08/technology/coronavirus-misinformation-social-media.html>
- Funder, & D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and non-sense. *Advances in Methods and Practices in Psychological Science*, *2*, 156–168. <https://doi.org/10.1177/2515245919847202>
- Geary, D. C., Scofield, J. E., Hoard, M. K., & Nugent, L. (2020). Boys' advantage on the fractions number line is mediated by visuospatial attention: Evidence for a parietal-spatial contribution to number line learning. *Developmental Science*, 1-19. <https://doi.org/10.1111/desc.13063>
- Gough, M. F. (1954). Mathemaphobia: Causes and treatments. *Clearing House*, *28*, 290–294.
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*, 1–51. <https://doi.org/10.1111/j.1529-1006.2007.00032.x>
- Händel, M., de Bruin, A. B. H., & Dresel, M. (2020). Individual differences in local and global metacognitive judgments. *Metacognition and Learning*, *15*, 51–75. <https://doi.org/10.1007/s11409-020-09220-0>
- Harari, R. R., Vukovic, R. K., & Bailey, S. P. (2013). Mathematics anxiety in young children: An exploratory study. *Journal of Experimental Education*, *81*(4), 538–555. <https://doi.org/10.1080/00220973.2012.727888>
- Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, *21*(1), 33–46. <https://doi.org/10.2307/749455>
- Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. A. (2003). The Abbreviated Math Anxiety Scale (AMAS): Construction, validity, and reliability. *Assessment*, *10*, 178–182. <https://doi.org/10.1177/1073191103010002008>
- Huang, X., Zhang, J., & Hudson, L. (2019). Impact of math self-efficacy, math anxiety, and growth mindset on math and science career interest for middle school students: The gender moderating effect. *European Journal of Psychological Education*, *34*, 621–640. <https://doi.org/10.1007/s10212-018-0403-z>
- Hunt, T. E., Clark-Carter, D., & Sheffield, D. (2011). The development and part validation of a U.K. scale for mathematics anxiety. *Journal of Psychoeducational Assessment*, *29*(5), 455–466. <https://doi.org/10.1177/0734282910392892>
- Hutchison, J. E., Lyons, I. M., & Ansari, D. (2019). More similar than different: Gender differences in children's basic numerical skills are the exception not the rule. *Child Development*, *90*, e66–e79. <https://doi.org/10.1111/cdev.13044>
- Jain, S., & Dowson, M. (2009). Mathematics anxiety as a function of multidimensional self-regulation and self-efficacy. *Contemporary Educational Psychology*, *34*(3), 240–249. <https://doi.org/10.1016/j.cedpsych.2009.05.004>
- Jameson, M. M. (2013). The development and validation of the children's anxiety in math scale. *Journal of Psychoeducational Assessment*, *31*(4), 391–395. <https://doi.org/10.1177/0734282912470131>
- Jamieson, J. P., Mendes, W. B., Blackstock, E., & Schmader, T. (2010). Turning the knots in your stomach into bows: Reappraising arousal improves performance on the GRE. *Journal of Experimental Social Psychology*, *46*, 208–212. <https://doi.org/10.1016/j.jesp.2009.08.015>
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. <https://doi.org/10.1037/0096-3445.126.4.349>
- Koriat, A. (2011). Subjective confidence in perceptual judgments: A test of the self-consistency model. *Journal of Experimental Psychology: General*, *140*(1), 117–139. <https://doi.org/10.1037/a0022171>
- Koriat, A., & Adiv, S. (2016). The self-consistency theory of subjective confidence. In J. Dunlosky and S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 127–148). <https://doi.org/10.1093/oxfordhb/9780199336746.013.18>
- Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 483–502). The Guilford Press.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 107–118. <https://doi.org/10.1037/0278-7393.6.2.107>
- Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp 117–135). Psychology Press.

- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning*, 5, 173–194. <https://doi.org/10.1007/s11409-010-9056-2>
- Låg, T., Bauger, L., Lindberg, M., & Friborg, O. (2014). The role of numeracy and intelligence in health-risk estimation and medical data interpretation. *Behavioral Decision Making*, 27(2), 95–108. <https://doi.org/10.1002/bdm.1788>
- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences*, 19, 355–365. <https://doi.org/10.1016/j.lindif.2008.10.009>
- Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(3), 464–470. <https://doi.org/10.1037/0278-7393.16.3.464>
- Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21(1), 37–44. <https://doi.org/10.1177/0272989X0102100105>
- Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, 30(5), 520–540. <https://doi.org/10.2307/749772>
- Miller-Cotto, D., & Lewis, N. A., Jr. (2020). Am I a “math person?” How classroom cultures shape math identity among black and latinx students. <https://doi.org/10.31219/osf.io/hcqt>
- Morony, S., Kleitman, S., Lee, Y. P., & Stankov, L. (2013). Predicting achievement: Confidence vs. self-efficacy anxiety, and self-concept in Confucian and European countries. *International Journal of Educational Research*, 58, 79–96. <https://doi.org/10.1016/j.ijer.2012.11.002>
- Morsanyi, K., Busdraghi, C., & Primi, C. (2014). Mathematical anxiety is linked to reduced cognitive reflection: A potential road from discomfort in the mathematics classroom to susceptibility to biases. *Behavioral and Brain Functions*, 10(1), 1–13.
- Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language*, 93, 245–258. <https://doi.org/10.1016/j.jml.2016.10.008>
- Namkung, J. M., Peng, P., & Lin, X. (2019). The relation between mathematics anxiety and mathematics performance among school-aged students: A meta-analysis. *Review of Educational Research*, 89(3), 459–496. <https://doi.org/10.3102/0034654319843494>
- Nelson, L. J., & Fyfe, E. R. (2019). Metacognitive monitoring and help-seeking decisions on mathematical equivalence problems. *Metacognition and Learning*, 14(2), 167–187. <https://doi.org/10.1007/s11409-019-09203-w>
- Ni, Y., & Zhou, Y. D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist*, 40, 27–52. https://doi.org/10.1207/s15326985ep4001_3
- Nolen, S. B., Ward, C. J., & Horn, I. S. (2014). In P. W. Richardson, H. M. G. Watt, & S. A. Karabenick (Eds.), *Teacher Motivation: Theory and Practice*. New York: NY Routledge.
- Núñez-Peña, M. I., Guilera, G., & Suárez-Pellicioni, M. (2014). The single-item math anxiety scale: An alternative way of measuring mathematical anxiety. *Journal of Psychoeducational Assessment*, 34(4), 306–317. <https://doi.org/10.1177/0734282913508528>
- Ooi, J., Dodd, H. F., & Walsh, J. (2015). Shared cognition in childhood anxiety: Interpretation bias in pre-school children and their parents. *Journal of Child and Family Studies*, 24, 3413–3422. <https://doi.org/10.1007/s10826-015-0143-5>
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543–578. <https://doi.org/10.3102/00346543066004543>
- Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, 86(2), 193–203. <https://doi.org/10.1037/0022-0663.86.2.193>
- Park, D., Ramirez, G., & Beilock, S. L. (2014). The role of expressive writing in math anxiety. *Journal of Experimental Psychology: Applied*, 20(2), 103–111. <https://doi.org/10.1037/xap0000013>
- Pennycook, G., McPhetres, J., Zhang, Y., Lu, G. L., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, 31(7), 770–780. <https://doi.org/10.1177/0956797620939054>
- Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning-Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin Review*, 24, 1774–1784. <https://doi.org/10.3758/s13423-017-1242-7>

- Peters, E. (2020). Innumeracy in the Wild. *Oxford University Press*. <https://doi.org/10.1093/oso/9780190861094.003.0001>
- Peters, E., Tompkins, M. K., Knoll, M. A. Z., Ardoin, S. P., Shoots-Reinhard, B., & Meara, A. S. (2019). Despite high objective numeracy, lower numeric confidence relates to worse financial and medical outcomes. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(39), 19386–19391. <https://doi.org/10.1073/pnas.1903126116>
- Pruden, S. M., Levine, S. C., & Huttenlocker, J. (2011). Children's spatial thinking: Does talk about the spatial world matter? *Developmental Science*, *14*(6), 1417–1430. <https://doi.org/10.1111/j.1467-7687.2011.01088.x>
- Punaro, L., & Reeve, R. (2012). Relationships between 9-year-olds' math and literacy worries and academic abilities. *Child Development Research*, *2*. <https://doi.org/10.1155/2012/359089>
- Ramirez, G., & Beilock, S. L. (2011). Writing about testing worries boosts exam performance in the classroom. *Science*, *331*, 211–213. <https://doi.org/10.1126/science.1199427>
- Reder, L. (1987). Strategy selection in question answering. *Cognitive Psychology*, *19*(1), 90–138. [https://doi.org/10.1016/0010-0285\(87\)90005-3](https://doi.org/10.1016/0010-0285(87)90005-3)
- Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 435–451. <https://doi.org/10.1037/0278-7393.18.3.435>
- Rhodes, M. (2019). Metacognition. *Teaching of Psychology*, *46*(2), 168–175. <https://doi.org/10.1177/0098628319834381>
- Rinne, L. F., & Mazzocco, M. M. M. (2014). Knowing right from wrong in mental arithmetic judgments: Calibration of confidence predicts the development of accuracy. *PLoS ONE*, *9*(7), e98663. <https://doi.org/10.1371/journal.pone.0098663>
- Rivers, M. L., Fitzsimmons, C. J., Fisk, S. R., Dunlosky, J., & Thompson, C. A. (2020). Gender differences in confidence during number-line estimation. *Metacognition and Learning*. <https://doi.org/10.1007/s11409-020-09243-7>
- Robinson, D. H., & Levin, J. R. (2019). Quackery in educational research. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 35–48). Boston: Cambridge University Press. <https://doi.org/10.1017/9781108235631.003>
- Robinson, D. H., Levin, J., Schraw, G., Patall, E. A., & Hunt, E. B. (2013). On going (way) beyond one's data: A proposal to restrict recommendations for practice in primary educational research journals. *Educational Psychology Review*, *25*(2), 291–302. <https://doi.org/10.1007/s10648-013-9223-5>
- Rolison, J. J., Morsanyi, K., & O'Connor, P. A. (2016). Can I count on getting better? Association between math anxiety and poorer understanding of medical risk reductions. *Medical Decision Making*, *36*, 876–886. <https://doi.org/10.1177/0272989X15602000>
- Russonello, G. (2020, March 13). Afraid of coronavirus? That might say something about your politics. *The New York Times*. Retrieved from <https://www.nytimes.com/2020/03/13/us/politics/coronavirus-trump-polling.html>
- Sanchez, C., & Dunning, D. (2020). Decision fluency and overconfidence among beginners. *Decision*, *7*(3), 225–237. <https://doi.org/10.1037/dec0000122>
- Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of number line estimation with mathematical competence: A meta-analysis. *Child Development*, *89*(5), 1467–1484. <https://doi.org/10.1111/cdev.13068>
- Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, *127*(11), 966–972. <https://doi.org/10.7326/0003-4819-127-11-199712010-00003>
- Sidney, P. G., Thalluri, R., Buerke, M. L., & Thompson, C. A. (2018). Who uses more strategies? Linking mathematics anxiety to adults' strategy variability and performance on fraction magnitude tasks. *Thinking and Reasoning*, *25*(1), 94–131. <https://doi.org/10.1080/13546783.2018.1475303>
- Sidney, P. G., Thompson, C. A., Fitzsimmons, C., & Taber, J. M. (2021). Children's and adults' math attitudes are differentiated by number type. *The Journal of Experimental Education*, 1-32. <https://doi.org/10.1080/00220973.2019.1653815>
- Siegler, R. S. (2016). Magnitude knowledge: The common core of numerical development. *Developmental Science*, *19*(3), 341–361. <https://doi.org/10.1111/desc.12395>
- Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*(2), 428–444. <https://doi.org/10.1111/j.1467-8624.2004.00684.x>
- Siegler, R. S., & Thompson, C. A. (2014). Numerical landmarks are useful—except when they're not. *Journal of Experimental Child Psychology*, *120*, 39–58. <https://doi.org/10.1016/j.jecp.2013.11.014>
- Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology*, *62*, 273–296. <https://doi.org/10.1016/j.cogpsych.2011.03.001>

- Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, 67, 415–437. <https://doi.org/10.1146/annurev-psych-073115-103235>
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *The state-trait anxiety inventory (test manual)*. Palo Alto, CA: Consulting Psychologists Press.
- Stankov, L., Kleitman, S., & Jackson, S. A. (2015). Measures of the trait of confidence. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of personality and social psychological constructs* (pp. 158–189). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-386915-9.00007-3>
- Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences*, 22, 747–758. <https://doi.org/10.1016/j.lindif.2012.05.013>
- Supekar, K., Iuculano, T., Chen, L., & Menon, V. (2015). Remediation of childhood math anxiety and associated neural circuits through cognitive tutoring. *The Journal of Neuroscience*, 35, 12574–12583. <https://doi.org/10.1523/JNEUROSCI.0786-15.2015>
- Sweeny, K., Melnyk, D., Miller, W., & Shepperd, J. A. (2010). Information avoidance: Who, what, when, and why. *Review of General Psychology*, 14(4), 340–353. <https://doi.org/10.1037/a0021288>
- Taber, J. M., & Klein, W. M. P. (2016). The role of conviction in personal disease risk perceptions: What can we learn from research on attitude strength? *Social and Personality Psychology Compass*, 10(4), 202–218. <https://doi.org/10.1111/spc3.12244>
- Thompson, C. A., Mielicki, M., Rivera, F., Fitzsimmons, C. J., Scheibe, D. A., Sidney, P. G., Taber, J. M., & Waters, E. A. (in press). Leveraging math cognition to combat health innumeracy. *Perspectives on Psychological Science*
- Thompson, C. A., & Opfer, J. E. (2008). Costs and benefits of representational change: Effects of context on age and sex differences in symbolic magnitude estimation. *Journal of Experimental Child Psychology*, 101(1), 20–51. <https://doi.org/10.1016/j.jecp.2008.02.003>
- Thompson, C. A., Taber, J. M., Sidney, P. G., Fitzsimmons, C. J., Mielicki, M. K., Matthews, P. G., Schemmel, E. A., Simonovic, N., Foust, J. L., Aurora, P., Disabato, D. J., Seah, T. H. S., & Coifman, K. G. (2021). Math matters: A novel, brief educational intervention decreases whole number bias when reasoning about COVID-19. *Journal of Experimental Psychology: Applied*, 27(4), 632–656. <https://doi.org/10.1037/xap0000403>
- Tomasetto, C. (2019). Gender stereotypes, anxiety, and math outcomes in adults and children. In I. C. Mammarella, S. Caviola, & A. Dowker (Eds.), *Mathematics Anxiety: What is known and what is still to be understood*. Routledge.
- Wall, J. L., Thompson, C. A., Dunlosky, J., & Merriman, W. E. (2016). Children can accurately monitor and control their number-line estimation performance. *Developmental Psychology*, 52(10), 1493–1502. <https://doi.org/10.1037/dev0000180>
- Waters, E. A., Biddle, C., Kaphingst, K. A., Schofield, E., Kiviniemi, M. T., Orom, H., Li, Y., & Hay, J. L. (2018). Examining the interrelations among objective and subjective health literacy and numeracy and their associations with health knowledge. *Journal of General Internal Medicine*, 33(11), 1945–1953. <https://doi.org/10.1007/s11606-018-4624-2>
- Waters, E. A., Fagerlin, A., & Zikmund-Fisher, B. J. (2016). Overcoming the many pitfalls of communicating risk. In M. A. Diefenbach, S. Miller-Halegoua, & D. J. Bowen (Eds.), *Handbook of Health Decision Science*, pp. 265–277. New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3486-7_19
- Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, 26, 198–212. <https://doi.org/10.1002/bdm.1751>
- Wigfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *Journal of Educational Psychology*, 80(2), 210–216. <https://doi.org/10.1037/0022-0663.80.2.210>
- Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, 27(4), 552–565. <https://doi.org/10.1037/0012-1649.27.4.552>
- Yáñez-Marquina, L., & Villardón-Gallego, L. (2017). Math anxiety, a hierarchical construct: Development and validation of the Scale for Assessing Math Anxiety in Secondary education. *Ansiedad y Estrés*, 23, 59–65. <https://doi.org/10.1016/j.anyes.2017.10.001>
- Zhang, J., Zhao, N., & Kong, Q. P. (2019). The relationship between math anxiety and math performance: A meta-analytic investigation. *Frontiers in Psychology*, 10, 1–17. <https://doi.org/10.3389/fpsyg.2019.01613>