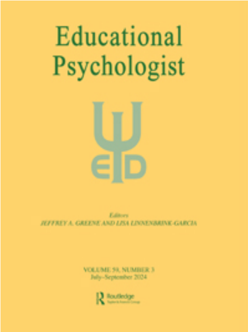

Educational Psychologist

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hedp20

The effectiveness of refutation text in confronting scientific misconceptions: A meta-analysis

Robert W. Danielson, Neil G. Jacobson, Erika A. Patall, Gale M. Sinatra, Olusola O. Adesope, Alana A. U. Kennedy, Bethany H. Bhat, Onur Ramazan, Blessing Akinrotimi, Gabriel Nketah, Gan Jin & Oluwafemi J. Sunday

To cite this article: Robert W. Danielson, Neil G. Jacobson, Erika A. Patall, Gale M. Sinatra, Olusola O. Adesope, Alana A. U. Kennedy, Bethany H. Bhat, Onur Ramazan, Blessing Akinrotimi, Gabriel Nketah, Gan Jin & Oluwafemi J. Sunday (16 Aug 2024): The effectiveness of refutation text in confronting scientific misconceptions: A meta-analysis, Educational Psychologist, DOI: 10.1080/00461520.2024.2365628

To link to this article: https://doi.org/10.1080/00461520.2024.2365628

EDUCATIONAL PSYCHOLOGIST
https://doi.org/10.1080/00461520.2024.2365628

The effectiveness of refutation text in confronting scientific misconceptions: A meta-analysis

Robert W. Danielson[a], Neil G. Jacobson[b], Erika A. Patall[b], Gale M. Sinatra[b], Olusola O. Adesope[a], Alana A. U. Kennedy[b*], Bethany H. Bhat[c], Onur Ramazan[a], Blessing Akinrotimi[a], Gabriel Nketah[d], Gan Jin[a] and Oluwafemi J. Sunday[a]

[a]College of Education, Washington State University, Pullman, WA, USA; [b]Rossier School of Education, University of Southern California, Los Angeles, SC, USA; [c]Department of Psychology, The University of Texas, Austin, TX, USA; [d]Voiland School of Chemical Engineering and Bioengineering, Washington State University, Pullman, WA, USA

ABSTRACT
Misinformation around scientific issues is rampant on social media platforms, raising concerns among educators and science communicators. A variety of approaches have been explored to confront this growing threat to science literacy. For example, refutations have been used both proactively as warning labels and in attempts to inoculate against misconceptions, and retroactively to debunk misconceptions and rebut science denialism. Refutations have been used by policy makers and scientists when communicating with the general public, yet little is known about their effectiveness or consequences. Given the interest in refutational approaches, we conducted a comprehensive, pre-registered meta-analysis comparing the effect of refutation texts to non-refutation texts on individuals' misconceptions about scientific information. We selected 71 articles (53 published and 18 unpublished) that described 76 studies, 111 samples, and 294 effect sizes. We also examined 26 moderators. Overall, our findings show a consistent and statistically significant advantage of refutation texts over non-refutation texts in controlled experiments confronting scientific misconceptions. We also found that moderators neither enhanced nor diminished the impact of the refutation texts. We discuss the implications of using refutations in formal and informal science learning contexts and in science communications from three theoretical perspectives.

Information has never been more easily accessible. A student with a smartphone and an internet connection can watch free lectures from the most prestigious universities or read innovative research from world class researchers on nearly any topic at any time. For example, with the advent of platforms like Twitter (X), anyone can reach out directly to climate change experts such as Katharine Hayhoe or Michael Mann and converse with them in real-time. However, the democratization of the internet has allowed for an explosion of inaccurate information. A student with a smart phone and internet connection who is searching for cutting-edge research on climate change must now contend with an information landscape that can be inaccurate at best (Allcott et al., 2019; Kata, 2012; Kortum et al., 2008; Scheufele & Krause, 2019) and malicious at worst (Fisher, 2022).

In this landscape of widespread misinformation, a schism between scientific consensus and public understanding of science topics is growing. Osborne et al. (2022) stated, "The threat to science from this new facility to disseminate misinformation so readily is, we argue, akin to the challenge posed by the launch of Sputnik in 1957" (p. 248). For example, although most scientists (88%) believe that genetically modified foods (GMFs) are safe to consume, the majority of the public disagrees (i.e., only 37% believe they are safe, Funk & Kennedy, 2016). Similarly, although nearly all scientists (97%) support the claim that climate change is real and human-caused, only 20% of Americans understand the strength of the scientific consensus (Leiserowitz et al., 2023). Thus, more research is needed to determine how best to push back on misinformation in these online contexts.

Scientific misinformation, malinformation, and disinformation

Misinformation has been defined as "any information that turns out to be false" (Ecker et al., 2022, p. 13). Other definitions include information that is, "initially processed as valid but that is subsequently retracted or corrected" (Lewandowsky et al., 2012, p. 124). Misinformation can range from innocuous, such as the controversy around classifying tomatoes as vegetables, to dangerous, such as the widely debunked claim that vaccines cause autism (DeStefano & Shimabukuro, 2019). Due to the vast amount of misinformation circulating online, individuals are likely to encounter multiple and conflicting accounts of scientific information. For example, searching Google with the phrase "is fluoridated

CONTACT Robert W. Danielson robert.danielson@wsu.edu College of Education, Washington State University, Pullman, WA.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/00461520.2024.2365628.
*Current affiliation: College of Education, Northern Arizona University, Flagstaff, AZ, USA

water safe" generates mixed responses. Links to the Centers for Disease Control and Prevention (CDC) and American Dental Association (ADA) support fluoride's safety. However, a link to a Time Magazine article entitled "Is Fluoride in Water Safe? A New Study Reignites the Debate" includes frightening information on skeletal fluorosis. This leaves many readers confused, skeptical, and misled, and this confusion can be amplified by using different search terms. For example, searching for the "truth" (for example, "is fluoride safe?") returns very different results than a query on the "controversy," which generates multiple conflicting points. It is not our position that these searches should be "censored" or "fact-checked" in real-time, but we do want to highlight the challenges that learners often encounter in their search for reliable information.

Another classic example is the discussion around the safety of vaccinations that has lasted for decades, with traditional public health education efforts failing to change minds. Although the World Health Organization (WHO) highlights that the measles vaccine has averted approximately 56 million deaths in the last twenty years, the percentage of children vaccinated for measles has declined significantly post-pandemic due to the proliferation of online misinformation about vaccines (WHO, 2023).

What can be done to confront the scourge of scientific misinformation that may cause significant harm? Some researchers have suggested that refutation texts may be a viable approach to combating misinformation (Schroeder & Kucera, 2022). Across multiple reviews (Guzzetti et al., 1993; Tippett, 2010) researchers have suggested that compared to expository (i.e., non-narrative informational) texts, refutation texts are an efficacious strategy for combating misconceptions. However, these findings have recently been met with skepticism (Zengilowski et al., 2021). Thus, the purpose of the current systematic meta-analytic review was to provide an up-to-date, comprehensive, and rigorous assessment of the efficacy of refutation texts compared to non-refutation texts.

What are refutation texts and how do they work?

Refutation texts typically contain three important components—first, they state the common but inaccurate knowledge that is assumed to be held by the reader; second, they explicitly indicate what is incorrect; and third, they provide the correct information, often with supporting explanations (see Jacobson et al., 2021; Kendeou et al., 2014). The advantage of refutation texts is thought to be derived from an explicit mention of assumed prior incorrect beliefs. This prompts the learner to recall their prior conceptions as they simultaneously process the new information (Tippett, 2010), allowing for a comparison between the two (Kendeou & O'Brien, 2014).

Evidence of the effectiveness of refutation texts has spanned decades (Guzzetti et al., 1993; Schroeder & Kucera, 2022; Tippett, 2010) and many educational topics. Refutation texts have been used for science topics such as force (Kendeou & Van Den Broek, 2007) and seasonal change (Cordova et al., 2014; Mason et al., 2017; Will et al., 2019), mathematics topics such as box plots (Lem et al., 2017), as well as socio-scientific topics including genetically modified foods (Heddy et al., 2017; Thacker et al., 2020; Trevors, 2016), climate change (Danielson et al., 2016), and vaccinations (Kessler et al., 2019; Trevors & Kendeou, 2020; Vaughn & Johnson, 2018). Recently, refutation texts have been successfully deployed to facilitate knowledge change in other domains, including history (Alongi et al., 2016; Donovan et al., 2018), public policy (Aguilar et al., 2019), immigration (Trevors, 2022), and dyslexia (Peltier et al., 2020).

Notably, Kozyreva et al. (2022) included refutations in their "toolbox of interventions against online misinformation and manipulation" (p. 1). This "toolbox" provides ten useful interventions for science communicators and practitioners to use in their fight against misinformation. Of these ten, refutations appear across 40% (i.e., second most prevalent) of these strategies. Kozyreva et al. (2022) claimed that refutation strategies can be used to: debunk misconceptions, inoculate against misconceptions, rebut science denialism, and/or as warning or fact-checking labels. Because these functions are not mutually exclusive (i.e., a refutation could debunk a misconception for one individual and act as an inoculation for another), scientists have recommended policymakers and science communicators use refutations when communicating with the general public.

Theoretical frameworks supporting refutation texts

We highlight three overarching theoretical frameworks that help explain the potential efficacy of refutation texts for addressing misconceptions. These frameworks include theories of conceptual change (Dole & Sinatra, 1998), the Knowledge Revision Components Framework (KReC, Kendeou & O'Brien, 2014), and the "medical" analogy of prebunking/debunking (Ecker et al., 2022; Lewandowsky et al., 2020). All three theories offer different explanations for why refutations may be effective. Next, we describe these theories and their differential explanations for the effectiveness of refutations, for whom, and under what circumstances.

Many researchers have drawn on a conceptual change framework to both explore and explain the efficacy of refutation texts. Theories around conceptual change have a rich and varied history (see Posner et al., 1982; Sinatra, 2022; Treagust & Duit, 2008, for example), and it is not within the scope of the present work to review this history. However, it is important to note that conceptual change researchers have used refutation texts as an intervention to overcome misconceptions (Heddy et al., 2017) and to facilitate conceptual change (Thacker et al., 2020). Some models of conceptual change posit that a fundamental element of this change is a *noticeable dissatisfaction* with one's current conceptual understanding (Posner et al., 1982). This perspective of conceptual change suggests that a refutation text should have an advantage over alternative types of texts because the refutation statement explicitly aims to create this dissatisfaction by juxtaposing the readers' conceptual (mis)understanding against an alternative conception that is typically scientifically valid.

Early researchers suggested that these conceptual shifts can be brought about rationally (Pintrich et al., 1993). A classic example is when a key aspect of a theory is undermined by new evidence, causing individual scientists to experience dissatisfaction with their current theories. These scientists would then abandon their old approach in favor of a new theory. For example, a person who is updating their understanding of topics like Newtonian Physics may do so rationally. That is, they are dissatisfied with their current conception, and in turn, they update their understanding without necessarily experiencing emotional turmoil in the process (described as "cold conceptual change," Pintrich et al., 1993). However, more contemporary views (Sinatra, 2005) have suggested that this dissatisfaction is inherently social and/or motivational (e.g., when a respected friend changes their stance on climate change, prompting you to consider new evidence). In this paradigm of conceptual change (see Dole & Sinatra, 1998), emotions and attitudes can play a central role in facilitating or inhibiting this change process. The extent of this dissatisfaction may be moderated by attitudinal values (e.g., a person who holds a negative attitude toward scientists may be less likely to revise their knowledge of climate change, Sinatra, 2005). In line with these conceptual change frameworks, we examined the role of both text characteristics (e.g., text topic) and reader characteristics (e.g., attitudes toward science) as potentially important moderators. In more recent work, researchers have examined the extent to which refutation texts can change more than prior incorrect beliefs. Some researchers have shown that in addition to shifting incorrect prior beliefs, refutation texts can shift attitudes (Heddy et al., 2017; Thacker et al., 2020) and emotions (Broughton et al., 2010).

Another relevant theoretical perspective is the Knowledge Revision Components Framework (KReC), posited by Kendeou and colleagues (Kendeou et al., 2014; Kendeou & O'Brien, 2014; Lassonde et al., 2016). The KReC framework provides a theoretical explanation of how prior knowledge is systematically restructured. There are several key components of the framework that are relevant here. First, once information is *encoded* in long-term memory, it stays in memory. Second, information in long-term memory can be *reactivated* through a learner's experience, such as reading a text. When two pieces of information are read in close proximity to one another within the text, both of those ideas may be active in working memory at the same time, or *co-activated*. Knowledge revision is said to be more likely when two ideas are *competing for activation* and one is more likely to be activated in the future (Kendeou et al., 2014, 2019; Will et al., 2019). The KReC framework provides a sound theoretical basis for the refutation text effect. As a misconception is activated and then refuted, it competes with the new explanation allowing the scientifically correct information the opportunity to "win" the activation competition. The KReC framework informs our approach in that the characteristics of the texts (e.g., structural cohesion) and the readers (e.g., prior knowledge) were included as moderators to elucidate the relative import of these factors as change facilitators.

Finally, a "medical" analogy has been suggested by Ecker et al. (2022), who distinguished between two frequently tested approaches to confronting misinformation: prebunking and debunking. Prebunking strategies are intended to warn readers about potential future exposure to misinformation that they may not have experienced, whereas debunking strategies are employed on the assumption that individuals have already been exposed to the misinformation and may believe it to be accurate. Prebunking is based on the notion that, much like a vaccination, individuals can be "inoculated" against misinformation by exposing them a priori to a weakened form of the misleading or incorrect information. In this approach, individuals are warned that they may soon hear incorrect information and then are given the correct information. For example, individuals could be told, *there are many false claims circulating about vaccine side effects, so in the coming weeks, you may hear claims that the COVID-19 vaccine causes infertility. However, this is not correct. There is no evidence of a link between vaccinations and infertility.* In addition to providing the warning and the correct information, prebunking communications may also provide instructive information about techniques commonly used by purveyors of misinformation, such as cherry-picking data or using non-experts who claim to have relevant expertise.

In contrast, debunking techniques are meant to "treat" those already exposed to misinformation (see Lewandowsky et al., 2020). Refutation texts are traditionally seen as a form of debunking because they are employed post hoc and in response to specific misinformation already in circulation and often widely believed to be true. Typically, debunking directly corrects the misinformation by stating that it is untrue, incorrect, false, or misleading, and then introducing the correct claim. Often, counter evidence is provided in support of the veracity of the alternative claim. This assumes that most readers hold misinformation. If they do not, refutations could be considered a form of prebunking.

A key factor in refutation text research is the background knowledge of the learner (i.e., researchers often use age and grade level of the reader as a proxy) because those with more background knowledge are less likely to hold misconceptions. However, those with more background knowledge are also expected to be less likely to shift their views when they do hold misconceptions as they are more committed to these prior conceptions (Sinatra & Mason, 2013). The prebunking/debunking framework also highlights an important consideration: the conditions in which the research is conducted. Some of this research is done in controlled lab settings whereas some is done in "the wild" in online free reading settings. The characteristics of the experimental conditions (e.g., lab-based experimental design or online free-reading) provide insights into the relative import of moderators as change facilitators, given the potential differences in attention between controlled and online settings (Creswell & Creswell, 2017).

In sum, three theoretical frameworks are relevant and inform our methodological approach to this meta-analysis. First, perspectives on conceptual change suggest that refutation texts can help readers juxtapose their prior beliefs against scientifically aligned ideas, thus creating dissatisfaction between their current concept and the new concept, giving them the opportunity to compare the two. Conceptual change perspectives suggest that readers' attitudes, motivations, and other characteristics could influence whether readers accept the new conception. In addition, the "hot" perspective of conceptual change would suggest that this change could shift attitudes and emotions as well as beliefs (Heddy et al., 2017; Thacker et al., 2020). Second, the KReC framework details how the battle of ideas is metaphorically fought and won within the constraints of working memory limitations. The KReC framework suggests that affordances around working memory (prior knowledge of the learner or textual cohesion) could influence the probability of learners updating their knowledge. Third, the prebunking/debunking literature suggests that readers can be metaphorically inoculated prior to exposure or "treated" for misinformation after the fact. It also suggests that background demographics of the learners, as well as conditions for learning, could determine the extent to which learners update their understanding. All three provide distinct, yet compellingly similar theoretical explanations of how refutation texts could help readers confront scientific misconceptions. Further, these theories provide the theoretical rationale for including specific text features and reader characteristics as moderators in our analysis.

Critiques of refutation texts

Regardless of theoretical positioning, refutation materials have been employed to increase readers' acceptance of scientific information or to promote conceptual change (Guzzetti et al., 1993; Sinatra & Broughton, 2011; Tippett, 2010). Research investigating the effects of refutation texts has been conducted in educational contexts (e.g., K-12 through post-secondary) in a wide range of domains, especially in the sciences (Schroeder & Kucera, 2022). Figure 1 illustrates that scholarly interest in the effects of refutation texts have grown significantly since 1981. Given that this theoretical scrutiny has also increased, now we turn to some theoretical and methodological critiques of employing refutation texts.

Theoretical critiques of refutation texts

The aforementioned frameworks lend theoretical support for the advantage of refutation texts over their expository counterparts. However, they also suggest possible critiques of refutation texts as a method to promote change. One noted challenge from the conceptual change perspective is that of granularity. Although change can be reliably produced around discreet propositions (e.g., the safety of GMOs, the greenhouse effect), most conceptual change researchers have argued that more deeply rooted and embedded concepts (e.g., the nature of science) may require longer, more interactive interventions than a single text could promote (see Lombardi et al., 2022; Sinatra, 2022; Vosniadou, 2008; for examples of larger interventions).

If the targeted misconception is foundational to a reader's misconceived knowledge, refutations may promote conceptual change (Danielson et al., 2016). However, because not all targeted misconceptions are foundational to conceptual knowledge, conceptual change may not occur, even if the refutation successfully upends the

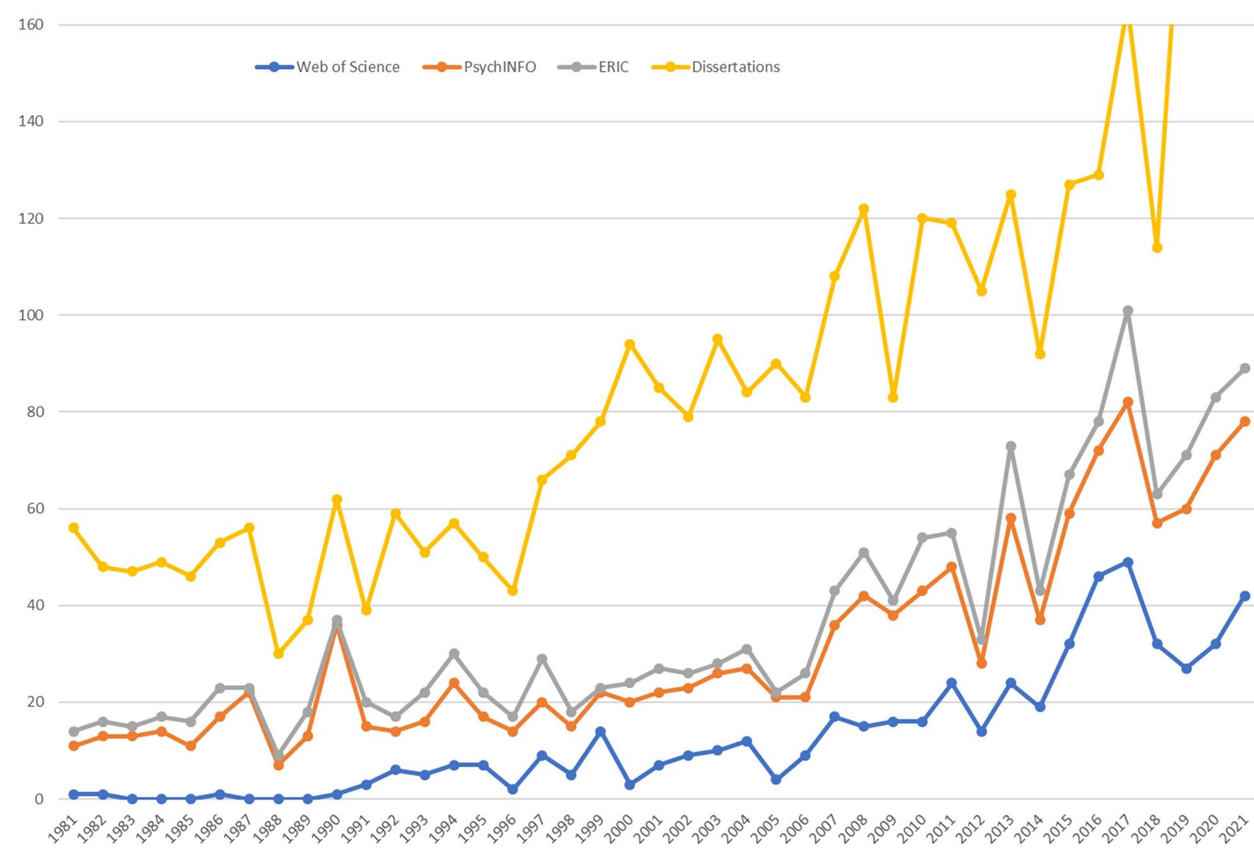

Figure 1. Refutation text publication trend (1981–2021).

misconception. For example, changing misconceptions about the mechanisms causing climate change (e.g., the greenhouse effect) may change a reader's conceptual understanding of how global warming works (Danielson et al., 2016; Ranney & Clark, 2016). In contrast, correcting misconceptions about COVID-19 vaccines causing infertility may not change a reader's conceptual knowledge about virology, infertility, or related concepts if the reader does not connect this to their conceptual understanding and instead perceives the misinformation as only an isolated fact.

Some have argued that perhaps the structure or style of refutation texts explains their effectiveness, rather than the refutation itself. By providing a causal explanation of the phenomenon in question, the KReC perspective suggests that the refutation structure is more coherent because the text affords greater argument overlap (Kendeou & O'Brien, 2014). Sinatra and Broughton (2011) noted that refutation texts have a greater "causal explanatory style" than typical expository texts, potentially leading to a more "coherent" text. This increased coherence may also promote surprise and allow readers to notice their prior knowledge is incorrect. These features may lead to interest and curiosity, which may in turn prompt active engagement with new information (D'Mello et al., 2014; Munnich & Ranney, 2019; Ranney & Clark, 2016; Thagard, 1989; Vogl et al., 2019). This active engagement is a key predictor of conceptual change (Dole & Sinatra, 1998). Thus, perhaps it is the text features rather than the refutation itself that provides benefits. When learners are presented with new information, researchers may assume that any discrepancies between their prior conceptions and the incoming information will be noticed. Thus, refutation texts must be carefully constructed to promote such evaluation. Along these lines, this synthesis sought to examine whether stylistic features of the texts predict variation in refutation text effects.

From the debunking perspective, one major concern is that refutations may "backfire." The backfire effect occurs when readers double down on their acceptance of misinformation targeted by the attempted correction (Ecker, 2017; Ecker et al., 2020; Jacobson et al., 2021). Lewandowsky et al. (2012) initially proposed three "backfire" effects, which include the "overkill," "familiarity," and "worldview" effects. The overkill effect arises when the refutation is overly complicated with excessive counterarguments. The familiarity effect arises when the presentation of a refutation or correction itself boosts the familiarity of the targeted misconception. In a worst-case scenario, if an individual did not initially hold the targeted misconception, the refutation text may call attention to misconceptions that were not otherwise known. This initial exposure could increase acceptance of the misconception. In the worldview effect, learners may resist and argue against correcting misconceptions that are supported by core identity values. Examples include teachers resisting reform efforts that threaten their core values, or students struggling to reconcile biological evolution with their religious worldviews.

Concerns about backfire effects were common in earlier work (Ecker et al., 2010; Schwarz et al., 2007). However, Lewandowsky et al. (2012) noted that the level of concern over backfire effects is unwarranted, as evidence suggests that these effects are rare and small. Our meta-analysis may further elucidate whether these effects are present as backfire effects would reduce our ability to detect the refutation text effect.

Methodological critiques of refutation texts

Typically, researchers conducting a refutation text study might administer a knowledge pretest, then randomly assign participants to either a treatment (i.e., refutation) or control (i.e., expository) text condition, and then administer a knowledge posttest. The knowledge pretest allows researchers to identify misconceptions that participants may hold, and the posttest allows researchers to identify whether the intervention was effective. This experimental design also allows researchers to empirically test whether participants in both conditions have comparable levels of misconceptions before the intervention and allows them to assess the reduction of misconceptions post intervention. This experimental design is so common that Zengilowski et al. (2021) noted only 14% of their reviewed articles used posttest only (i.e., no pretest) designs. However, if the co-activation of the misconception facilitates conceptual change (Kendeou et al., 2014; Kendeou, 2024), then it is possible that the pretest could facilitate this co-activation. There is little empirical evidence to directly address this concern. However, a recent direct test of this hypothesis utilizing a Solomon Four Group Design (Jin et al., under review) revealed that even though there was an advantage for individuals who received a pretest, refutation texts still increased accurate beliefs and reduced inaccurate beliefs over expository text. Given this concern, in this synthesis we explored the refutation text effects from independent groups designs with and without pretests separately in order to determine if this design feature predicts different results.

Another concern raised by Zengilowski et al. (2021) was that the science topics investigated with refutation texts were limited in scope. For example, Newtonian physics, genetically modified foods, evolution, energy, seasonal change, and vaccines accounted for nearly half of their reviewed articles. Although these texts differed in content, Zengilowski et al. cautioned against overgeneralizing findings from these topics to the refutation literature in general. Given this concern, we explored the extent to which science discipline moderated the refutation text effect.

Despite these concerns, a recent meta-analysis of 33 studies examining the effectiveness of refutation texts revealed a moderate effect size in favor of refutation texts (Schroeder & Kucera, 2022). Schroeder and Kucera (2022) also examined several moderators, including whether the effect was moderated by text topic, and found no differences in the effectiveness of refutation texts among science, mathematics, or social science topics. Refutation texts were found to be moderately effective across all topics. They also found that the refutation text effect held up against other text and reader moderators including text length, text format (i.e., whether images were included or not), text medium (i.e., print versus electronic), the age of participants, and other factors. However, that meta-analysis included only 33 studies, with limited power for tests of moderation.

Thus, research on refutation text is at a crossroads. Although empirical evidence suggests that refutation texts are effective in confronting misconceptions, several theoretical and methodological issues remain. Previous meta-analyses failed to fully resolve these issues or to adequately test moderators for these effects. Therefore, we conducted a rigorous test of the refutation text effect to discover whether they are effective, for whom and under what circumstances.

Do refutation texts work: For whom and under what conditions?

Building on the efforts of previous meta-analyses and reviews (Kozyreva et al., 2022; Schroeder & Kucera, 2022; Zengilowski et al., 2021), we aimed to comprehensively examine the effects of refutation texts. The primary goal of this meta-analysis was to investigate the overall impact of refutation texts on learning outcomes (i.e., accurate and inaccurate beliefs) to see if our findings aligned with previous meta-analyses and systematic reviews (e.g., Guzzetti et al., 1993; Schroeder & Kucera, 2022; Tippett, 2010). Then we extended our examination of refutation texts to other outcomes examined in prior research (e.g., positive/negative emotions and attitudes; Dole & Sinatra, 1998; Pintrich et al., 1993; Sinatra, 2005) that have not been included in prior meta-analyses (Schroeder & Kucera, 2022). Given the literature reviewed above, we expected to find that refutation texts would increase accurate beliefs and decrease inaccurate beliefs. We also expected to find that refutation texts would increase positive attitudes and emotions and reduce negative attitudes and emotions. This prediction was based on both theory (Dole & Sinatra, 1998) and previous empirical studies (Heddy et al., 2017; Thacker et al., 2020).

Finally, we examined the largest set of moderators to date. With more studies now available, we were able to examine theoretical and practical moderators related to text features, as well as characteristics of the design, samples, settings, and topics. In line with our pre-registration of this research, hypotheses for each moderator were designated as either confirmatory—that is, testing a strong theory-based hypothesis, or exploratory—that is, examining the moderator was expected to be informative for guiding existing theories and there was only limited (or no) theoretical guidance for making a prediction. We describe these moderators in more detail below. In addition, Table 1 describes all the information that was retrieved about studies in the synthesis, in particular, how moderators were coded and categorized.

Text characteristics that may influence the refutation effect

We examined whether features of the texts may influence their effectiveness in addressing misconceptions. These factors included the discipline and topic of the text, as well as relative differences between, and average levels of, the stylistic features and length of the texts.

Text discipline and topic. As noted by Zengilowski et al. (2021), most refutation studies examined a narrow set of science topics, creating concerns about the generalizability of the refutation text approach across disciplines and topics. We sought to break down science from a monolithic construct into its sub-disciplines. We compared four sub-disciplines of science—social science, geoscience, life science, and physics. Although we had no *a priori* reason to believe that differences across science disciplines *should* be present, this is an empirically important question to explore given concerns about generalizability across disciplines. Furthermore, we explored whether refutation text effects varied depending on whether the topic was or was not a socio-scientific topic. According to the definition provided by Sadler (2004), socio-scientific topics are those which "encompass social dilemmas with conceptual or technological links to science" (p. 1) and are often seen as controversial (e.g., evolution/natural selection, climate change, vaccinations, etc.). Though we had little theory or prior research findings to guide our predictions, we hypothesized (exploratory) that the refutation effect would be smaller for controversial topics vs. non-controversial topics because individual difference characteristics (e.g., political beliefs, religious beliefs, etc.) might make controversial topics more resistant to change.

Text length and style. Many, but not all, studies have attempted to control the relative differences between the refutation text and the control text. This often includes balancing the text length or written style. For example, if the control text is approximately 1000 words in length, many researchers have aimed to keep the reputational text within a 20% range (about 850–1150 words) in their studies. This is important because the relative difference in length could potentially confound the results. Given this possibility, we explored whether the relative differences predicted variation in refutation text effects.

As previously discussed, desirable stylistic features of the texts may also play a role in refutation text effects. We examined the following measures of text structure: referential cohesion (i.e., "ideas that overlap across sentences and the entire text, forming explicit threads that connect the text for the reader," McNamara & Graesser, 2012, p. 17), causal cohesion (i.e., "text…contains causal, intentional, and temporal connectives" McNamara & Graesser, 2012, p. 18), narrativity (i.e., "text tells a story, with characters, events, places, and things that are familiar to the reader" McNamara & Graesser, 2012, p. 17), syntactic simplicity (i.e., "the degree to which sentences in the text contain fewer words and use more simple, familiar syntactic structures, which are less challenging to process" McNamara & Graesser, 2012, p. 18), and concreteness (i.e., "texts that contain content words that are concrete, meaningful, and evoke mental images," McNamara & Graesser, 2012, p. 18). Based on KReC, we hypothesized (confirmatory) that when the refutation text was either *relatively*- or

Table 1. Description of information retrieved from studies and moderator variables used for analysis.

Code categories	Description	Moderator categories/ variable for analysis
Research report characteristics		
Author/year	Citation information for report/study	
Type of report	Journal article, book/chapter, dissertation, thesis, private report, government report, conference paper, other report	Published (includes journal article, book/chapter), unpublished (all else)
Text characteristics		
Text subject	Chemistry, computer and information science and engineering, engineering, geosciences, life sciences, materials research, mathematical sciences, physics and astronomy, psychology, social sciences, STEM education and learning research	Chemistry, Geosciences, life sciences, physics, and applied physics (includes physics, astronomy, engineering), social sciences (includes psychology, STEM education and learning research, social science)
Text topic	Astronomy and astrophysics, economics, evolutionary biology, genetics, public policy, particle physics, physical oceanography, theoretical physics, ecology, history and philosophy of science, mathematics education, physical and dynamic meteorology, climate and large-scale atmospheric dynamics, energy, biomedical, science education, electrical and electronic engineering, zoology, microbiology, social psychology, cognitive psychology, acids and bases, perception, and psychophysics	Controversial (includes history and philosophy of Science, climate, and large-scale atmospheric dynamics, biomedical, economics, public policy, science education, zoology, social psychology, cognitive Psychology, evolutionary biology, genetics), not controversial (includes astronomy and astrophysics, ecology. mathematics education, physical and dynamic meteorology, energy, electrical and electronic, microbiology, acids & bases, perception & psychophysics, particle physics, physical oceanography, theoretical physics)
Text length	Length (in words) of refutation and control text.	Relative: Continuous moderator variable created by taking the difference between refutation text referential cohesion score and control text lengths.
Text referential cohesion[a]	Rating of extent to which refutation and control text contains words and ideas that overlap across sentences and the entire text, forming explicit threads that connect the text for the reader as reported by Coh-Metrix	Relative: Continuous moderator variable created by taking the difference between refutation text referential cohesion score and control text score. Mean: Continuous moderator using the refutation text score using only studies where referential cohesion was approximately equal across texts (difference in scores is <10%).
Text causal cohesion[b]	Rating of the extent to which refutation and control text contains connectives that help the reader form a more coherent and deeper understanding of the causal events, processes, and actions in the text as reported by Coh-Metrix.	Relative: Continuous moderator variable created by taking the difference between refutation text causal cohesion score and control text score. Mean: Continuous moderator using the refutation text score using only studies where causal cohesion was approximately equal across texts (difference in scores is <10%).
Text narrativity[b]	Rating of the extent to which refutation and control text tells a story with characters, events, places, and things that are familiar to the reader as reported by Coh-Metrix	Relative: Continuous moderator variable created by taking the difference between refutation text narrativity score and control text score. Mean: Continuous moderator using the refutation text score using only studies where narrativity was approximately equal across texts (difference in scores is <10%).
Text syntactic simplicity[b]	Rating of the extent to which the sentences contain fewer words and use simpler, familiar syntactic structures as reported by Coh-Metrix	Relative: Continuous moderator variable created by taking the difference between refutation text syntactic simplicity score and control text score. Mean: Continuous moderator using the refutation text score using only studies where syntactic simplicity was approximately equal across texts (difference in scores is <10%).
Text word concreteness[b]	Rating of the extent to which refutation and control texts contain content words that are concrete, meaningful, and evoke mental images are easier to process and understand as reported by Coh-Metrix	Relative: Continuous moderator variable created by taking the difference between refutation text word concreteness score and control text score. Mean: Continuous moderator using the refutation text score using only studies where word concreteness was approximately equal across texts (difference in scores is <10%).
Design characteristics		
Random assignment	Random, nonrandom	Random, nonrandom
Comparison condition	Expository, one-sided persuasive, non-refutational (two-sided), narrative, repetition text, other	One-sided non-refutational (includes expository, one-sided persuasive, narrative text), two-sided non-refutational
Sample characteristics		
Belief screening	Inaccurate beliefs screened and sampled, accurate beliefs screened and sampled, beliefs not screened	Sample excludes individuals with accurate beliefs, sample does not exclude individuals with accurate beliefs
Current education level	Elementary (1st–5th), middle/high school (6th–12th), college	Elementary, middle/high school, college
Gender	Percentage of female students	Continuous variable
Ethnicity	Ethnic composition and percentage of each ethnicity reported	Continuous variable; insufficient variability for analysis
Setting characteristics		
Educational setting	Content embedded in curriculum, not embedded	Embedded with authority, Embedded without authority, Not embedded with authority, Not embedded without authority
Authority setting	Person in authority present, authority not present	
Region: Continent/Country of study	USA, non-USA, specified country	North America (includes USA, Canada), Europe (includes Finland, Belgium, Germany, Spain, France, Netherlands, Cyprus, Spain, other (all else)
Country science trust	Specified country	Country science trust[b]: High science trust (includes Australia, Finland) vs. low science trust (includes USA., Canada, Turkey, Indonesia, Belgium, Israel, Germany, Spain, France, Netherlands, Cyprus, Italy)
Outcome characteristics		
Outcome type	Accurate beliefs, inaccurate beliefs, positive attitudes, negative attitudes, positive emotions, negative emotions, other motivation (specified)	Accurate beliefs, inaccurate beliefs, positive emotions (includes positive affect, topic interest, task intrinsic motivation), negative emotions, positive attitudes (includes negative attitudes with sign flipped on effect)
Question type	Multiple choice, true/false, explanation, Likert scale, essay, short-answer, free recall, recognition	Open (includes explanation, essay, short answer, free recall, and any combination of these four), closed (multiple choice, true and false, Likert scale and any combination of these three), includes both open and closed questions
Outcome reliability	Cronbach's alpha	Continuous variable
Timing of outcome	Time between reading text and completion of outcome measure	Continuous variable
Effect size information		
Effect size information	Standardized mean difference (Hedges' *g*), group means, group standard deviations, group sample sizes, and inferential statistics (for pre- and post-intervention measures as available)	

[a]Coh-metrix indices described in more detail in Graesser et al., 2004, 2011.
[b]Countries were categorized as low and high trust based on survey findings of 140,000 people from 140 counties of the Wellcome Global Monitor (see Rabesandratana, 2019; https://www.science.org/content/article/global-survey-finds-strong-support-scientists).

absolutely- more cohesive or concrete than the control text, the refutation effect would increase.

Design, sample, setting, and outcome features that may influence the refutation effect

Studies also varied in the characteristics of their designs, sample, settings, and outcomes. We found strong theoretical or methodological reasons for examining whether the participants' prior levels of misconceptions, the educational setting, the type of questions in the outcome measure, and outcome measure reliability predicted variation in refutation effects based on confirmatory hypotheses. We also felt there were important practical reasons for exploring the extent to which the nature of the comparison text, how participants were assigned to condition, participants' educational level, participants' gender, the geographic region, the county's overall trust in science, the nature of the outcome variable, and the length of delay in the outcome measure predicted variation in refutation effects.

Comparison condition and assignment to condition. Both Guzzetti et al. (1993) and Tippett (2010) discussed the efficacy of refutation texts over standard expository texts. However, researchers have examined other forms of comparison texts, including persuasive texts presenting either a single or dual (i.e., presenting both sides of an issue) perspective (Buehl et al., 2001). We hypothesized (exploratory) that the refutation text effect would be larger when compared to one-sided and smaller when compared to two-sided non-refutational texts based on previous research showing a stronger advantage for refutations for single-sided texts (Thacker et al., 2020). Moreover, given that studies varied in their experimental approach and studies using random assignment typically reveal smaller effects (St. Pierre, 2001), we also thought it was important to explore whether the use of random assignment to condition or a nonrandom assignment procedure explained variation in the effects and to control for this core design feature when examining other moderators.

Sample characteristics. A common feature of some studies has been to include learners who were prescreened for holding relevant misconceptions. Drawing on all three frameworks (i.e., conceptual change, KReC, and debunking), we hypothesized (confirmatory) that the refutation text effect would be larger when participants' beliefs were screened to exclude individuals with high levels of accurate beliefs. Given that screening may guard against a ceiling effect, excluding individuals with correct conceptions may increase the power of refutation texts to shift beliefs. That is, refutation texts are more likely to be effective if readers hold the targeted misconceptions prior to reading. For practical reasons we also explored whether students' educational level may predict variation in refutation text effects. Though prior theory (Dole & Sinatra, 1998) and empirical findings (Kardash & Scholes, 1996) were limited for guiding our predictions, we hypothesized (exploratory) that as students' age and educational level increased, the refutation effect would decrease because older students may hold more entrenched beliefs that are more resistant to change (Sinatra & Mason, 2013). We were also interested in understanding the extent to which participants' gender (exploratory) might predict variation in refutation text effects. To our knowledge, no meta-analysis has examined whether there is variation in refutation text effects by gender and we had no hypothesis related to gender.

Setting. We also examined whether aspects of the setting predicted variation in refutation text effects. Some studies were designed such that exposure to the refutation texts were embedded in the curriculum of an authentic educational setting (e.g., in the curriculum of a real course conducted in the classroom), with an on-site facilitator. In contrast, other studies took place in laboratories or completely online with no facilitator. We hypothesized that the refutation text effect would be largest when the study was embedded in the curriculum of an authentic educational setting with a facilitator present. Likewise, the refutation text effect would be smallest when the study was not embedded and a facilitator was not present, with the size of effects of studies with one feature or the other falling in between the two extremes. We predicted this given that prior research suggests settings can influence effect sizes (Kraft, 2020). Beyond this core setting feature, we thought it was practically important to explore whether geographic region (exploratory) and a country's overall trust in science (exploratory) predicted variation in refutation text effects. To our knowledge, no meta-analysis has explored whether a country's overall trust in science predicts variation in refutation text effects. We had no hypothesis related to the region, but we expected that as a country's overall trust in science increased, the effect of the refutation text would also increase (Wellcome, 2018; World Economic Forum, 2020). The largest effects were expected in high trust countries (i.e., Australia and Northern Europe). Smaller effects were expected for studies conducted in Northern America, Middle East, Western Europe, Southern Europe, and Eastern Europe.

Outcome characteristics. Previous refutation studies have employed a wide range of outcome measures to determine the effectiveness of refutation texts. As previously discussed, at the broadest level, we expected to find that refutation texts would increase accurate beliefs, emotions, and positive attitudes and would decrease inaccurate beliefs. We also examined whether effects varied depending on whether the outcome was accurate beliefs increasing or inaccurate beliefs decreasing, although we did not expect to find a difference. Nonetheless, we felt it was important to explore whether the nature of beliefs explained variation in refutation text effects, and further, control for this covariate in analyses of other moderators. We also examined whether the reliability and nature of the outcome measure questions explained variation in the refutation text effects. Specifically, we hypothesized (confirmatory) that the refutation text effect would be larger when measured using open-ended questions (e.g., essays, short-answers) compared to closed-ended questions (e.g., multiple choice, true/false) because open-ended questions traditionally measure recall, deeper conceptual knowledge, and limit guessing, whereas close-ended questions traditionally measure recognition, factual knowledge, and afford participants the opportunity to guess the correct answer (Sychev et al., 2020). Moreover, we predicted that stronger effects would be revealed when relatively more reliable outcome measures were used given that low measurement reliability can attenuate effects (Allen, 2017). From a conceptual change perspective, we also hypothesized (exploratory) that the refutation text advantage would remain even if there was a delay in the measurement of the outcome over time. For example, Paynter et al. (2019) attempted to debunk myths around autism with a population comprised of paraprofessionals, behavior analysts, teachers, and speech pathologists who had various levels of knowledge about autism. The refutation text was effective in confronting this myth when measured immediately after reading, but at delay the effect faded even for those who were knowledgeable about autism. However, some studies have shown the refutation effect decays over time, but a statistically significant advantage remains in comparison to control conditions (Danielson et al., 2016), and Lombardi et al. (2013) found that the advantages of refutation texts remained six months after instruction.

Need for a new meta-analysis on effects of refutation text

Given that misconceptions about science are rampant online, it is crucial to evaluate the success of debunking those misconceptions through refutation texts. Although other reviews have made considerable efforts toward uncovering the effects of refutation texts, several issues remain. By thoroughly examining moderators and including several meta-analytic methodological advancements, our investigation elucidates the extant literature on refutational texts (Schroeder & Kucera, 2022; Zengilowski et al., 2021). Our meta-analysis contributes substantially to the extant literature in the following ways.

First, we conducted a comprehensive set of searches for refutation text studies, including unpublished manuscripts via listservs, robust personal communications with prolific scholars in the field, and funder archives. Therefore, we screened and included significantly more studies (i.e., more than double the studies included in the Schroeder & Kucera, 2022 meta-analysis), reducing concerns about publication biases and substantially increasing our ability to evaluate the effects of moderators. Second, our meta-analysis examined several moderators of interest to science and practitioner communities. These moderators included whether participants were screened for misconceptions prior to inclusion in the study, the educational authenticity of the experimental environment, the topic of study and whether this topic was seen as publicly controversial (climate change) or not (seasonal change), learner characteristics (gender) and country characteristics (general acceptance of science), and a more nuanced examination of textual effects (the relative differences of narrativity, length, cohesion, etc.).

Third, our meta-analysis utilized a more rigorous methodological approach. The most recent meta-analysis of refutation text effects (Schroeder & Kucera, 2022) used traditional random effects meta-analytic techniques that did not account for the dependency of effects within studies. Moreover, that meta-analysis did not focus on moderators that might have accounted for changes in beliefs and did not control for covariates in analyses. However, our meta-analysis extended the results of prior syntheses by computing both posttest only effects (i.e., effects based on independent groups [IG] with posttests only) and posttest effects adjusted by pretest scores (i.e., effects based on independent groups with repeated measures [IGRM] in which the outcome is measured before and after the refutation text intervention). This partially addresses comments by Zengilowski et al. (2021) around the lack of posttest-only designs and allows for comparison of results across designs. We also limited bias in our results by leveraging a random effects multilevel modeling approach with robust variance estimation to account for the multilevel nature and dependency of the data, and controlled for methodological covariates that are related to study quality or bias in effects in moderator models.

Fourth, our meta-analysis included “hot” outcome variables. These hot constructs, including interest, attitudes, and emotions, have long been considered essential for conceptual change (Dole & Sinatra, 1998; Pintrich et al., 1993; Posner et al., 1982; Sinatra, 2005). Despite this conceptual link between “hot” variables and refutation texts, previous meta-analyses focused solely on learning outcomes (Schroeder & Kucera, 2022). Finally, our research questions, hypotheses, protocol, and analysis approach were pre-registered, and provide robust objectivity and openness (Van den Akker et al., 2019).

Given these advances in our approach, the present meta-analysis advances understanding of refutation text effects. In sum, this work was guided by the following confirmatory and exploratory research questions.

RQ1. To what extent do refutation texts compared with non-refutation texts predict greater accurate and less inaccurate beliefs, as well as greater emotion (positive and negative) and more positive attitudes? Moreover, is there any evidence of publication bias in these effects?

RQ2A. To what extent can we confirm that characteristics of the text, particularly the relative stylistic characteristics of texts (e.g., cohesiveness, narrativity, simplicity, etc.), moderate the effect of refutation texts on beliefs even after accounting for key methodological covariates (type of belief, publication status, random assignment, use of belief screening)?

RQ2B. To what extent do we find in exploratory analyses that other characteristics of the text, including text discipline, controversial nature of the topic, relative text length, and the mean level of text stylistic characteristics (e.g., cohesiveness, narrativity, simplicity, etc.) moderate the effect of refutation texts on beliefs even after accounting for key methodological covariates?

RQ3A. To what extent can we confirm that design, sample, setting, and outcome characteristics, specifically whether there was screening for participants’ prior levels of misconceptions, the educational setting, the type of questions in the outcome

measure, and outcome measure reliability, moderate the effect of refutation texts on beliefs even after accounting for key methodological covariates?
RQ3B. To what extent do we find in exploratory analyses that other characteristics of the design, sample, setting, and outcome, including the nature of the comparison text, participants' educational level, participant's gender, the geographic region, the county's overall trust in science, and the length of delay in the outcome measure, moderate the effect of refutation texts on beliefs even after accounting for key methodological covariates?

Methods

We adopted well-established meta-analysis protocols (Borenstein et al., 2021; Lipsey & Wilson, 2001; Pustejovsky & Tipton, 2022; Viechtbauer, 2010). This section delineates our procedures for study inclusion criteria, data collection, search strategies, synthesis, and analyses. This meta-analysis was pre-registered at Open Science Framework (https://osf.io/8kfue/).

Study inclusion criteria

We began by reviewing several meta-analyses and literature reviews on the topic (Guzzetti et al., 1993, Tippett, 2010, etc.). Following this review, the research team developed eligibility criteria for studies investigating the effect of refutational texts to be included in the current research synthesis. Each study that we included was required to meet all of the following inclusion criteria.

1. The study contrasted the effects of a refutational or conceptual change text with a non-refutational control text condition (i.e., either one- or two-sided expository, persuasive, narrative text) focused on the same informational content as the refutation text.
2. The text focused on a science topic (e.g., Chemistry, Computer and Information Science Engineering, Engineering, Geosciences, Life Sciences, Materials Research, Mathematical Sciences, Physics and Astronomy, Psychology, Social Sciences, or STEM Education and Learning).
3. The study assessed and reported the effect of text type on post-intervention beliefs. If another outcome of interest, including emotion, attitudes, or motivation was assessed, it was also coded.
4. The study reported descriptive or inferential statistics to calculate effect sizes or researchers provided this information when contacted by email.
5. The study was conducted with a sample of "typical functioning students" (i.e., those without behavioral or emotional challenges). Few, if any, studies were conducted with students other than typically functioning students and limiting studies by sample reduced unexplained heterogeneity in the database.
6. Participants needed to be randomly or non-randomly assigned to separate refutation and control conditions and could not serve as their own control. This criteria allowed us to easily compare the results of our meta-analysis to prior ones (Schroeder & Kucera, 2022). Moreover, given the limited number of studies using within-person designs, excluding this design reduced unexplained heterogeneity in the database and eliminated the confounding effect of exposure.
7. The refutation text intervention could not include interactive instruction (i.e., studies that included constructive turn-taking dialogue with a teacher, student, or intelligent tutor) (Chi & Wylie, 2014). This criterion allowed us to easily compare the results of our meta-analysis to prior ones.
8. The report of the study was written in the English language.

We did not have eligibility criteria regarding the report or publication year, reporting outlet, or the publication status of the report, and we included both published and unpublished research.

Search strategies

Six different approaches were used to conduct comprehensive and systematic searches to locate all relevant studies that met the inclusion criteria. First, we conducted comprehensive and systematic searches involving the following electronic databases and search strategies: Google Scholar, PsycINFO, ERIC, ProQuest Dissertations & Theses Global, Sociological Abstracts; Worldwide Political Science Abstracts, Medline, Biology Database, Science Database, Biological Science Database. Table 2 shows search queries used on different databases:

Second, descendent searches were conducted using the Social Science Citation Index within the Web of Science for reports that cited any of the following seminal review or empirical reports (Alvermann & Hague, 1989; Dole, 2000; Dole & Sinatra, 1998; Guzetti, 2000; Guzetti et al., 1993, 1997; Hynd & Alvermann, 1986; Kendeou, 2005; Kendeou et al., 2013; Kendeou & van den Broek, 2007; Sinatra, 2005; Sinatra & Broughton, 2011; Tippett, 2010). Third, we contacted professional organizations through listserv emails to request additional relevant data. We posted to several research communities including the American Psychological Association's (APA) Division 15 Educational Psychology, the American Educational Research Association's (AERA) Division C Learning and Instruction, the Society for Text & Discourse Processes (ST&D), and two special interest groups (i.e., SIG-2 text and graphic comprehension, and SIG-3 conceptual change, from the European Association for Research on Learning and Instruction). We contacted research communities via Twitter that we could not contact via listserv including the Psychonomic Society, Cognitive Science Society (CSS), European Association for Research on Learning and Instruction (EARLI), Association for Psychological Science (APS), and Society for the Scientific Study of Reading (SSSR).

Fourth, we searched the award databases of funding agencies, including National Science Foundation (NSF), Institute for Education Sciences (IES), the National Institute of Health (NIH), Spencer Foundation, Bill and Melinda

Gates Foundation, The Ford Foundation, and The Alfred P. Sloan Foundation. We searched the terms "refutation," "refutation text," "conceptual change," "conceptual change text" in all databases. However, each funder database had a different structure for searching. For NSF, we specified the keyword location as "Title and Abstract." For IES, we searched "all words." Because the Spencer Foundation, Bill and Melinda Gates Foundation, NIH, Ford Foundation, and the Alfred P. Sloan Foundation did not allow specific keyword locations, we conducted a global search on those sites. PIs of relevant projects were contacted and asked to share relevant research reports.

Fifth, we conducted an ancestry search in which the reference sections of all articles that met inclusion criteria from all other search strategies were checked to recover additional studies not captured with the database searches. Finally, we directly contacted (by email) prolific authors, defined as any author whose name appeared twice in the included articles (regardless of authorship position) to request any additional or unpublished research that met our inclusion criteria.

Document retrieval, screening, and data extraction

We adopted two selection phases to determine whether articles returned by the searches should be included or excluded from the meta-analysis. In the first phase, two of four trained coders (Ph.D. students) read the titles and abstracts of 7,741 reports found in the search to examine their potential for inclusion by applying the selection criteria. Coders were trained in screening procedures before coding reports for inclusion. The second author and all coders met to discuss inclusion criteria and practice screening based on titles, abstracts, and full-text together. Coders then practiced coding a set of records independently, met to verify their coding, and were given additional records to practice screening. Once coders established 80% agreement with a lead researcher (the second author), they were given independent coding assignments, with two coders screening every record or full-text report. Reports that were selected for further exploration by either coder in the first phase were thoroughly examined for final inclusion in the second phase (full-text review). In the first phase, for cases in which the abstract did not provide sufficient information to include or exclude the report based on our selection criteria, the coders defaulted to advancing to full-text review to more rigorously examine potential for inclusion. In this phase (title and abstract screening), 927 records were retained.

In the second phase (full-text review), two coders reviewed the entire text of all 927 records retained in the first phase to determine whether any study met our predefined inclusion criteria using a short inclusion criteria coding form. Disagreement about whether a report included at least one study met inclusion criteria and should be retained was resolved through discussion. In cases where a joint decision could not be made by the PhD student coders, additional senior researchers on the team reviewed the abstracts and entire studies. Through these robust discussions, collective decisions were made. In several cases, the team read the method, procedure, and data collection sections of such borderline studies to obtain more information that helped us retain or exclude such papers. Of 927 records examined in the second phase, only 71 met all selection criteria and were included in our meta-analysis. Figure 2 shows a PRISMA flow diagram (Liberati et al., 2009) of how studies were advanced through the coding process, from search to title and abstract screening to full-text review to data extraction, as well as the reasons for excluding 856 studies after full-text reviews.

Like the screening process, coders worked together to ensure a rigorous coding process. This coding process was as follows. First, all reports that met inclusion criteria were double-coded by two independent coders (from a pool of 7 graduate student coders). Second, these codes were discussed to reach consensus across the two coders. Finally, this consensus was validated by a third expert coder. This process was followed for all studies included in the present meta-analysis, and the inter-rater agreement for the process of discussion, consensus, and validation was high (95%). To train coders on data extraction, the second author and all coders met together to discuss the meaning of all coding characteristics and rules for consistent coding. Coders then practiced independently coding 10 practice articles and met to verify their coding. Once coders established 80% agreement with a lead researcher (the second author), they were given coding assignments. The research team coded numerous different characteristics of each study. These characteristics encompass seven broad distinctions among studies: (a) the research report, (b) the research design, (c) the setting, (d) the sample, (e) the refutation text and control text characteristics, (f) the outcome measure, and (g) the estimate of the effect of refutation text on the outcome. See Table 1

Table 2. Search queries and databases searched.

Database:	(Medline, Biology Database, Science Database)
Search string:	all(concept* change")
Search string:	all(refutation* text")
Database:	Sociological Abstracts, PsycINFO, ProQuest Diss & Theses Global, ERIC
Search string:	all("conceptual* change")
Search string:	all("conceptual* change" OR misconceptions OR "knowledge revision" OR "science learning" OR "concept formation" OR "belief revision" OR "cognitive construction of knowledge" AND all("refutation* text" OR intervention OR "persuasive discourse")
Search string:	ab("refut* text*") AND AB(belief* OR knowledge* OR misconception*)
Database:	Google Scholar
Search string:	allintitle: "refutation text"

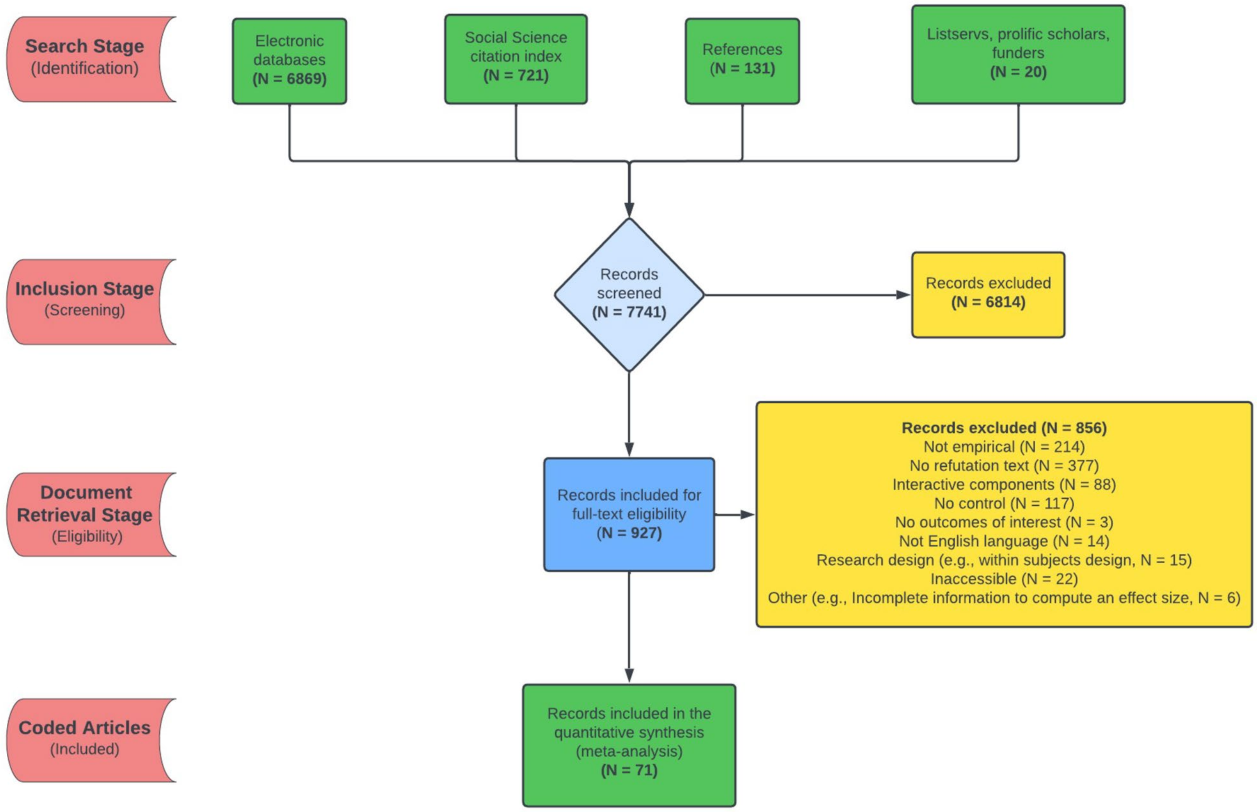

Figure 2. PRISMA flow diagram showing how studies were selected for the meta-analysis.

below for a list of all codes and how they were used in analyses, particularly for moderator analyses.

Computing effect sizes

We calculated effect sizes as standardized mean differences (SMD) on outcome measures between refutation text and control text groups. We calculated effect sizes directly from the means, standard deviations, and sample sizes for the intervention and control groups whenever possible. When effect sizes could not be calculated in this way, we computed them from F ratios, *t*-statistics, or chi-square statistics (see Lipsey & Wilson, 2001 for conversion formulas). For studies that included multiple refutation or control text conditions, we calculated the effect size for each comparison separately.

Many studies assessed an intervention's effect on student outcomes using an independent groups' (IG) posttest only design. However, others used an IG repeated measures (IGRM) design in which the outcome was measured before and after the text intervention. This required the use of an alternative formula that involves taking the difference between separate RM effects computed for intervention and control groups (Morris & DeShon, 2002). For studies reporting ANCOVA results, we used the equations in Borenstein (2009). If a study did not report the correlation between pre and post-test scores, we imputed a correlation of $r = 0.534$ that was found from another meta-analysis of the correlation between prior knowledge and learning (Simonsmeier et al., 2022). We converted all intervention effect sizes to bias corrected Hedge's *g*, a standardized effect size that corrects for a slight positive bias in effects present with small samples (Hedges, 1981). Regardless of the specific formula, a positive *g* value indicates that the outcome was greater in the refutation text condition compared to the control text condition.

We coded effects separately for studies with multiple samples or multiple outcomes. We computed effects in *R* either directly or using the *metafor* R package (Viechtbauer, 2010). Prior to analysis, we also examined the distribution of effect sizes to determine if any studies contained outliers using Tukey's (1977) definition where values more than 1.5 times the interquartile range from the quartiles were considered outliers. We found zero outliers among the IGRM effects. Among the IG effects, we found two outliers (with values of 2.52 and 2.83). We left outliers as is since the magnitude of the outliers could be explained by moderators in meta-regression.

Analysis strategy

We meta-analyzed refutation text intervention data separately using the *metafor* and *clubSandwich* R packages (Pustejovsky, 2019; Viechtbauer, 2010). We used random-effects modeling throughout the analyses. To account for the dependency between multiple effect size estimates within studies and samples and to guard against model misspecification, we adopted multilevel modeling approach in conjunction with robust variance estimator (RVE; Pustejovsky & Tipton, 2022). Specifically, we used correlated and hierarchical effects (CHE+) type working model for RVE, which entails both correlated and hierarchical dependency structures among effect size estimates, and we nested effects within subsamples within studies. The CHE + working model is appropriate to choose when there is little information about correlations between effect size estimates (Pustejovsky & Tipton, 2022). We assumed a correlation of r = .80 for effect sizes for multiple outcomes nested within a sample.

We fitted different random-effects models to estimate the pooled effect sizes for each outcome category (accurate beliefs, inaccurate beliefs, positive emotions, negative emotions, positive attitudes) separately for IGRM and IG effects. We also assessed the heterogeneity among effect sizes, indicated by Q, and τ^2. We reported 95% confidence intervals (CI) and 95% prediction intervals (PI) for each weighted average effect (Borenstein et al., 2021). For both CI and PI, we incorporated cluster-robust variance estimation (CRVE) for the standard errors and a small sample correction for the critical values (Tipton, 2015).

To further explain heterogeneity in the effect size estimates, we utilized mixed-effects meta-regression models for belief outcomes only, combining accurate and inaccurate beliefs by reversing the sign of inaccurate belief effects. We examined the effect of each moderator in univariate models that controlled for four covariates, (1) type of belief (accurate or inaccurate), (2) publication status (published or unpublished), (3) random versus nonrandom assignment to condition, and (4) whether participants were screened to exclude individuals with highly accurate beliefs or not.

Finally, we examined the possibility of publication bias and funnel plot asymmetry by creating funnel plots for IG and IGRM effects and each outcome separately and conducting a modified version of Egger's regression (Egger et al., 1997) for each outcome that accounted for the dependent effect sizes (Rodgers & Pustejovsky, 2020). We also conducted selection models as another selective reporting analysis approach where the probability of reporting is based on the p-value of the intervention effect (Vevea & Hedges, 1995). We used the *weightr* package in R (Coburn, 2018) which implements a step-function model. Currently, selection models assume independence among the effect sizes. Given that our data structure consists of dependent effects, there is likely inflated Type-I error rates and we only evaluate the adjusted mean estimates of the intervention effects. The results are included in our supplementary materials.

Results

We identified 71 total articles (53 published and 18 unpublished). These articles contained 76 studies, 111 samples, and 294 effect sizes (both IG and IGRM versions of effects were able to be computed in 292 cases, only IG effects could be computed in 274 cases, and only IGRM effects could be computed in 123 cases). Seventy-five studies with 237 effects focused on accurate beliefs as the outcome (both IG and IGRM versions of effects were able to be computed in 98 cases, only IG effects could be computed in 222 cases, and only IGRM effects could be computed in 113 cases). Ten studies with 15 effects focused on inaccurate beliefs (both IG and IGRM versions of effects were able to be computed in 9 cases, only IG effects could be computed in 14 cases, and only IGRM effects could be computed in 10 cases). In addition, 7 studies provided 13 IG effects focused on positive emotions, 7 studies provided 14 IG effects focused on negative emotions, and 7 studies provided 11 IG effects focused on attitudes. The authors, sample sizes, and effect sizes for these studies are listed in Table S1 and S2 of the supplemental materials, along with other study, sample, text, and outcome characteristics. Articles appeared between 1986 and 2021 and sample sizes ranged from 18 to 600, with the total sample size of 10,265.

Research question 1: overall effects of refutation texts

For our first research question we asked, to what extent do refutation texts compared with non-refutation texts predict greater accurate and less inaccurate beliefs, as well as greater emotion (positive and negative) and more positive attitudes? In line with the hypothesis that refutation texts enhance accurate beliefs, the overall mean weighted IG and IGRM effect sizes for accurate beliefs were in the small-to-medium range, g_{IG} (number of studies [k] = 67, number effect sizes [N_{ES}] = 222) = 0.37, p < .001 and g_{IGRM} (k = 49, N_{ES} = 113) = 0.36, p < .001. Likewise, in line with the hypothesis that refutation texts reduce inaccurate beliefs, the overall mean weighted IG and IGRM effect sizes for inaccurate beliefs were also in the small-to-medium range, g_{IG} (k = 10, N_{ES} = 14) = −0.38, p < .001 and g_{IGRM} (k = 6, N_{ES} = 10) = −0.31, p < .002. In contrast to our hypotheses, the effect of refutation texts on positive emotions (g_{IG}[k = 7, N_{ES} = 13] = 0.24, p = .35), negative emotions (g_{IG} [k = 7, N_{ES} = 14] = −0.04, p = .71), and attitudes (g_{IG} [k = 7, N_{ES} = 11] = 0.08, p = .36) did not differ significantly from zero. The effect size estimates for accurate beliefs were heterogeneous, with the between-study variance (τ^2) estimated as .11 for IG effects and .04 for IGRM effects for accurate beliefs, respectively, indicating that there was variation in effects for accurate beliefs that moderators might have been able to explain. However, the between-study variance (τ^2) was essentially 0 for both IG effects and IGRM effects for inaccurate beliefs (see Table 3), suggesting little variation in effects for inaccurate belief outcomes.

Publication bias

Moreover, as part of our first research question we asked if any evidence of publication bias in these effects is present. To assess potential publication bias due to missing studies and small samples, we generated funnel plots (see Figure 3 and 4). The funnel plots for accurate and inaccurate beliefs appeared to be symmetrical, so we used Egger's regression test to assess the asymmetry. Results showed no evidence of funnel plot asymmetry for any outcome, using either IGRM (IGRM accurate beliefs: b = −0.67, SE = 1.22, t(13.14) = −0.55, p = .59; IGRM inaccurate beliefs: b = 0.70, SE = 4.88, t(1.42) = 0.14, p = .90) or IG effects (IG Accurate beliefs: b = −0.67,

$SE = 0.78$, $t(16.77) = -0.86$, $p = .40$; IG inaccurate beliefs: $b = 4.93$, $SE = 5.23$, $t(2.93) = 0.94$, $p = .42$; positive emotions: $b = -12.97$, $SE = 10.62$, $t(2.37) = -1.22$, $p = .33$; negative emotions: $b = -2.50$, $SE = 1.61$, $t(2.60) = -1.56$, $p = .23$; attitudes: $b = -1.75$, $SE = 3.41$, $t(3.40) = -0.51$, $p = .64$). Taken together, funnel plot and "Egger's" regression models show little evidence of publication bias in this synthesis. Details on the results of the selection model analyses are provided in the supplemental materials (see Table S3).

Moderator analyses

To address research questions 2 and 3, we examined moderators of refutation text effects across both accurate and inaccurate beliefs combined in one set of analyses using IG effects, as well as a second set using IGRM effects. In these analyses, the direction of effects on inaccurate beliefs was reversed so that all positive effects would indicate greater accuracy in the refutation text condition. We chose this strategy to maintain statistical power, given the identical hypotheses across accurate and inaccurate beliefs and the limited number of effects for inaccurate beliefs. Results are presented starting with moderator analyses of covariates to be controlled in subsequent moderator analyses (see Table 4). Then, we present results in order of the separate confirmatory and exploratory research questions, in alignment with the organization of the pre-registration. Table 5 presents the results of moderator analyses addressing confirmatory research questions (2A and 3A) and Table 6 presents the results of moderator analyses addressing exploratory research questions (2B and 3B).

Covariates

We began by examining whether a set of covariates that we intended to include in all subsequent models explained heterogeneity in the effect of refutation texts on beliefs (Table 4). We examined the following moderators: (1) type of belief (accurate or inaccurate), (2) publication status, (3) random versus nonrandom assignment to condition, and (4) whether or not participants were screened to exclude individuals with highly accurate beliefs. We examined each of the four covariates separately, using IG effects in one set of analyses, and IGRM effects in a second set. None of the covariates predicted statistically significant differences in the refutation effects using either IG or IGRM effects. Nevertheless, we retained and included all four as covariates in subsequent moderator analyses. It is important to note that only a few studies use nonrandom assignment, and thus there were not enough studies with variation on this factor to draw a firm conclusion. Although this analysis should not be interpreted on its own, it is still a useful covariate when examining other moderators.

Research question 2A: confirmatory text characteristic moderators

For research question 2 A, five text characteristic moderators for which we had confirmatory hypotheses were specified.

Table 3. Overall average effects of refutation texts on outcomes.

					95% CI			95% PI
Outcome	k	N_S	N_{ES}	g	Low/High	τ^2	Q	Low/High
IGRM (Change) Effects								
Accurate beliefs	49	68	113	.36***	.24/.47	.04	562.14***	−.41/1.12
Inaccurate beliefs	6	7	10	−.31**	−.44/−.18	.00	9.77	−.38/−0.24
IG (Post Only) Effects								
Accurate beliefs	67	102	222	.37***	.26/.47	.11	1315.23***	−.53/1.26
Inaccurate beliefs	10	11	14	−.38**	−.55/−.21	.00	29.32**	−.83/.07
Positive emotions	7	9	13	.24	−.33/.82	.33	88.52***	−1.24/1.72
Negative emotions	7	9	14	−.04	−.32/.23	.05	26.48*	−.64/.55
Positive attitudes	7	10	11	.08	−.12/.28	.00	27.30**	−.36/.50

Note. k = number of studies. N_S = number of samples. N_{ES} = number of effects. g = Hedges' g (average pooled effect). CI = confidence interval. PI = prediction interval. Low = lower estimate. High = upper estimate. *$p < .05$, **$p < .01$, ***$p < .001$.

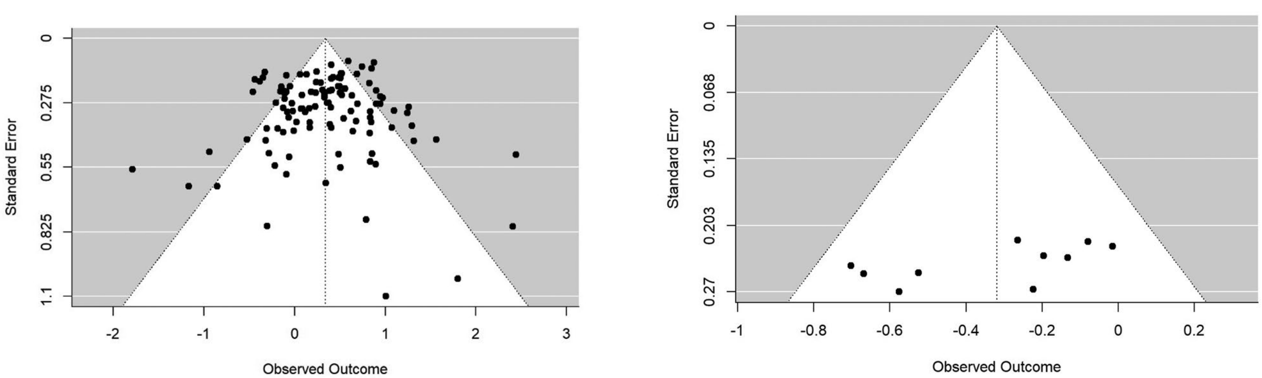

Figure 3. Funnel Plots for IGRM Effects.
Note. Funnel plot for accurate beliefs on left and inaccurate beliefs on right.

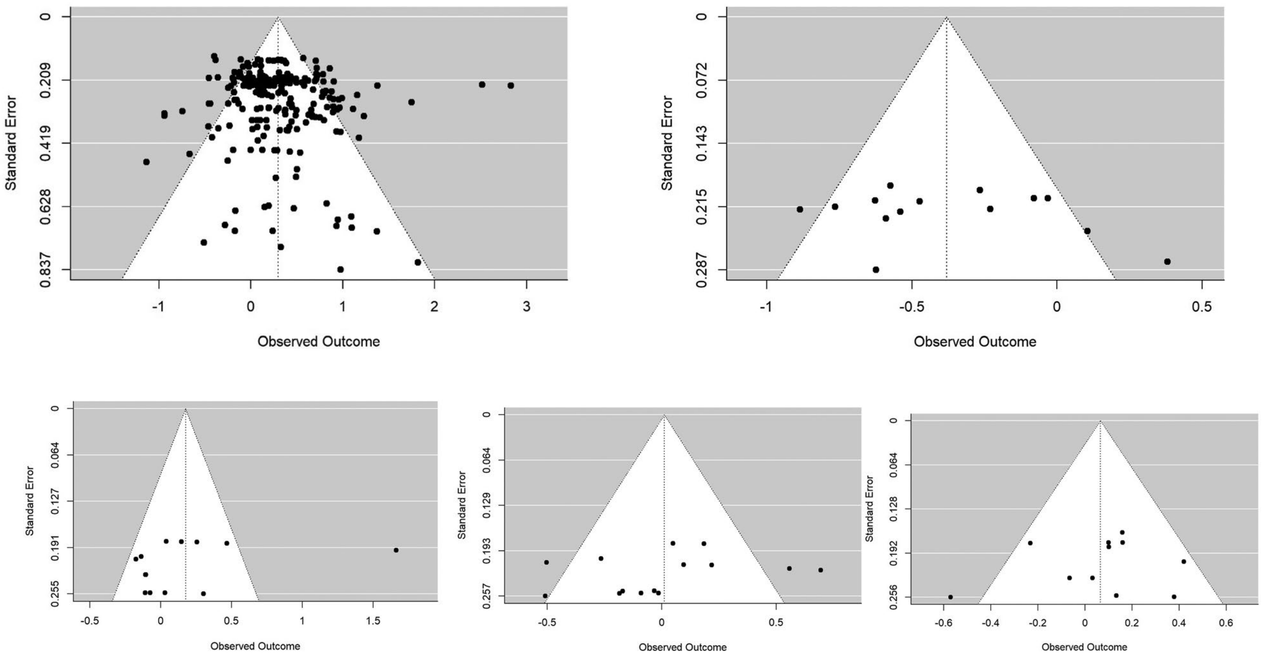

Figure 4. Funnel Plots for IG Effects.
Note. Funnel plot for accurate beliefs on top left and inaccurate beliefs on top right. Funnel plot for positive emotions bottom left, negative emotions bottom middle, and positive attitudes bottom right.

Table 4. Results of covariate moderator analyses.

	IGRM refutation text effects on changes in beliefs							IG refutation text effects on post intervention beliefs						
						95% CI	95% PI						95% CI	95% PI
Moderator	*k*	N_S	N_{ES}	*b(SE)*	*g*	Low/High	Low/High	*k*	N_S	N_{ES}	*b(SE)*	*g*	Low/Hi	Low/Hi
Type of belief														
Accurate	48	67	112	–	0.35***	0.24/0.47	−0.04/0.74	67	103	221	–	0.37***	0.26/0.47	−0.28/1.02
Inaccurate	6	7	10	0.05(0.10)	0.40*	0.15/0.65	−0.16/0.97	10	11	14	0.11(0.11)	0.47**	0.23/0.72	−0.30/1.24
Publication status														
Published	39	53	92	–	0.40***	0.27/0.53	0.02/0.77	51	79	173	–	0.39***	0.29/0.51	−0.26/1.05
Unpublished	10	15	30	−0.18(0.10)	0.22*	0.03/0.40	−0.23/0.66	17	25	62	−0.08(0.13)	0.31*	0.05/0.57	−0.42/1.04
Random assignment														
Random	46	64	110	–	0.38***	0.28/0.48	0.03/0.73	63	98	221	–	0.38***	0.27/0.49	−0.27/1.04
Not random	3	3	12	−0.22(0.30)	0.16	−1.46/1.78	−1.69/2.02	5	6	14	−0.15(0.14)	0.23	−0.17/0.64	−0.83/1.30
Screening														
Does not exclude	41	52	92	–	0.35***	0.22/0.48	−0.06/0.75	55	74	157	–	0.38***	0.25/0.50	−0.29/1.04
Excludes highly acc	10	16	30	0.06(0.08)	0.41***	0.29/0.53	−0.04/0.86	16	30	78	−0.02(0.08)	0.36***	0.24/0.48	−0.35/1.07

Note. k = number of studies. N_S = number of samples. N_{ES} = number of effects. b = unstandardized regression slope coefficient (moderator effect). SE = standard error. g = Hedges' g (average pooled effect). CI = confidence interval. PI = prediction interval. Low = lower estimate. High = upper estimate. Acc = accurate beliefs. Dashes (–) in *b(SE)* column for reference category of moderator analyses. Each covariate was tested in a separate moderator model. The signs of inaccurate belief effects were reversed so effects could be compared to inaccurate beliefs. *$p < .05$, **$p < .01$, ***$p < .001$.

Specifically, we examined the extent to which the relative level of referential cohesion, causal cohesion, narrativity, syntactic simplicity, and word concreteness across refutation and control texts (computed by taking the difference) moderated the effect of refutation texts on beliefs even after accounting for covariates. Referential cohesion, causal cohesion, narrativity, syntactic simplicity, and word concreteness indices were generated by applying Coh-Metrix (Graesser et al., 2004, 2011) to both the refutation and control texts. Moderator patterns were analyzed for each of the five text characteristic moderator variables separately, using a model for each that included the four covariates. None of these moderators explained heterogeneity in the intervention effects, suggesting that the positive effect of refutation text on beliefs is relatively robust across circumstances.

Research question 2B: exploratory text characteristic moderators

For research question 2B, we proposed to explore the extent to which eight additional characteristics of texts moderated the effect of refutation texts on beliefs even after accounting for covariates. Specifically, we explored the role of discipline (five levels—geosciences, life sciences, physics/applied physics, social science), controversial nature of the topic (two levels—controversial, not controversial), the relative text

Table 5. Results of confirmatory moderator analyses.

	IGRM refutation text effects on changes in beliefs							IG refutation text effects on post intervention beliefs						
						95% CI	95% PI						95% CI	95% PI
Moderator	*k*	N_S	N_{ES}	*b(SE)*	*g*	Low/High	Low/High	*k*	N_S	N_{ES}	*b(SE)*	*g*	Low/Hi	Low/Hi
Text characteristics														
Rel referential cohesion	22	27	54	0.001 (0.002)	–	–	–	29	41	102	−0.002 (0.005)	–	–	–
Rel causal cohesion	23	29	56	−0.007 (0.007)	–	–	–	30	43	104	0.003 (0.006)	–	–	–
Rel narrativity	23	29	56	−0.003 (0.03)	–	–	–	30	43	104	0.008 (0.01)	–	–	–
Rel syntactic simplicity	22	27	54	0.008 (0.008)	–	–	–	29	41	102	0.01 (0.005)	–	–	–
Rel word concreteness	23	29	56	−0.004 (0.007)	–	–	–	30	43	104	−0.02 (0.02)	–	–	–
Sample characteristics														
Screening[a]														
Does not exclude	41	52	92	–	0.42***	0.28/0.56	0.08/0.75	55	74	157	–	0.42***	0.27/0.57	−0.25/1.09
Excludes highly acc	10	16	30	0.02 (0.08)	0.45***	0.30/0.59	0.08/0.81	16	30	78	−0.04 (0.09)	0.37***	0.24/0.51	−0.33/1.08
Setting characteristics														
Educational setting														
Embedded/ Authority	8	14	28	0.16 (0.16)	0.50*	0.16/0.85	−0.02/1.04	13	21	49	0.04 (0.15)	0.47**	0.15/0.78	−0.34/1.28
Not emb/Auth	27	35	59	–	0.35**	0.16/0.53	−0.04/0.74	35	53	110	–	0.43***	0.28/0.59	−0.27/1.13
Not emb/No auth	4	5	8	0.23 (0.15)	0.58*	0.19/0.97	−0.10/1.26	4	6	17	−0.19 (0.14)	0.24	−0.09/0.58	−0.80/1.28
Outcome characteristics														
Question type														
Closed	33	48	82	–	0.45***	0.28/0.62	0.08/0.82	46	74	129	–	0.45***	0.29/0.62	−0.21/1.11
Open	12	13	21	−0.14 (0.10)	0.31**	0.10/0.51	−0.11/0.72	26	38	60	−0.06 (0.08)	0.39***	0.21/0.57	−0.28/1.06
Both	6	9	17	0.02 (0.17)	0.48*	0.04/0.91	−0.11/1.06	16	29	42	−0.08 (0.06)	0.37***	0.20/0.54	−0.31/1.05
Outcome reliability	26	36	60	−0.009 (0.01)	–	–	–	35	56	96	0.005 (0.002)	–	–	–

Note. *k* = number of studies. N_S = number of samples. N_{ES} = number of effects. *b* = unstandardized regression slope coefficient (moderator effect). SE = standard error. *g* = Hedges' *g* (average pooled effect). CI = confidence interval. PI = prediction interval. Low = lower estimate. High = upper estimate. Acc = accurate beliefs. Emb = embedded in authentic setting. Auth = Authority. Rel = Relative level. Dashes (–) in *b(SE)* column for reference category of moderator analyses and for continuous moderators, where average effect information for levels of a moderator is not relevant. Each moderator was tested in separate models that included the four covariates (type of belief, publication status, random assignment to condition, and screening); covariates results omitted from tables. [a]In contrast to the examination of screening in Table 4, this analysis screening controlled for the other three covariates. *$p < .05$, **$p < .01$, ***$p < .001$.

length, and the overall level of referential cohesion, causal cohesion, narrativity, syntactic simplicity, and word concreteness among texts for which the relative difference between refutation and control was less than 10%. Again, none of these moderators predicted the refutation text effects. Average effects were typically positive and statistically significant across various levels of these moderators.

Research question 3A: confirmatory design, sample, setting and outcome characteristic moderators

For research question 3 A, four design, sample, setting, and outcome characteristic moderators for which we had confirmatory hypotheses were specified. Specifically, we examined the extent to which screening (2 levels—accurate beliefs screened/excluded, not screened/included individuals with highly accurate beliefs), the educational setting (three levels—embedded in an authentic setting/person of authority present, not embedded in an authentic setting/person of authority present, not embedded in an authentic setting/person of authority not present), the type of questions in the outcome measure (three levels—open-ended, closed-ended, both open and closed), and reliability (alpha) of the closed-ended outcome measure as reported in the primary article moderated the overall effectiveness of refutation texts even after accounting for our set of covariates.

None of these moderators explained heterogeneity in the intervention effects. Although average effects were typically positive and statistically significant across various levels of these moderators, the positive average refutation text effect did not significantly differ from zero in one instance. Specifically, the refutation effect was not statistically significantly different from zero when the study content was not embedded and a person in authority was not present using IG effects.

Research question 3B: exploratory design, sample, setting and outcome characteristic moderators

For research question 3B, we proposed to explore the extent to which six additional characteristics of the design, sample, setting, and outcome moderated the effect of refutation texts on beliefs even after accounting for covariates. Specifically, we explored the role of type of control text (two levels—one-sided, two-sided), education level (three levels—elementary, middle/high school, college; eight adult or multilevel samples excluded), gender (percent female in the sample), geographic region (three levels—North America, Europe, Other), country's level of science trust (low, high, Rabesandratana, 2019), and the timing of the outcome measurement relative to the administration of the text intervention in minutes.

Again, none of these moderators explained heterogeneity in the intervention effects. Refutation text effects were robust across circumstances, including across control text type, education levels, gender of the sample, continent, country level of trust in science, and outcome measurement timing. Although average effects were typically positive and statistically significant across various levels of these moderators, the positive average refutation text effect on beliefs did not significantly differ from zero in a few instances. Specifically, using IGRM effects, the average effect of refutation text was not statistically significantly different from zero for either elementary or secondary students but was for college

Table 6. Results of exploratory moderator analyses.

	IGRM refutation text effects on changes in beliefs							IG refutation text effects on post intervention beliefs						
						95% CI	95% PI						95% CI	95% PI
Moderator	*k*	N_S	N_{ES}	*b(SE)*	*g*	Low/High	Low/High	*k*	N_S	N_{ES}	*b(SE)*	*g*	Low/Hi	Low/Hi
Text Characteristics														
Discipline														
Physics	10	14	24	–	0.49*	0.10/0.88	−0.09/1.07	17	34	76	–	0.54**	0.19/0.90	−0.26/1.33
Geosciences	12	19	41	−0.12 (0.20)	0.37*	0.10/0.64	−0.12/0.86	13	20	56	−0.14 (0.14)	0.40***	0.20/0.60	−0.34/1.14
Life sciences	15	19	29	−0.05 (0.17)	0.44***	0.25/0.63	−0.01/0.89	21	28	54	−0.11 (0.18)	0.43***	0.21/0.65	−0.30/1.17
Social science	12	16	28	−0.09 (0.19)	0.40**	0.15/0.65	−0.09/0.89	17	22	49	−0.20 (0.18)	0.34**	0.14/0.55	−0.39/1.08
Controversial topic														
No	20	26	46	–	0.44***	0.25/0.59	0.05/0.83	35	55	115	–	0.39***	0.19/0.58	−0.32/1.09
Yes	29	42	76	0.04 (0.11)	0.40***	0.22/0.63	0.02/0.78	33	49	120	−0.06 (0.10)	0.46***	0.30/0.63	−0.24/1.16
Rel text length	37	51	91	−0.0006 (0.0004)	–	–	–	48	72	174	−0.0006 (0.0004)	–	–	–
Mn referential cohesion	20	25	51	−0.0005 (0.003)	–	–	–	27	38	97	−0.005 (0.006)	–	–	–
Mn causal cohesion	16	22	40	0.002 (0.01)	–	–	–	19	32	76	−0.001 (0.005)	–	–	–
Mn narrativity	15	20	36	−0.01 (0.006)	–	–	–	17	27	67	0.004 (0.007)	–	–	–
Mn syntactic simplicity	9	13	28	−0.009 (0.007)	–	–	–	16	27	74	0.003 (0.006)	–	–	–
Mn word concreteness	21	27	53	0.0005 (0.003)	–	–	–	25	37	94	−0.002 (0.003)	–	–	–
Design Characteristics														
Control Text														
One-sided	41	56	100	–	0.44***	0.28/0.59	0.09/0.78	54	77	172	–	0.39***	0.24/0.55	−0.30/1.09
Two-sided	10	12	22	−0.10 (0.14)	0.34*	0.05/0.62	−0.10/0.77	16	27	63	0.16 (0.15)	0.55**	0.24/0.86	−0.21/1.32
Sample Characteristics														
Education Level														
Elementary	3	5	14	−0.20 (0.17)	0.21	−0.49/0.91	−0.85/1.27	4	6	22	−0.16 (0.09)	0.28*	0.02/0.53	−0.74/1.29
Secondary	7	12	20	0.10 (0.23)	0.51	−0.03/1.04	−0.17/1.18	9	14	30	0.17 (0.21)	0.61*	0.10/1.12	−0.29/1.51
College	35	44	78	–	0.41***	0.23/0.58	0.006/0.80	52	78	169	–	0.43***	0.29/0.57	−0.25/1.12
% Female	40	54	88	$-0.009^{\dagger}$(0.004)	–	–	–	52	77	161	$-0.007^{\dagger}$ (0.004)	–	–	–
Setting Characteristics														
Region														
North America	29	40	74	–	0.38***	0.19/0.57	0.03/0.74	40	66	159	–	0.40***	0.25/0.56	−0.20/1.01
Europe	16	23	39	−0.02 (0.12)	0.36**	0.17/0.55	0.005/0.71	23	32	66	−0.10 (0.08)	0.31***	0.19/0.42	−0.29/0.90
Other	4	5	9	0.38 (0.27)	0.76	−0.24/1.76	−0.37/1.89	5	6	10	0.62 (0.45)	1.02	−0.33/2.38	−0.57/2.62
Country science trust														
Low	45	63	110	–	0.44***	0.27/0.60	0.08/0.80	62	97	218	–	0.45***	0.28/0.62	−0.22/1.12
High	4	5	12	0.09 (0.12)	0.34*	0.01/0.68	−0.31/0.99	6	7	17	0.22 (0.11)	0.23*	0.05/0.41	−0.66/1.13
Outcome characteristics														
Timing of outcome	49	68	120	−0.000 (0.000)	–	–	–	68	104	233	−0.000 (0.000)	–	–	–

Note. *k* = number of studies. N_S = number of samples. N_{ES} = number of effects. *b* = unstandardized regression slope coefficient (moderator effect). SE = standard error. *g* = Hedges' *g* (average pooled effect). CI = confidence interval. PI = prediction interval. Low = lower estimate. High = upper estimate. Acc = accurate beliefs. Emb = embedded in authentic setting. Auth = Authority. Rel = Relative level. Mn = Mean level. Dashes (–) in *b(SE)* column for reference category of moderator analyses and for continuous moderators, where average effect information for levels of a moderator is not relevant. Each moderator was tested in separate models that included the four covariates (type of belief, publication status, random assignment to condition, and screening); covariates results omitted from tables. $^{\dagger}p \le 0.10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

students. The refutation text effect was also not statistically significant across other continents outside of North America and Europe using either IG or IGRM effects.

Discussion

Although refutations have long been touted as a means of correcting misconceptions, the evidence of their effectiveness has been mixed (Guzzetti et al., 1993; Tippett, 2010). Two recent efforts to systematically explore the efficacy of refutation texts (Schroeder & Kucera, 2022; Zengilowski et al., 2021) have left many questions unanswered. We set out to address some shortcomings in prior systematic and meta-analytic reviews on refutation texts by using a more comprehensive search that yielded more studies. We specifically focused on scientific information, due to its proliferation online, as well as the fact that most studies have focused on science topics. In addition to ensuring that our search was as comprehensive as possible, we also explored new moderators of theoretical and practical interest and used more exacting methods that employed state-of-the art statistical approaches and differentiated post-intervention effects from pre-post change effects.

Overall, our findings show a consistent and statistically significant advantage of refutation text (g_{IG} = 0.37, g_{IGRM} = 0.36) over non-refutation texts in confronting scientific misconceptions. These findings support the findings of previous reviews, particularly those of Schroeder and Kucera (2022), which showed a similar effect size advantage for refutation texts. Some have pushed back on the use of refutation texts over concerns about potential backfire effects. Our findings are in line with Ecker et al.'s (2022) findings in that our meta-analysis found no evidence of widespread backfire effects. Given that a substantial amount of science communication is online via web pages, social media, and public service announcements, we recommend incorporating refutational approaches into communication on these platforms.

Moderators

Past systematic reviews of refutation texts included few to no moderators (Guzzetti et al., 1993; Schroeder & Kucera, 2022; Tippett, 2010; Zengilowski et al., 2021). We aimed to move beyond the question of efficacy alone to answer a more nuanced question: *If refutation texts are effective, with whom and under what circumstances do they have an effect*?

We examined the broadest set of moderators to date, including reader characteristics (i.e., gender, education level), topic and text characteristics, (i.e., science discipline, relative cohesion across texts), setting characteristics (i.e., embeddedness in an authentic setting, geographic region), outcome measurement characteristics (i.e., open/closed assessments, timing of assessments), and research design characteristics (i.e., control text type), as well as covariates (i.e., use of random assignment to condition, prior knowledge screening). Despite our best efforts to elucidate the circumstances in which these texts are and are not effective, it was not possible to definitively answer that question because our moderators did not explain variability in the observed effects. Said another way, our findings suggest that refutation texts consistently outperform their non-refutational counterparts with little variation in effects across different designs, groups of individuals, settings, topics, text features, or measures.

There are several possible reasons why the proposed moderators did not have a statistically significant effect. Despite a large sample of effects across 76 studies, perhaps there were too few studies examining individual moderators of interest to detect a significant moderator effect, especially for certain moderators. For example, only five studies took place outside of America/Europe and only six countries in the analysis could be categorized as "high trust." Despite our assumption that these populations would differ significantly on their acceptance of science, it is likely that the within-country variance outweighed the between-country differences.

Interestingly, Schroeder and Kucera (2022) also found no moderator effects. It is therefore possible that the impact of refutation texts is modest but stable across populations, content, and text features. That is, the refutational structure itself may provide an advantage to learners compared to other text formats. As one of the authors of this manuscript quipped—they seem to work in a house and with a mouse, in a box and with a fox, on a train and in the rain, here or there or anywhere. Moreover, the lack of moderators may indicate that something deeper is at play. Perhaps the structure of a refutation text affords the opportunity for co-activation, which is the mechanism for change. Following the KReC framework, when properly constructed with argument overlap and in a causal, explanatory style (Sinatra & Broughton, 2011), the refutation may "tee up" the opportunity for readers to notice the discrepancy and correct their misconceptions. Given that a host of moderators in this and other studies have not provided an alternative explanation of when and how refutation texts are effective, perhaps an Occam's razor argument is that structure itself accounts for the effect.

Therefore, further avenues for research may focus more on the enhancement of refutational texts than determining whether they are effective. We look to the work on Multimedia Learning Theory as a framework, wherein after decades of determining *whether* graphics can facilitate greater learning outcomes (Mayer & Sims, 1994), the discussion became more nuanced on the principles that educators could use to maximize their communication effectiveness (Mayer, 2002, 2009, 2019) or which gestures are most effective with pedagogical agents (Li et al., 2019). Recent examples in the refutation text literature include the work of Danielson et al. (2016) and Thacker et al. (2020), who showed that refutation texts can be enhanced by adding graphics and a metaphor or persuasive text features, respectively, as well as Trevors and Kendeou (2020), who embedded emotional content into refutation texts.

Limitations

All studies have limitations, and ours is no exception. First, although we sought to examine several outcomes that were deemed critical for theories of conceptual change (i.e., attitudes, emotions), very few studies have examined outcomes other than beliefs. We encourage future researchers to include these measures, because individuals may react strongly when presented with attempts to correct their misconceptions. Second, although the texts that we examined included a large variety of domains, we did not deeply analyze specific concepts within these domains. There may be distinctions among topics that we did not drill down far enough to reveal. Third, several of our moderators had little variability or were only measured in a small number of studies, which limited our power to detect differences. Some of these limitations point to promising future directions. For example, few studies engaged with students in elementary school or outside of the US/Europe. Including these populations may greatly enhance understanding of the effectiveness of refutation text, and therefore we strongly encourage researchers to engage with these populations in future studies. Fourth, we could not examine the characteristics of the sample in much detail because there was limited information and variability in the samples. For example, race information was often not reported, and researchers provided limited information about other characteristics of their samples. We implore researchers to collect and publish this information in the future to more accurately determine for whom refutation texts are effective, and under what circumstances.

Finally, we included experimental studies that focused on text alone and restricted our sample to studies without an interactive component or discussion. This methodological decision was made to reduce heterogeneity in findings that may not be explained through moderator analyses. This is because few studies included an interactive or discussion component, and the nature of that component differed greatly between studies. However, this choice comes with consequences for external validity. When used in classrooms for educational purposes, we suspect that refutation texts will often be used in combination with other activities, such as whole or small group discussion. Our team identified a sufficient number of studies to examine the effects of refutational strategies that go beyond text in a systematic way. However, as these were not included here, this limits the generalizability of our findings.

Implications for theory

Three theoretical frameworks informed our study, the KReC framework (e.g. Kendeou & O'Brien, 2014), conceptual change (e.g. Posner et al., 1982; Sinatra, 2022; Treagust &

Duit, 2008), and the "medical" analogy to combating misinformation (e.g. Ecker et al., 2022; Lewandowsky et al., 2020). Our findings have implications for all three but may show the strongest support for the KReC framework. Because moderators failed to predict refutation text effects in this review, and in any other systematic reviews, we suggest that the refutation text effect is likely due to a more "basic" mechanism—that of working memory. According to the KReC framework, the advantage of refutation text is that its structure affords the comparison of prior and new information within working memory. The co-activation of competing ideas allows both to be compared within the limits of working memory. This should free up cognitive resources that might otherwise be allocated to parsing the text or searching long term memory for relevant prior knowledge. To test this, refutation text researchers should include measures of working memory and reading fluency. Findings may show that a refutation text affords readers of *lower* working memory capacity the opportunity to compare information within their working memory limitations. Such studies would have to control for factors such as background knowledge, age, and grade level.

Conceptual change researchers should focus on refutation texts' role in correcting specific misconceptions that have the potential to restructure knowledge. We have argued that misconception correction cannot be equated with conceptual change unless readers can be shown to have restructured their conceptual knowledge. For example, fish have vertical fins whereas whales and dolphins have horizontal fins. This surface-level difference (i.e., orientation of fins) belies a deeper conceptual difference (i.e., the concept of "mammals" includes whales and dolphins but not fish). Overcoming a misconception of fin orientation might lead to the type of conceptual change Chi (2009) refers to as ontological shifts.

Given the import of "warm" constructs to the conceptual change framework, more of these constructs should be tested as outcome measures, not just predictors of change. Although Heddy et al. (2017) found a shift in attitudes and emotions with a reduction in misconceptions, conceptual change researchers should explore whether such shifts resulting from refutation texts occur when misconceptions are overcome, and knowledge is restructured. Schroeder and Kucera (2022) mention the importance of these constructs throughout their meta-analysis yet did not include them as moderators, perhaps due to how infrequently these variables are measured. Although researchers (Trevors et al., 2016; Sinatra, 2005) have argued for the importance of emotional/attitudinal variables, we only located seven studies including these constructs as outcome measures that met our selection criteria.

Finally, those employing the "medical" analogy approach to correcting misinformation have not focused on the structural issues of the text. These researchers could more systematically evaluate the three-part structure as described by Kendeou et al. (2014). This includes stating the misinformation, refuting it, and explaining the correct information. For example, future research might examine how alterations to this structure might impact the effectiveness of the text. Inspiration might be drawn from metacognition studies manipulating the relative distance between text titles and textual information (Lippmann et al., 2021) or multimedia studies examining how texts and graphics are paired to maximize or disrupt learning (Clark & Mayer, 2023). Debunking research (Ecker et al., 2020) has cautioned about backfire effects but more carefully crafted texts might allow researchers to understand what features of the text or the reader prompted those backfire effects (Jacobson et al., 2021). Further, this literature also suggests that repeating misinformation is harmful as it may result in the reader remembering the misinformation, but not necessarily remembering that it is incorrect. Refutation texts that include multiple refutation statements should be evaluated to examine whether they cause readers to double down on incorrect beliefs or if, when carefully crafted, whether they can be an effective debunking strategy.

Practice and policy implications

One question to consider is: why is it necessary to confront misconceptions at all? Might it be better to build scientifically accurate knowledge? We believe this is a false choice, as it is no choice at all. We posit that both instructional options are useful. A focus on reduction in misconceptions might be "best" when misconceived ideas are particularly harmful (e.g., putting individuals' health at risk), create confusion, prevent the development of conceptual understanding, prevent conceptual change (Danielson et al., 2016) or prevent epistemic conceptual change (thinking differently about knowledge and knowing, Sinatra & Chinn, 2011). A focus on building scientifically accurate knowledge is a laudable goal, especially when background knowledge is limited or non-existent. Indeed, building accurate science knowledge is the primary goal of most science instruction, as it should be.

One critique of refutation texts is that they are not an effective means of teaching science in K-12 settings (Zengilowski et al., 2021). We find this to be a straw man argument, as we know of no one who suggests that refutations should be a *primary* method for teaching science in K-12, higher education, or any learning environment. Although our work and that of others provides evidence that refutational approaches are helpful for debunking misinformation, teaching science involves much more than debunking flawed scientific ideas. It requires engaging students in active learning (Lombardi et al., 2022) and the practices of science, as suggested by the Next Generation Science Standards (National Research Council, 2013). Science education scholars (Brown, 2021; Bybee, 1997; Windschitl et al., 2020) have provided the field with excellent strategies, methods, and recommendations (National Research Council, 2013; Schwarz et al., 2009), which we endorse for effective science pedagogy. We do not recommend teaching science through text alone, even effective texts, in any level of education. We do see refutations as playing an important role when students hold misconceptions. Teachers may consider leveraging refutation texts as a starting point for interactive discussions when misconceptions are identified.

We recommend three guidelines for educators and education policy makers. First, there is a need for small-scale interventions that teachers can develop and use, as well as guidelines for teachers on how to implement them in their classroom. Targeted, small-scale, flexible interventions can be used to confront misconceptions. Although they do not take the place of more interactive deep learning opportunities through engaging instruction, they are tools with utility. Second, teachers and practitioners can take more formal ownership over these interventions to develop and deploy them in their classrooms. Third, policy makers should support the development of refutational resources for use in classrooms with funded research for testing the efficacy of these approaches, provide guidance for their use, and fund and support professional development for teachers to employ them effectively.

Future directions

Despite the emerging consensus of systematic reviews all showing a positive impact of refutation texts, there is still much to be done. In this meta-analysis, we explored experimental studies, mostly outside of "authentic" instructional contexts. The question remains as to whether refutations are effective when they are embedded within instructional contexts. We have argued that refutation texts should play only a limited role in K-12 science education. However, even limited use would need to be justified with evidence of their effectiveness in instructional settings. If the effect is modest in experimental contexts as we saw here, it is worth exploring whether the effect holds up in real-world contexts such as a high school biology classroom that include discussions and interactions among students and teachers. Our team is currently exploring the impact of refutations in instructional contexts and encourages others to do so as well.

Refutations are often used in other authentic contexts, such as in science communication. One of our coauthors has received periodic updates about COVID-19 from her doctor, much of this in a three-part refutational style (i.e., you may have heard that you are not at risk of a COVID-19 infection at this time, but this is not the case…). It is unclear whether these communications are effective. Moving beyond anecdotal examples is important because a systemic or meta-analytic review has not yet examined the effectiveness of science communication messaging structured to refute medical or public policy science misinformation, beyond an experimental setting (e.g., a doctor debunking misinformation for her patients).

Efforts to modify or enhance the impact of refutation texts have shown modest success. Researchers have added graphics, pictures, metaphors, and persuasive arguments (i.e., Danielson et al., 2016, Mason et al., 2017, Thacker et al., 2020). Other enhancements or modifications may also have added value. For example, although refutations typically mirror expository texts, the power of narrative enhancement has not been deeply explored. In other words, could a refutation text be more persuasive if information was debunked with a narrative element in the text (e.g., a clearly labeled fictional scenario of a peer who changed their minds and why)? Other examples might include recent work in "gamification" (Trevors & Ladhani, 2022) and large-scale public health campaigns, where interventions with small effects may still reach hundreds of thousands of individuals. These modifications might contain more interactive, game-like effects, or more traditional television/website content aimed at recognizing and refuting common misconceptions. In short, creative enhancements to the texts themselves and to the scale of interventions are worth exploring to determine whether the modest impact of refutation texts can be amplified.

Finally, we recommend employing careful design and reporting standards so that meta-analysts can include those studies in future systematic reviews. Research on power in meta-analyses with dependent effect size models suggests that the small to medium overall average effects typically found in education research will be adequately powered (i.e., above .80) when at least 10 studies are included (Vembye et al., 2022). Therefore, we might speculate that moderator analyses that have fewer than 10 studies are not adequately powered. However, research on power for overall average effects may not generalize to models with moderators. Researchers have yet to explore power for moderator analyses with dependent effect models. Those conducting refutation text studies should include information on the demographic characteristics of their samples and other textual characteristics to explore their potential as moderators. Moreover, researchers should expand refutation text effects to populations that have been underrepresented in science, including participants from countries outside the US and historically marginalized groups.

Conclusions

Our findings show that refutation texts are a modest but effective means of confronting misinformation. Given the widespread and sometimes deadly proliferation of scientific misinformation online, even a modest effect is worth employing to debunk scientific misconceptions. For example, consider the number of deaths of individuals who tried dangerous and untested COVID-19 remedies and those who refused to get a life-saving vaccine due to misinformation about side effects. Science literacy is a dynamic and complex learning process that requires a range of effective approaches. Refutation texts should be considered as one, albeit limited, useful approach to employ to target misinformation in a variety of settings. In future research, we aim to examine the extent to which they are effective in other contexts, and we encourage others to explore the utility of this approach in a variety of communication and learning environments.

Disclosure statement

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

Funding

This research was supported by a grant from the National Science Foundation10.13039/100000001 [Project #1920348].

ORCID

Robert W. Danielson http://orcid.org/0000-0002-6288-5719
Erika A. Patall http://orcid.org/0000-0003-3448-8932
Gale M. Sinatra http://orcid.org/0000-0002-6545-587X
Alana A. U. Kennedy http://orcid.org/0000-0002-1464-0813
Onur Ramazan http://orcid.org/0000-0002-0805-7103
Gan Jin http://orcid.org/0000-0002-4058-2397

References

*Indicates study included in the meta-analysis

*Adesope, O. O., Cavagnetto, A., Hunsu, N. J., Anguiano, C., & Lloyd, J. (2017). Comparative effects of computer-based concept maps, Refutational texts, and expository texts on science learning. *Journal of Educational Computing Research*, *55*(1), 46–69. https://doi.org/10.1177/0735633116654163

*Aguilar, S., Polikoff, M., & Sinatra, G. M. (2019). Refutation texts: A new approach to changing public misconceptions about education policy. *Educational Researcher*, *48*(5), 263–272. https://doi.org/10.3102/0013189X19849416

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics*, *6*(2), 205316801984855. https://doi.org/10.1177/2053168019848554

Allen, M. (2017). Errors of measurement: Attenuation. In *The SAGE encyclopedia of communication research methods* (Vol. 4, pp. 430–433). SAGE Publications, Inc. https://doi.org/10.4135/9781483381411

Alongi, M. D., Heddy, B. C., & Sinatra, G. M. (2016). Real-world engagement with controversial issues in history and social studies: Teaching for transformative experiences and conceptual change. *JSSE-Journal of Social Science Education*, *15*(2), 26–41.

*Alvermann, D. E., & Hague, S. A. (1989). Comprehension of counter-intuitive science text: Effects of prior knowledge and text structure. *The Journal of Educational Research*, *82*(4), 197–202. https://doi.org/10.1080/00220671.1989.10885893

*Alvermann, D. E., & Hynd, C. R. (1989). Effects of prior knowledge activation modes and text structure on Nonscience majors' comprehension of physics. *The Journal of Educational Research*, *83*(2), 97–102. https://doi.org/10.1080/00220671.1989.10885937

*Ariasi, N., Hyönä, J., Kaakinen, J., & Mason, L. (2017). An eye-movement analysis of the refutation effect in reading science text. *Journal of Computer Assisted Learning*, *33*(3), 202–221. https://doi.org/10.1111/jcal.12151

*Ariasi, N., & Mason, L. (2014). From covert processes to overt outcomes of refutation text reading: The interplay of science text structure and working memory capacity through eye fixations. *International Journal of Science and Mathematics Education*, *12*(3), 493–523. https://doi.org/10.1007/s10763-013-9494-9

*Asterhan, C. S., & Resnick, M. S. (2020). Refutation texts and argumentation for conceptual change: A winning or a redundant combination? *Learning and Instruction*, *65*, 101265. https://doi.org/10.1016/j.learninstruc.2019.101265

*Beerenwinkel, A., Parchmann, I., & Gräsel, C. (2011). Conceptual change texts in chemistry teaching: A study on the particle model of matter. *International Journal of Science and Mathematics Education*, *9*(5), 1235–1259. https://doi.org/10.1007/s10763-010-9257-9

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 221–235). Russell Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons. https://doi.org/10.1002/9780470743386

*Braasch, J. L., Goldman, S. R., & Wiley, J. (2013). The influences of text and reader characteristics on learning from refutations in science texts. *Journal of Educational Psychology*, *105*(3), 561–578. https://doi.org/10.1037/a0032627

*Brandts, J., Busom, I., Lopez-Mayan, C., & Panadés, J. (2019). Dispelling misconceived beliefs about rent control: Insights from a field and a laboratory experiment. *SSRN Electronic Journal*, 1–14. https://doi.org/10.2139/ssrn.3384538

*Broughton, S. H., Sinatra, G. M., & Reynolds, R. E. (2010). The nature of the refutation text effect: An investigation of attention allocation. *The Journal of Educational Research*, *103*(6), 407–423. https://doi.org/10.1080/00220670903383101

Brown, B. A. (2021). *Science in the city: Culturally relevant STEM education*. Harvard Education Press.

Buehl, M. M., Alexander, P. A., Murphy, P. K., & Sperl, C. T. (2001). Profiling persuasion: The role of beliefs, knowledge, and interest in the processing of persuasive texts that vary by argument structure. *Journal of Literacy Research*, *33*(2), 269–301. https://doi.org/10.1080/10862960109548112

*Butterfuss, R. (2020). *Revising misconceptions using multiple documents* [Doctoral dissertation]. https://hdl.handle.net/11299/216409.

Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices*. Heinemann, 88 Post Road West, PO Box 5007, Westport, CT 06881.

*Cameron, K. A., Roloff, M. E., Friesema, E. M., Brown, T., Jovanovic, B. D., Hauber, S., & Baker, D. W. (2013). Patient knowledge and recall of health information following exposure to "facts and myths" message format variations. *Patient Education and Counseling*, *92*(3), 381–387. https://doi.org/10.1016/j.pec.2013.06.017

*Chambers, S. K., & Andre, T. (1997). Gender, prior knowledge, interest, and experience in electricity and conceptual change text manipulations in learning about direct current. *Journal of Research in Science Teaching*, *34*(2), 107–123. https://doi.org/10.1002/(sici)1098-2736(199702)34:23.0.co;2-x

Chi, M. T. (2009). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In *International handbook of research on conceptual change* (pp. 89–110). Routledge.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Clark, R. C., & Mayer, R. E. (2023). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons.

Coburn, K. M. (2018). *A weight-function model for moderators of publication bias*. University of California.

*Cordova, J. R. (2017). *Gender, prior knowledge, interest, and experience in electricity and conceptual change text manipulations in learning about direct current* [Doctoral dissertation]. https://doi.org/10.34917/10985844

Cordova, J., Sinatra, G. M., Jones, S. H., Taasoobshirazi, G., & Lombardi, D. (2014). Self-efficacy, confidence in prior knowledge, and conceptual change. *Contemporary Educational Psychology*, *39*(2), 164–174. https://doi.org/10.1016/j.cedpsych.2014.03.006

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.

Danielson, R. W., Sinatra, G. M., & Kendeou, P. (2016). Augmenting the refutation text effect with analogies and graphics. *Discourse Processes*, *53*(5-6), 392–414. https://doi.org/10.1080/0163853X.2016.1166334

*Dersch, A. S., Renkl, A., & Eitel, A. (2021, August). Do refutation texts foster teachers' conceptual change? Only when designed in a personalized way [Conference session]. European Association for Research on Learning and Instruction, Virtual. https://earli.org/assets/files/BOA-2021.pdf.

DeStefano, F., & Shimabukuro, T. T. (2019). The MMR vaccine and autism. *Annual Review of Virology*, *6*(1), 585–600. https://doi.org/10.1146/annurev-virology-092818-015515

*Diakidoy, I. N., Kendeou, P., & Ioannides, C. (2003). Reading about energy: The effects of text structure in science learning and conceptual change. *Contemporary Educational Psychology*, *28*(3), 335–356. https://doi.org/10.1016/s0361-476x(02)00039-5

*Diakidoy, I. N., Mouskounti, T., & Ioannides, C. (2011). Comprehension and learning from refutation and expository texts. *Reading Research Quarterly*, *46*(1), 22–38. https://doi.org/10.1598/RRQ.46.1.2

*Diakidoy, I. N., Mouskounti, T., Fella, A., & Ioannides, C. (2016). Comprehension processes and outcomes with refutation and expository texts and their contribution to learning. *Learning and Instruction*, *41*, 60–69. https://doi.org/10.1016/j.learninstruc.2015.10.002

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153–170. https://doi.org/10.1016/j.learninstruc.2012.05.003

Dole, J. A., & Sinatra, G. M. (1998). Reconceptualizing Change in the Cognitive Construction of Knowledge. *Educational Psychologist*, *33*(2-3), 109–128. https://doi.org/10.1080/00461520.1998.9653294
Dole, J. A. (2000). Readers, texts and conceptual change learning. *Reading & Writing Quarterly*, *16*(2), 99–118. https://doi.org/10.1080/105735600277980
*Donovan, A. M., Zhan, J., & Rapp, D. N. (2018). Supporting historical understandings with refutation texts. *Contemporary Educational Psychology*, *54*, 1–11. https://doi.org/10.1016/j.cedpsych.2018.04.002
*Ecker, U. K. H. (2017). Why rebuttals may not work: The psychology of misinformation. *Media Asia*, *44*(2), 79–87. https://doi.org/10.1080/01296612.2017.1384145
Ecker, U. K. H., Lewandowsky, S., & Chadwick, M. (2020). Can corrections spread misinformation to new audiences? Testing for the elusive familiarity backfire effect. *Cognitive Research*, *5*(1), 41. https://doi.org/10.1186/s41235-020-00241-6
Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y
Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*(8), 1087–1100. https://doi.org/10.3758/MC.38.8.1087
Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical Research ed.)*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629
*Frède, V. (2008). Teaching astronomy for pre-service elementary teachers: A comparison of methods. *Advances in Space Research*, *42*(11), 1819–1830. https://doi.org/10.1016/j.asr.2007.12.001
Fisher, M. (2022). *The Chaos machine: The inside story of how social media rewired our minds and our world*. Hachette UK.
Funk, C., & Kennedy, B. (2016). *The new food fights: US public divides over food science* (pp. 1–100). Pew Research Center.
*Geller, E. H., & Fukuda, K. (2021). *Surprise, confusion, and conceptual change* [Conference session].
*Gill, M. G., Trevors, G., Greene, J. A., & Algina, J. (2020). Don't take it personally? The role of personal relevance in conceptual change. *The Journal of Experimental Education*, *90*(1), 1–22. https://doi.org/10.1080/00220973.2020.1754152
Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234. https://doi.org/10.3102/0013189X11413260
Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, *36*(2), 193–202. https://doi.org/10.3758/bf03195564
Guzzetti, B. J. (2000). Learning counter-intuitive science concepts: What have we learned from over a decade of research? *Reading & Writing Quarterly*, *16*(2), 89–98. https://doi.org/10.1080/105735600277971
Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly*, *28*(2), 116. https://doi.org/10.2307/747886
Guzzetti, B. J., Williams, W. O., Skeels, S. A., & Wu, S. M. (1997). Influence of text structure on learning counterintuitive physics concepts. *Journal of Research in Science Teaching*, *34*(7), 701–719. https://doi.org/10.1002/(SICI)1098-2736(199709)34:7<701::AID-TEA3>3.0.CO;2-Q
Heddy, B. C., Danielson, R. W., Sinatra, G. M., & Graham, J. (2017). Modifying knowledge, emotions, and attitudes about genetically modified foods. *The Journal of Experimental Education*, *85*(3), 513–533. https://doi.org/10.1080/00220973.2016.1260523
Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107
*Hunsu, N. J. (2016). *Effects of text format, reader factors and comprehension strategies on science learning and conceptual change* [Doctoral dissertation]. https://hdl.handle.net/2376/112001
*Hynd, C., & Alvermann, D. E. (1986). The role of refutation text in overcoming difficulty with science concepts. *Journal of Reading*, *29*(5), 440–446. https://www.jstor.org/stable/40025804
Index, G. S. M. (2020). World Economic Forum. 2020. http://www3.weforum.org/docs/Global_Social_Mobility_Report.pdf
Jacobson, N. G., Thacker, I., & Sinatra, G. M. (2021). Here's hoping it's not just text structure: The role of emotions in knowledge revision and the backfire effect. *Discourse Processes*, *59*(1-2), 13–35. https://doi.org/10.1080/0163853X.2021.1925059
Jin, G., Danielson, R. W., & Ramazan, O. (*under review*). Resurrecting a classic design: Examining the refutational text effect using a solomon four-group design.
Kardash, C. M., & Scholes, R. J. (1996). Effects of preexisiting beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. *Journal of Educational Psychology*, *88*(2), 260–271. https://doi.org/10.1037/0022-0663.88.2.260
Kata, A. (2012). Anti-vaccine activists, Web 2.0, and the postmodern paradigm–An overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine*, *30*(25), 3778–3789. https://doi.org/10.1016/j.vaccine.2011.11.112
Kendeou, P. (2005). *Processing of scientific texts: Effects of prior knowledge and text structure*. University of Minnesota.
Kendeou, P. (2024). A theory of knowledge revision: The development of the KReC framework. *Educational Psychology Review*, *36*(2), 44. https://doi.org/10.1007/s10648-024-09885-y
Kendeou, P., & O'Brien, E. J. (2014). The Knowledge Revision Components (KReC) framework: Processes and mechanisms. In D. N. Rapp & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 353–377). MIT Press.
Kendeou, P., Smith, E. R., & O'Brien, E. J. (2013, May) Updating during reading comprehension: Why causality matters. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(3), 854–865. https://doi.org/10.1037/a0029468
Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014). Knowledge revision processes in refutation texts. *Discourse Processes*, *51*(5-6), 374–397. https://doi.org/10.1080/0163853X.2014.913961
Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition*, *35*(7), 1567–1577. https://doi.org/10.3758/bf03193491
Kessler, E. D., Braasch, J. L., & Kardash, C. M. (2019). Individual differences in revising (and maintaining) accurate and inaccurate beliefs about childhood vaccinations. *Discourse Processes*, *56*(5-6), 415–428. https://doi.org/10.1080/0163853X.2019.1596709
Kortum, P., Edwards, C., & Richards-Kortum, R. (2008). The impact of inaccurate Internet health information in a secondary school learning environment. *Journal of Medical Internet Research*, *10*(2), e986. https://doi.org/10.2196/jmir.986
Kozyreva, A., Lorenz-Spreen, P., Herzog, S. M., Ecker, U. K. H., Lewandowsky, S., Hertwig, R., … Wineburg, S. (2022, December 16). Toolbox of Interventions Against Online Misinformation. https://doi.org/10.31234/osf.io/x8ejt
Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798
*Kutza, D. M. (2000). *The role of prior beliefs, refutational text, intrinsic and extrinsic goals, and extrinsic reward structure, in the conceptual change of preservice teachers (9984446)* [Doctoral dissertation]. ProQuest Dissertations and Theses Global. https://www.proquest.com/openview/1137bfa5d95995546da6c27622efb7f3/1?pq-origsite=gscholar&cbl=18750&diss=y.
*Lally, M., Goldsworthy, R., Sarr, M., Kahn, J., Brown, L., Peralta, L., & Zimet, G. (2014). Evaluation of an intervention among adolescents to reduce preventive misconception in HIV vaccine clinical trials. *The Journal of Adolescent Health*, *55*(2), 254–259. https://doi.org/10.1016/j.jadohealth.2014.01.006
Lassonde, K. A., Kendeou, P., & O'Brien, E. J. (2016). Refutation texts: Overcoming psychology misconceptions that are resistant to change. *Scholarship of Teaching and Learning in Psychology*, *2*(1), 62–74. https://doi.org/10.1037/stl0000054

Lem, S., Baert, K., Ceulemans, E., Onghena, P., Verschaffel, L., & Van Dooren, W. (2017). Refutational text and multiple external representations as a method to remediate the misinterpretation of box plots. *Educational Psychology*, *37*(10), 1281–1300. https://doi.org/10.1080/01443410.2017.1333574

*Lem, S., Kempen, G., Ceulemans, E., Onghena, P., Verschaffel, L., & Van Dooren, W. (2015). Combining multiple external representations and Refutational text: An intervention on learning to interpret box plots. *International Journal of Science and Mathematics Education*, *13*(4), 909–926. https://doi.org/10.1007/s10763-014-9604-3

*Lem, S., Onghena, P., Verschaffel, L., & Van Dooren, W. (2016). The power of refutational text: Changing intuitions about the interpretation of box plots. *European Journal of Psychology of Education*, *32*(4), 537–550. https://doi.org/10.1007/s10212-016-0320-y

*Leon, T. M. (2009). *Middle-school students comprehending, analyzing, and evaluating persuasive texts* [Doctoral dissertation]. http://hdl.handle.net/1903/8480.

Leiserowitz, A., Maibach, E., Rosenthal, S., Kotcher, J., Lee, S., Verner, M., Ballew, M., Carman, J., Myers, T., Goldberg, M., Badullovich, N., & Marlon, J. (2023). *Climate Change in the American Mind: Beliefs & Attitudes, Spring 2023*. Yale University and George Mason University. Yale Program on Climate Change Communication.

Lewandowsky, S., Cook, J., Ecker, U., Albarracín, D., Amazeen, M., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E., Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., … Zaragoza, M. S. (2020). *The Debunking Handbook 2020*.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Li, W., Wang, F., Mayer, R. E., & Liu, H. (2019). Getting the point: Which kinds of gestures by pedagogical agents improve multimedia learning? *Journal of Educational Psychology*, *111*(8), 1382–1395. https://doi.org/10.1037/edu0000352

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*, *339*(jul21 1), b2700–b2700. https://doi.org/10.1136/bmj.b2700

*Lin, Y. C. (2011). *Comprehension of scientific texts during reading: The effects of prior knowledge and text structure* [Master's thesis]. https://etda.libraries.psu.edu/catalog/11557.

*Liu, Q., & Nesbit, J. C. (2018). Conceptual change with refutational maps. *International Journal of Science Education*, *40*(16), 1980–1998. https://doi.org/10.1080/09500693.2018.1515513

Lippmann, M., Danielson, R. W., Schwartz, N. H., Körndle, H., & Narciss, S. (2021). Effects of keyword tasks and biasing titles on metacognitive monitoring and recall. *Metacognition and Learning*, *16*(1), 233–253. https://doi.org/10.1007/s11409-020-09246-4

Lipsey, M. W., & Wilson, D. (2001). *Practical meta-analysis*. Sage Publications Inc.

Lombardi, D., Shipley, T. F., Bailey, J. M., Bretones, P. S., Prather, E. E., Ballen, C. J., Knight, J. K., Smith, M. K., Stowe, R. L., Cooper, M. M., Prince, M., Atit, K., Uttal, D. H., LaDue, N. D., McNeal, P. M., Ryker, K., St. John, K., van der Hoeven Kraft, K. J., & Docktor, J. L. (2021). The curious construct of active learning. *Psychological Science in the Public Interest*, *22*(1), 8–43. https://doi.org/10.1177/1529100620973974

Lombardi, D., Sinatra, G. M., & Nussbaum, E. M. (2013). Plausibility reappraisals and shifts in middle school students' climate change conceptions. *Learning and Instruction*, *27*, 50–62. https://doi.org/10.1016/j.learninstruc.2013.03.001

*Lombardi, D., Danielson, R. W., & Young, N. (2016). A plausible connection: Models examining the relations between evaluation, plausibility, and the refutation text effect. *Learning and Instruction*, *44*, 74–86. https://doi.org/10.1016/j.learninstruc.2016.03.003

Lombardi, D., Matewos, A. M., Jaffe, J., Zohery, V., Mohan, S., Bock, K., & Jamani, S. (2022). Discourse and agency during Scaffolded Middle School Science Instruction. *Discourse Processes*, *59*(5-6), 379–400. https://doi.org/10.1080/0163853X.2022.2068317

*Maria, K., & Johnson, J. M. (1990). *Correcting misconceptions: Effect of type of text*. National Reading Conference Yearbook. https://psycnet.apa.org/record/1991-15002-001

*Mason, L., Baldi, R., Di Ronco, S., Scrimin, S., Danielson, R. W., & Sinatra, G. M. (2017). Textual and graphical refutations: Effects on conceptual change learning. *Contemporary Educational Psychology*, *49*, 275–288. https://doi.org/10.1016/j.cedpsych.2017.03.007

*Mason, L., Zaccoletti, S., Carretti, B., Scrimin, S., & Diakidoy, I. N. (2019). The role of inhibition in conceptual learning from refutation and standard expository texts. *International Journal of Science and Mathematics Education*, *17*(3), 483–501. https://doi.org/10.1007/s10763-017-9874-7

Mayer, R. E. (2002). Multimedia learning. In *Psychology of learning and motivation* (Vol. 41, pp. 85–139). Academic Press.

Mayer, R. E. (2009). *Multimedia learning*. Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Mayer, R. E. (2019). Thirty years of research on online learning. *Applied Cognitive Psychology*, *33*(2), 152–159. https://doi.org/10.1002/acp.3482

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, *86*(3), 389–401. https://doi.org/10.1037/0022-0663.86.3.389

McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In *Applied natural language processing: Identification, investigation and resolution* (pp. 188–205). IGI Global.

*McCuin, J. L., Hayhoe, K., & Hayhoe, D. (2014). Comparing the effects of traditional vs. misconceptions-based instruction on student understanding of the greenhouse effect. *Journal of Geoscience Education*, *62*(3), 445–459. https://doi.org/10.5408/13-068.1

*Mikkilä-Erdmann, M. (2001). Improving conceptual change concerning photosynthesis through text design. *Learning and Instruction*, *11*(3), 241–257. https://doi.org/10.1016/s0959-4752(00)00041-4

*Mikkilä-Erdmann, M. (2002). Science learning through text: The effect of text design and text comprehension skills on conceptual change. In M. Limón & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 337–353). Springer Science & Business Media. https://doi.org/10.1007/0-306-47637-1_17

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105–125. https://doi.org/10.1037/1082-989x.7.1.105

*Muis, K. R., Etoubashi, N., & Denton, C. A. (2020). The catcher in the lie: The role of emotions and epistemic judgments in changing students' misconceptions and attitudes in a post-truth era. *Contemporary Educational Psychology*, *62*, 101898. https://doi.org/10.1016/j.cedpsych.2020.101898

*Muis, K. R., Sinatra, G. M., Pekrun, R., Winne, P. H., Trevors, G., Losenno, K. M., & Munzar, B. (2018). Main and moderator effects of refutation on task value, epistemic emotions, and learning strategies during conceptual change. *Contemporary Educational Psychology*, *55*, 155–165. https://doi.org/10.1016/j.cedpsych.2018.10.001

Munnich, E., & Ranney, M. A. (2019). Learning from surprise: Harnessing a metacognitive surprise signal to build and adapt belief networks. *Topics in Cognitive Science*, *11*(1), 164–177. https://doi.org/10.1111/tops.12397

*Nasim, A. M. (2018). *Using refutation texts to change attitudes and knowledge concerning auditory verbal hallucinations (10813433)* [Doctoral dissertation]. ProQuest Dissertations and Theses Global. https://media.proquest.com/media/hms/ORIG/2/4PUsG?_s=vKxbjOZBrKJzkc1D5D3S6U2IOK8%3D.

National Research Council. (2013). *Next generation science standards: For states, by states*. NRC.

Osborne, J., Pimentel, D., Alberts, B., Allchin, D., Barzilai, S., Bergstrom, C., … Wineburg, S. (2022). Science Education in an Age of Misinformation.

*Palmer, D. H. (2003). Investigating the relationship between refutational text and conceptual change. *Science Education*, *87*(5), 663–684. https://doi.org/10.1002/sce.1056

Paynter, J., Luskin-Saxby, S., Keen, D., Fordyce, K., Frost, G., Imms, C., Miller, S., Trembath, D., Tucker, M., & Ecker, U. (2019). Evaluation of a template for countering misinformation—Real-world Autism treatment myth debunking. *PLoS One*, *14*(1), e0210746. https://doi.org/10.1371/journal.pone.0210746

*Peltier, T. K., Heddy, B. C., & Peltier, C. (2020). Using conceptual change theory to help preservice teachers understand dyslexia. *Annals of Dyslexia*, *70*(1), 62–78. https://doi.org/10.1007/s11881-020-00192-z

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, *63*(2), 167–199. https://doi.org/10.3102/00346543063002167

*Poehnl, S., & Bogner, F. X. (2013). Cognitive load and alternative conceptions in learning genetics: Effects from provoking confusion. *The Journal of Educational Research*, *106*(3), 183–196. https://doi.org/10.1080/00220671.2012.687790

Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research*, *86*(1), 207–236. https://doi.org/10.3102/0034654315582067

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

*Prinz, A., Golke, S., & Wittwer, J. (2017, July). Refuting overconfidence: Refutation texts prevent detrimental effects of misconceptions on text comprehension and metacomprehension accuracy in the domain of statistics [Paper presentation]. The Cognitive Science Society. https://cogsci.mindmodeling.org/2017/papers/0555/paper0555.pdf.

*Prinz, A., Kollmer, J., Flick, L., Renkl, A., & Eitel, A. (2020). Refuting student teachers' misconceptions about multimedia learning. *Instructional Science*, *50*(1), 89–110. https://doi.org/10.1007/s11251-021-09568-z

Pustejovsky, J. E. (2019). *clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections* (Version 0.5.3). The Comprehensive R Archive Network. https://CRAN.R-project.org/package=clubSandwich.

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science*, *23*(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3

Rabesandratana. (2019). These are the countries that trust scientists the most – and the least. Science. https://www.science.org/content/article/global-survey-finds-strong-support-scientists

Ranney, M. A., & Clark, D. (2016). Climate change conceptual change: Scientific information can transform attitudes. *Topics in Cognitive Science*, *8*(1), 49–75. https://doi.org/10.1111/tops.12187

*Ramsburg, J. T. (2018). [*The lesson and the learner: The effect of individual differences and task scaffolding on category learning*]. [Doctoral dissertation]. https://hdl.handle.net/10027/21199

Rodgers, M. A., & Pustejovsky, J. E. (2020). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, Advance online publication. https://doi.org/10.1037/met0000300

*Rodrigues, J., & Thacker, I. (2019). Refuting a fraction misconception: A brief intervention promotes teachers' conceptual change. Conference Papers – Psychology of Mathematics & Education of North America, 781–785.

*Salisbury-Glennon, J. D., & J. Stevens, R. (1999). Addressing preservice teachers' conceptions of motivation. *Teaching and Teacher Education*, *15*(7), 741–752. https://doi.org/10.1016/s0742-051x(99)00023-2

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Shwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632–654. https://doi.org/10.1002/tea.20311

Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(16), 7662–7669. https://doi.org/10.1073/pnas.1805871115

Schroeder, N. L., & Kucera, A. C. (2022). Refutation text facilitates learning: A meta-analysis of between-subjects experiments. *Educational Psychology Review*, *34*(2), 957–987. https://doi.org/10.1007/s10648-021-09656-z

Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology*, *39*, 127–161.

*Seyedmonir, M. (2000). *The development and validation of Science Learning Inventory (SLI): A conceptual change framework (3039331)* [Doctoral dissertation]. ProQuest Dissertations and Theses Global. https://www.proquest.com/docview/304632635

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2022). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist*, *57*(1), 31–54. https://doi.org/10.1080/00461520.2021.1939700

Sinatra, G. M. (2005). The "Warming Trend" in conceptual change research: The Legacy of Paul R. Pintrich. *Educational Psychologist*, *40*(2), 107–115. https://doi.org/10.1207/s15326985ep4002_5

Sinatra, G. M. (2022). Motivational and emotional impacts on public (mis) understanding of science. *Educational Psychologist*, *57*(1), 1–10. https://doi.org/10.1080/00461520.2021.1975121

Sinatra, G. M., & Broughton, S. H. (2011). Bridging reading comprehension and conceptual change in science education: The promise of refutation text. *Reading Research Quarterly*, *46*(4), 374–393. https://doi.org/10.1002/RRQ.005

Sinatra, G. M., & Chinn, C. (2011). Thinking and reasoning in science: Promoting epistemic conceptual change. In K. Harris, S. Graham, T. Urdan (Eds.), *Application to learning and teaching. APA Educational Psychology Handbook* (Volume 3, pp. 257–282). APA Publications.

Sinatra, G. M., & Mason, L. (2013). Beyond knowledge: Learner characteristics influencing conceptual change. In *International handbook of research on conceptual change* (pp. 377–394). Routledge.

*Sota, M. (2012). *The effect of contrasting analogies on understanding of and reasoning about natural selection* [Doctoral dissertation]. https://diginole.lib.fsu.edu/islandora/object/fsu%3A183124/datastream/PDF/view.

St. Pierre, R. G (2001). Random assignment: Implementation in complex field settings. In J. S. Neil & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences*. Pergamon.

*Suma, K., Suriyasmini, N. M., & Pujani, N. M. (2018). The effect of conceptual change text on improving student understanding of electricity concepts and learning motivation. *International Research Journal of Engineering, IT & Scientific Research*, *4*(6), 33–43. https://doi.org/10.21744/irjeis.v4n6.337

*Södervik, I., Mikkilä-Erdmann, M., & Vilppu, H. (2014). Promoting the understanding of photosynthesis among elementary school student teachers through text design. *Journal of Science Teacher Education*, *25*(5), 581–600. https://doi.org/10.1007/s10972-013-9373-9

*Södervik, I., Virtanen, V., & Mikkilä-Erdmann, M. (2014). Challenges in understanding photosynthesis in a university introductory biosciences class. *International Journal of Science and Mathematics Education*, *13*(4), 733–750. https://doi.org/10.1007/s10763-014-9571-8

Sychev, O., Anikin, A., & Prokudin, A. (2020). Automatic grading and hinting in open-ended text questions. *Cognitive Systems Research*, *59*, 264–272. https://doi.org/10.1016/j.cogsys.2019.09.025

*Taşdelen, U., & Köseoğlu, F. (2008). Learner-friendly textbooks: Chemistry texts based on a constructivist view of learning. *Asia Pacific Education Review*, *9*(2), 136–147. https://doi.org/10.1007/BF03026494

Thacker, I., Sinatra, G. M., Muis, K. R., Danielson, R. W., Winne, P., Pekrun, R., & Chevrier, M. (2020). Using persuasive refutation texts to prompt attitudinal and conceptual change. *Journal of Educational Psychology*, *112*(6), 1085–1099. https://doi.org/10.1037/edu0000434

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, *12*(3), 435–467. https://doi.org/10.1017/S0140525X00057046

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

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375–393. https://doi.org/10.1037/met0000011

Treagust, D., & Duit, R. (2008). Conceptual change: A discussion of theoretical, methodological and practical challenges for science education. *Cultural Studies of Science Education*, *3*(2), 297–328. https://doi.org/10.1007/s11422-008-9090-4

*Trevors, G. (2012). *Learner, text, and context factors on conceptual change in biology* [Master's thesis]. https://escholarship.mcgill.ca/concern/theses/6m311t49g.

*Trevors, G. (2016). *Controversial science knowledge: A multi-study examination of how epistemic cognition relates to the ways we learn science* [Doctoral dissertation]. https://escholarship.mcgill.ca/concern/theses/cz30pw462.

Trevors, G. J. (2022). The roles of identity conflict, emotion, and threat in learning from refutation texts on vaccination and immigration. *Discourse Processes*, *59*(1-2), 36–51. https://doi.org/10.1080/0163853X.2021.1917950

Trevors, G., & Kendeou, P. (2020). The effects of positive and negative emotional text content on knowledge revision. *Quarterly Journal of Experimental Psychology*, *73*(9), 1326–1339. https://doi.org/10.1177/1747021820913816

Trevors, G., & Ladhani, F. (2022). It's contagious! Examining gamified refutation texts, emotions, and knowledge retention in a real-world public health education campaign. *Discourse Processes*, *59*(5-6), 401–416. https://doi.org/10.1080/0163853X.2022.2085477

*Trevors, G., & Muis, K. R. (2015). Effects of text structure, reading goals and epistemic beliefs on conceptual change. *Journal of Research in Reading*, *38*(4), 361–386. https://doi.org/10.1111/1467-9817.12031

Trevors, G. J., Muis, K. R., Pekrun, R., Sinatra, G. M., & Winne, P. H. (2016). Identify and epistemic emotions during knowledge revision: A potential account for the backfire effect. *Discourse Processes*, *53*(5-6), 339–370. https://doi.org/10.1080/0163853X.2015.1136507

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2, pp. 131–160). Addison-wesley.

Van den Akker, O., Weston, S. J., Campbell, L., Chopik, W. J., Damian, R. I., Davis-Kean, P., … Bakker, M. (2019, November 20). *Preregistration of Secondary Data Analysis: A Template and Tutorial.* https://doi.org/10.31234/osf.io/hvfmr

van den Akker, O., Weston, S., Campbell, L., Chopik, B., Damian, R., Davis-Kean, P., Hall, A., Kosie, J., Kruse, E., Olsen, J., Ritchie, S., Valentine, K. D., Veer, A. V., & Bakker, M. (2021). Preregistration of secondary data analysis: A template and tutorial. *Meta-Psychology*, *5*, 2625. https://doi.org/10.15626/MP.2020.2625

*Vaughn, A. R., & Johnson, M. L. (2018). Communicating and enhancing teachers' attitudes and understanding of influenza using refutational text. *Vaccine*, *36*(48), 7306–7315. https://doi.org/10.1016/j.vaccine.2018.10.040

Vembye, M. H., Pustejovsky, J. E., & Pigott, T. (2022, January 6). Power approximations for overall average effects in meta-analysis with dependent effect sizes. https://doi.org/10.3102/10769986221127379

*Versteeg, M., Van Loon, M. H., Wijnen-Meijer, M., & Steendijk, P. (2020). Refuting misconceptions in medical physiology. *BMC Medical Education*, *20*(1), 250. https://doi.org/10.1186/s12909-020-02166-6

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*(3), 419–435. https://doi.org/10.1007/BF02294384

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

*Vilppu, H., Mikkilä-Erdmann, M., & Ahopelto, I. (2013). The role of regulation and processing strategies in understanding science text among University students. *Scandinavian Journal of Educational Research*, *57*(3), 246–262. https://doi.org/10.1080/00313831.2011.637229

*Vivian, J. A. (2018). *The mediating roles of attitudes and epistemic beliefs in knowledge revision* [Master's thesis]. https://escholarship.mcgill.ca/downloads/70795b01d.

Vogl, E., Pekrun, R., Murayama, K., Loderer, K., & Schubert, S. (2019). Surprise, curiosity, and confusion promote knowledge exploration: Evidence for robust effects of epistemic emotions. *Frontiers in Psychology*, *10*, 2474. https://doi.org/10.3389/fpsyg.2019.02474

Vosniadou, S. (Ed.). (2008). *International handbook of research on conceptual change* (1st ed.). Erlbaum.

*Wang, T., & Andre, T. (1991). Conceptual change text versus traditional text and application questions versus no questions in learning about electricity. *Contemporary Educational Psychology*, *16*(2), 103–116. https://doi.org/10.1016/0361-476x(91)90031-f

*Watanabe, M. N. (2020). *Enhancing students' ability to correct misconceptions in natural selection with refutational texts and self-explanation training* [Master's thesis]. https://keep.lib.asu.edu/_flysystem/fedora/c7/234442/Watanabe_asu_0010N_20471.pdf.

*Weingartner, K. M., & Masnick, A. M. (2019). Refutation texts: Implying the refutation of a scientific misconception can facilitate knowledge revision. *Contemporary Educational Psychology*, *58*, 138–148. https://doi.org/10.1016/j.cedpsych.2019.03.004

Wellcome global monitor. (2018). How does the world feel about science and health? https://wellcome.org/sites/default/files/wellcome-global-monitor-2018.pdf

Will, K. K., Masad, A., Vlach, H. A., & Kendeou, P. (2019). The effects of refutation texts on generating explanations. *Learning and Individual Differences*, *69*, 108–115. https://doi.org/10.1016/j.lindif.2018.12.002

Windschitl, M., Thompson, J., & Braaten, M. (2020). *Ambitious science teaching*. Harvard Education Press.

World Health Organization. (2024, April). Measles. *World Health Organization (WHO).* https://www.who.int/news-room/fact-sheets/detail/measles.

*Yürük, N., & Eroğlu, P. (2016). The effect of conceptual change texts enriched with metaconceptual processes on pre-service science teachers' conceptual understanding of heat and temperature. *Journal of Baltic Science Education*, *15*(6), 693–705. https://doi.org/10.33225/jbse/16.15.693

Zengilowski, A., Schuetze, B. A., Nash, B. L., & Schallert, D. L. (2021). A critical review of the refutation text literature: Methodological confounds, theoretical problems, and possible solutions. *Educational Psychologist*, *56*(3), 175–195. https://doi.org/10.1080/00461520.2020.1861948