



Refutation Text Facilitates Learning: a Meta-Analysis of Between-Subjects Experiments

Noah L. Schroeder¹ · Aurelia C. Kucera²

Accepted: 29 December 2021 / Published online: 24 January 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Scientific misconceptions are ubiquitous, and in our era of near-instant information exchange, this can be problematic for both public health and the public understanding of scientific topics. Refutation text is one instructional tool for addressing misconceptions and is simple to implement at little cost. We conducted a random-effects meta-analysis to examine the effectiveness of the refutation text structure on learning. Analysis of 44 independent comparisons ($n=3,869$) showed that refutation text is associated with a positive, moderate effect ($g=0.41$, $p<.001$) compared to other learning conditions. This effect was consistent and robust across a wide variety of contexts. Our results support the implementation of refutation text to help facilitate scientific understanding in many fields.

Keywords Refutation text · Meta-analysis · Conceptual change · Learning · Refutation

Misconceptions are rampant in daily life. This is particularly true in the age of the internet and smart devices, as individuals have rapid access to a wide variety of resources of varying degrees of reputability. Individuals may engage with inaccurate information and then build inaccurate conceptions based on this inaccurate information. For the purposes of this paper, we will refer to inaccurate conceptions as misconceptions. In some cases, the implications of misconceptions may be minimal. For example, not having a scientifically accepted understanding of why tides occur may not tangibility influence an inland city dweller's daily life. However, in other cases, there can be tangible costs to failing to hold scientifically accepted understandings

✉ Noah L. Schroeder
noah.schroeder@wright.edu

¹ Department of Leadership Studies in Education and Organizations, Wright State University, 3640 Colonel Glenn Highway, Dayton, OH 45435, USA

² Lake Erie College of Osteopathic Medicine at Seton Hill, 20 Seton Hill Dr, Greensburg, PA 15601, USA

of some topics. For example, the world is currently fighting the coronavirus disease 2019 (COVID-19) and its causative agent, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Misconceptions about this disease may have disastrous consequences, particularly if they are held and promoted by people involved in public health policy. Reputable sources of health information are attempting to combat myths about SARS-CoV-2 through websites with information presented in an expository format (Mayo Clinic Staff, 2020; World Health Organization, n.d.). Unfortunately, expository text may not be the most effective approach to provide these scientific facts, as it does not always promote learning effectively or efficiently (Sinatra et al., 2011). We then face a critical question: how should we address misconceptions and help students obtain scientifically accepted understandings?

Misconceptions are inaccurate ideas and beliefs (Kendeou & Van Den Broek, 2007; Tippett, 2010; Will et al., 2019) which can influence an individual's conceptual change and learning (Dole & Niederhauser, 1990). Conceptual change is thought to occur spontaneously or intentionally (Sinatra et al., 2011), and the extent to which it occurs is context-dependent (Posner et al., 1982). However, the process by which conceptual change occurs is incompletely understood (Rusanen, 2014). A classic model of the conceptual change process requires the learner to realize the information they hold is incorrect (*dissatisfaction*) and then be presented with *understandable* and *plausible* alternative information that offers some benefit for explaining the phenomenon (*fruitful*) (Pintrich et al., 1993; Posner et al., 1982). It is also likely that social, emotional, and motivational variables can influence the conceptual change process, and the interaction between these factors is complex (Pintrich et al., 1993; Sinatra, 2005). As such, it is important to identify instructional strategies that can facilitate the conceptual change process.

A variety of different text formats have been used to facilitate conceptual change and learning more broadly. For instance, expository texts are based on scientific facts, as one may often see in a textbook. Conceptual change texts are slightly different, beginning with a question and following it with explanations of why common misconceptions are incorrect and providing the scientifically accepted explanation (Çalik et al., 2007; Roth, 1985). Similar to conceptual change text is another promising text format, refutation text, which is the focus of this research. A refutation text states a common misconception about a subject, refutes it, and provides an explanation of the correct conception (Hynd, 2001; Mason et al., 2019; Tippett, 2010). Importantly, there is occasionally overlap between the format of a conceptual change text and a refutation text, with the difference being that conceptual change texts ask learners to make predictions by posing a question, whereas a refutation text does not require this (Chambers & Andre, 1997). An example of a refutation text is as follows:

Many people believe that the peregrine falcon is on the United States Fish and Wildlife Service's (USFWS) endangered and threatened species list. However, this is not true. The last subspecies of peregrine falcon were removed from the USFWS endangered and threatened species list in 1999.

Refutation text has been investigated in relation to the learning of general, non-politicized, and well-accepted science concepts such as tides (Ariasi & Mason, 2014) and energy (Diakidoy et al., 2016). Recently however, researchers have begun investigating more politicized subjects (Zengilowski et al., 2021) and medical topics such as influenza (Vaughn & Johnson, 2018). In this meta-analysis, we broadly analyze the work around refutation text to build an understanding of under what conditions and for whom it is an effective instructional strategy.

Literature Review

The Conceptual Change Process

There are a variety of theoretical perspectives as to how conceptual change may occur, ranging from strictly cognition-oriented theories through those that incorporate aspects of affect or other aspects of their persons and varying from epistemological views to ontological views (for a history and review, see Treagust & Duit, 2008). Since the purpose of this paper is not to validate one theory as opposed to another, we refer readers to existing works dedicated to these theories (see, for example, Pintrich et al., 1993; Posner et al., 1982; Sinatra, 2005; Sinatra & Broughton, 2011; Treagust & Duit, 2008; Vosniadou, 2008).

For the purposes of this paper, it is important to broadly understand the theorized cognitive mechanisms behind how learners' mental models can change over time. This includes transitions from inaccurate preconceptions through scientifically accepted understandings. As such, rather than explore the many different views of conceptual change present in the literature (e.g., Pintrich et al., 1993; Posner et al., 1982; Sinatra, 2005; Sinatra & Broughton, 2011; Treagust & Duit, 2008; Vosniadou, 2008), we focus on one cognition-oriented framework that explains why refutation text may be effective for supporting conceptual change and learning and is largely complementary to the aforementioned explanations of conceptual change that incorporate social, emotional, and motivational variables.

The Knowledge Revision Components Framework (KReC) (Kendeou et al., 2014, 2019; Will et al., 2019) provides a cognition-oriented model of knowledge revision. The KReC framework is based on two assumptions. First, information within the long-term memory cannot be eradicated; it is always present. Second, this information held in the long-term memory can be activated by a learner as they read a text. These are known as the encoding and passive activation principles, respectively (Kendeou et al., 2014, 2019; Will et al., 2019). For knowledge revision to occur, the KReC framework suggests that novel information gained from text must interact with existing schema in the working memory. This activation of long-term memories can promote integration of the novel information with the learner's prior knowledge, thus creating an integrated representation in the learner's mental model (Kendeou et al., 2014, 2019; Will et al., 2019). KReC posits that the new information and prior knowledge then compete for activation, and knowledge revision occurs when the new information wins this competition for activation (Kendeou et al., 2014, 2019; Will et al., 2019).

With this cognition-oriented understanding of how knowledge revision may occur, a critical question remains: how does refutation text facilitate knowledge revision? Recall that refutation text states a common misconception about a subject, refutes it, and provides an explanation of the correct conception (Hynd, 2001; Mason et al., 2019; Tippett, 2010). If one were to align this with the KReC framework, presumably the statement of the misconception directly activates the existing schema, and an alternative conception is immediately offered and explained. Butterfuss and Kendeou (2020) suggested that this format “draws activation towards the correct idea and away from the misconception” (p. 790). Thus, conceptual change can result.

Learning with Refutation Text

What do we know about learning with refutation texts? Examining the research syntheses that have occurred, we can gain insights from both meta-analyses and qualitative reviews. For instance, Guzzetti et al. (1993) meta-analyzed the refutation text literature and concluded that “text can be used effectively to eradicate misconceptions either when text is refutational or when text is used in combination with other strategies that cause cognitive conflict” (p. 130). These results were promising, and a more recent meta-analysis of studies published between 1995 and 2010 examined conceptual change texts and found a very strong overall effect (Armağan et al., 2017). However, as noted, conceptual change texts are not the same as refutation texts, and this meta-analysis was not very clearly reported (e.g., the abstract and results report different overall effect sizes).

When we examine qualitative review efforts rather than meta-analyses, we gain additional understandings as to when and to what extent refutation texts may facilitate learning. For instance, Tippett’s (2010) review highlighted differential effects of refutation text depending on the age group reading the texts, and Zengilowski et al. (2021) noted that there are a number of variables that may influence the effectiveness of refutation texts.

While research synthesis efforts have supported the idea that refutation text can facilitate learning in some situations, it seems as though there are variables that may moderate this effect. Throughout this paper, we categorize these potentially moderating variables as features of the text and features of the research.

Features of the Text that May Influence Learning

We first examine the features of the text itself that may influence learning outcomes. These variables include the domain of the learning materials, the use of additional materials, the format of the instructional media, the length of the text, and the reading level of the text.

The Domain of the Learning Materials Zengilowski et al. (2021) noted that the refutation text literature often examines only one particular refutation text compared to a comparable expository text rather than testing the effects of refutation texts across

multiple texts. They further highlighted how, in many cases, the same few topics were being examined. Through meta-analysis, we can broadly examine the effects of refutation texts across various subject domains, as we will in this study. However, we note, as Zengilowski et al. did, that this is only a partial solution to the problem of many studies only examining one text in relation to another comparable text. But alas, a moderation analysis examining the effects of refutation texts across various knowledge domains would be a starting point which may provide insights into if there are varied outcomes depending on the knowledge domain.

The Use of Additional Learning Materials In their review of the refutation text literature, Tippett (2010) noted that it is important for researchers to examine the effects of refutation texts when they are accompanied by images. This suggestion is supported by the multimedia principle, which suggests that people learn more when words and images are present compared to only text or narration (Mayer, 2017, 2021). The multimedia principle has been found to be robust in the multimedia learning literature; thus, it is an open question as to whether providing images will add benefits to refutation texts.

The Format of the Instructional Media Recent meta-analyses have found that reading from screens was not as effective as reading from paper (Clinton, 2019; Kong et al., 2018). However, Clinton (2019) found that this effect was moderated by the type of text. More specifically, expository texts were associated with lower (i.e., worse) effects than narrative texts when reading on an electronic device. Given that refutation texts are in some ways more similar to expository texts than narrative texts, in this meta-analysis, we examine if there are differential effects of refutation text depending on whether it is presented on paper or electronically.

The Length of the Text One design limitation of refutation text is that it is often longer than expository texts. One potential view on this phenomenon is that a longer text may encourage more time engaged with the content, which may aid learning. Alternatively, refutation texts may provide additional information (e.g., a misconception and direct refutation) not present in the expository text, in addition to all the text in the expository text. It seems plausible that this extra amount of information may benefit learning, as opposed to the actual misconception and refutation itself. An open question then is whether refutation texts which are of similar length to the control condition texts vary in effectiveness compared to those that are longer. If so, it would imply that it is the text structure that facilitates learning rather than the simple addition of more information. Thus, we sought to investigate how a difference in length between the refutation text and the control condition moderated the effectiveness of refutation text.

The Reading Level of the Text A final feature of the text is the reading level of the text. It seems plausible that texts which are lower in reading level may be easier to understand, whereas those that are higher in reading level are more

complex. We question whether refutation text works differentially across these textual complexities.

Features of the Research Design that May Influence Learning

Researchers have also used a variety of research designs when investigating the use of refutation text. Thus, we sought to examine how features of the research design moderate learning with refutation text by investigating the following variables: the comparison condition, the age of the learner, the learners' prior knowledge, how participants were assigned to condition, the type of outcome test, the timing of the outcome test, the type of outcome test, and the location of the study.

The Comparison Condition While dated meta-analytic results have shown that refutation text can facilitate learning (Guzzetti et al., 1993), one outstanding question is how the effects of refutation text vary compared to other specific interventions such as narrative or expository texts. Tippett (2010) found that learners preferred refutation text to expository text, but it is unclear whether this carries over to learning outcomes. Further, it is not clear to what extent refutation text may benefit, or hinder, learning compared to other interventions, such as narrative texts.

The Age of the Learner Given that refutation texts are simple to create, a critical question is for what age groups are they effective? Tippett's (2010) review of 22 studies provides some initial insights into this question. Tippett found that only one study investigated the use of refutation text in grades K-2, and it did not find any significant benefit to reading a refutation text. Similarly, 11 studies investigated refutation texts with students in grades 11 or higher, and they also found no significant benefit. However, 10 studies investigated the effects of refutation texts with students in grades 3–10, and refutation text was found to be more likely to result in conceptual change than expository texts (Tippett, 2010). Thus, a critical question remains: a decade later, do we still see differential patterns of effectiveness from refutation text depending on the learners' age?

The Learner's Prior Knowledge As noted above, a learner's existing misconceptions can influence an individual's conceptual change and learning (Dole & Niederhauser, 1990). These misconceptions represent a learner's prior knowledge and beliefs about a subject. Thus, it is reasonable, and perhaps expected, that a learner's prior knowledge may significantly moderate the effects of refutation text on learning.

How Participants Were Assigned to Conditions It is plausible that the experimental design could have some influence on the outcomes found in an experiment. This is important contextual information to understand when interpreting a meta-analysis. For instance, if studies that used non-random and non-stratified assignment are associated with statistically higher effects than those that used random assignment, it may indicate that methodological choices in study design may be attributable to the

observed effects rather than the intervention. Accordingly, in this meta-analysis, we examine how the type of participant assignment influenced learning from refutation texts.

The Type of Outcome Test Various types of tests have been used to measure learning in the refutation text literature. It seems plausible that different types of tests may illuminate differential benefits of refutation text. For instance, if multiple choice retention tests were associated with small effects, while open-ended transfer tasks were associated with larger effects, this may imply that while refutation text may benefit retention to some degree, its benefits can be more strongly seen in tests of learning transfer. Given that many researchers may focus on how well students transfer knowledge to new domains whereas others are more interested in an immediate change in understanding, in this meta-analysis, we examine if the type of test moderates the effectiveness of refutation text.

The Timing of the Outcome Test A critical question around the efficacy of any educational intervention is how well the benefits of the intervention persist over time. Tippett (2010) found in their review that refutation text maintained its effectiveness over time. However, in their critical review of the refutation text literature, Zengilowski et al. (2021) found that the evidence was not so clear cut, having found evidence that both supported and undermined the learning benefits of refutation text over time. Thus, it seems pertinent to examine whether the effects of refutation text are maintained over time, and if so, how well they are maintained.

The Type of Publication Publication bias is a consistent issue in meta-analysis, as researchers have argued that statistically significant results are more likely to be published than non-significant results (Rosenthal, 1979). As such, in this study, we examine whether the type of publication, such as a journal article, conference paper, or dissertation, is associated with statistically different effects on learning.

The Location of the Study Different cultures often have varied norms and expectations as well as different written languages. It seems plausible that these may carry over into learning environments. In an effort to broadly analyze whether locale influences learning with refutation text, we categorize studies by continent in which the study took place. We acknowledge that this is a coarse measurement, but it provides a starting point to look for trends, which, if found, could lead to a variety of potentially interesting research questions.

As shown, there are many factors that could potentially influence the efficacy of refutation text. The purpose of this meta-analysis is to help understand the effects of refutation text across these various potentially moderating variables.

The Present Meta-analysis

In this era of near-instant information exchange, misinformation and misconceptions can spread rapidly. Refutation text is a simple and low-cost intervention to help correct misconceptions which could easily be implemented across print and electronic media, but research on the effects of conceptual change texts has not been meta-analyzed in more than 10 years (Armağan et al., 2017, whose literature search was conducted in 2010); the effects of refutation texts specifically have not been quantitatively synthesized in nearly 20 years (Guzzetti et al., 1992, 1993), nor qualitatively synthesized in ten (Tippett, 2010). In fact, the only recent review of the refutation text literature is Zengilowski et al.'s (2021) critical review, which highlighted a number of the field's limitations. In addition to the lack of recent synthesis efforts, it is still unclear what factors can influence the effects of refutation text. As such, we performed a random-effects meta-analysis to address the following questions about refutation text:

RQ1: How effective is refutation text in promoting learning, and is there any evidence of publication bias in the analysis?

RQ2: To what extent do features of the refutation text moderate its effectiveness?

RQ3: To what extent do methodological features of the research moderate the effectiveness of refutation text?

Methods

Literature Search

Databases were searched on March 28, 2020, for the key term refutation* text*. We did not apply any starting time limits on the searches, meaning the databases were searched for articles as old as they contained. The databases produced the following number of research items (number per database in parenthesis): Academic Search Complete (182), APA PsychINFO (137), CINAHL Plus with Full Text (11), Education Research Complete (101), ERIC (97), MEDLINE (29), and Web of Science Core Collection (266). In addition, we searched the reference list of Tippett's (2010) review, which provided another 32 research items for consideration. In total, our search revealed 855 research items, which after the removal of duplicates yielded 473 individual research items.

Inclusion Criteria

In order to be considered in this meta-analysis, studies had to include a between-subjects comparison of a refutation text condition and a separate non-refutation text condition. In addition, the non-refutation text condition had to contain the same general instructional information as the refutation text, meaning studies with no instructional materials (Heddy et al., 2017) or filler texts (Kendeou et al., 2014) as the

non-refutation text conditions were excluded. Additional inclusion criteria were that the study had to quantitatively measure learning outcomes of factual information rather than beliefs or opinions; report enough data to accurately calculate an effect size; report the number of participants in the experimental and control groups, as this allowed for a meta-analysis examining clear, transparent data and the ability to report the refutation text and comparison group numbers for each comparison; use a research design with few or minimal confounding variables (e.g., Zielinski, 2017 was excluded as it seemed one could not plausibly conclude that the refutation text alone caused any difference that may have occurred); be published in English; and be publicly available.

Study Screening and Coding

Phase I, Initial Study Review

During the first phase of study screening, we examined the titles and abstracts of the research items to see if they met the inclusion criteria. After excluding the studies which did not meet the inclusion criteria, we were left with 129 research items.

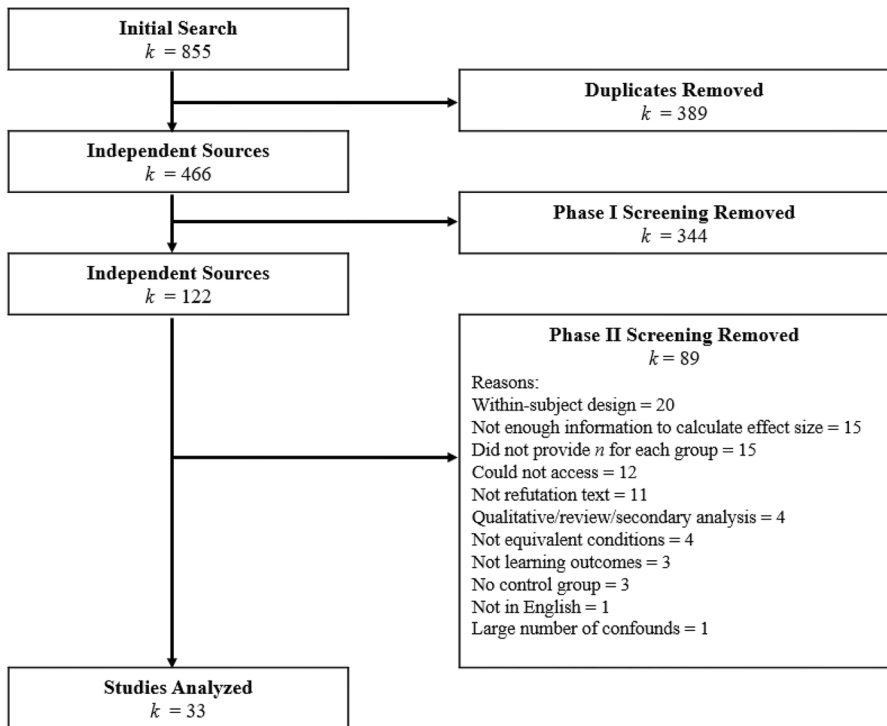


Fig. 1 PRISMA (Moher et al., 2009) inspired flow chart of coding process

Phase II, Full-Text Review

The 129 research items from phase I were retrieved in their full-text form. After reviewing the full text of the items which could be retrieved, 33 research items met the inclusion criteria.

A flow chart of the coding process appears in Fig. 1.

Phase III, Data Extraction

The final phase of our study screening and coding process was extracting the relevant information from each study (Supp. Data 1). Below, we delineate our data extraction paradigm.

Selection of Relevant Comparisons We took steps to ensure the comparisons examined had as few confounding variables as possible. For example, if a study contained three groups, such as an expository text group, a refutation text group, and a refutation text with image group, then we analyzed the refutation text group compared to the expository text group. In this case, we would exclude the refutation text with an image group because the image represents a confounding variable compared to the expository text condition it would be compared to.

In practical terms, we preferred comparisons within studies that compared refutation texts to other types of texts (often expository) while ignoring other treatment conditions present in the study. For example, some studies contained multiple comparisons, such as Adesope et al. (2017), who examined the use of a concept map, refutation text, and scientific text. In this case, we coded the comparison of the refutation text and scientific text, ignoring the concept map condition.

Reconciling Multiple Publications of the Same Data If a conference proceeding was located that appeared to have the same data as a journal article, the journal article was coded, unless it did not contain enough information to meet the inclusion criteria. For example, Diakidoy et al. (2002) had the same title and much of the same text as Diakidoy et al. (2003), but Diakidoy et al. (2003) did not specify the number of individuals in groups on post-testing; hence, Diakidoy et al. (2002) was coded.

Outcome Tests Coded In order to abide by the principle of statistical independence, we analyzed one outcome test from each independent comparison. We coded learning outcomes preferentially as follows: conceptual change tests regardless of format were preferred, followed by transfer tests, comprehension or retention tests, and finally unspecified learning tests. If more than one test was present, we coded the most delayed test to capture learning maintenance over time. For example, Aguilar et al. (2019) reported tests at the time of instruction and 1 week later. For our analysis, we examined the results for test administered at 1 week after the intervention. If more than one test took place at relatively the same time and the types of test could not be preferentially coded based on the basis outlined above, we coded open-ended

transfer tests rather than Likert or multiple-choice tests. For example, Liu and Nesbit (2018) used measures of free recall, short answer transfer, and a multiple-choice measure. We coded the short answer transfer measure.

Moderating Variables We categorized potential moderator variables as either features of the experimental text or as methodological features of the research. Note that regardless of the potentially moderating variable, if the details needed to code for the specific moderator variable were not reported in the primary study, we coded that specific variable as not reported for that study. Table 1 describes the individual variables and coding schemes used for each variable.

Analysis

Inter-rater Reliability

One author coded all of the studies. The second author coded eight randomly selected studies from those that met the inclusion criteria (24.24% of the sample). IBM SPSS 26 was used to calculate Cohen's Kappa and the percent agreement. The Kappa statistic showed strong inter-rater reliability ($k = 0.89$, $p < 0.001$), and inter-rater agreement was strong, with raters agreeing on 90.4% of the coded variables. Disagreements were reconciled through discussion.

Meta-analyzing and Interpreting the Results

We used Comprehensive Meta-Analysis (version 2.2.064) to run the meta-analysis, moderator analyses, publication bias tests, and sensitivity analysis. We used the random-effects model for the meta-analysis due to the nature of our sample (Borenstein et al., 2010). Borenstein et al. (2011, p. 83–84) suggest that random-effects analyses are appropriate when the studies within the analysis have been conducted by independent researchers, resulting in studies that are not functionally identical, and the goal is to generalize the results across a wide variety of scenarios outside of a narrow population. We report the effect size Hedges g , the standardized mean difference effect size with an adjustment for small sample size bias. Positive effects indicate a benefit of refutation text compared to non-refutation text conditions. Effect sizes were interpreted as follows: $g = 0.20$ being a small effect, $g = 0.40$ being a moderate effect, and $g = 0.60$ being a large effect (Hattie, 2015).

Table 1 The list of variables coded as potential moderators. Note that all categories had a “not reported” option if the specific variable was not reported in the primary study

Features of the Text	
<i>Variable name</i>	<i>Coding scheme</i>
Domain of learning materials	We categorized studies as falling within mathematics, science, or social science domains
Additional materials or instruction	We coded whether the refutation text included any additional materials or instruction other than the text itself, such as images or a mixture of different instructional techniques
Media format	We coded whether the text appeared as print text, appeared electronically, or a mixture of both
Text length	When enough information was provided, we documented the difference between the length of the refutation text and the control condition text. We report the difference between the two as a percentage with the refutation text as the reference point. In few cases was the control text longer than the refutation text
Reading level	Authors often reported the reading level of their text passage(s) by grade level. Accordingly, we coded the reading level of the refutation text as either middle level (grades 6–8), high school level (grades 9–12), or post-secondary level (grades 13+)
Features of the research	
<i>Variable name</i>	<i>Coding scheme</i>
Comparison condition	We coded the non-refutation text conditions as the authors reported them. If the specific type of text (e.g., expository, narrative) was not specified, it was coded as “non-specified text”
Learner age	Learners’ ages were largely categorized by grade level, including primary (grades K-5), middle (grades 6–8), secondary (grades 9–12), and post-secondary (students in university or similar settings). We also included studies that had mixed adult age groups (post-secondary setting and other settings) or a mixture of age groups
Prior knowledge	We coded the level of prior knowledge as it was reported in the primary source, such as low, high, or mixed
Participant assignment	We coded the way the researchers assigned participants to conditions as it was reported in the primary study. Our sample included studies that used random assignment and convenience assignment
Type of outcome test	We coded the format of the outcome test as it was reported in the primary study, including Likert scales, multiple choice tests, open-ended questions, true–false questions, and mixed item types
Timing of outcome test	The timing of the outcome test was coded as occurring the same day as the intervention, 2 days to 1 week after the intervention, 8 days to 1 month after the intervention, or more than 1 month after the intervention
Publication type	We coded studies as journal articles, dissertations/theses, or conference proceedings
Location of study	We categorized studies by the continent in which the research took place, locating studies from North America and Europe

Meta-analysis of Refutation Text on Learning

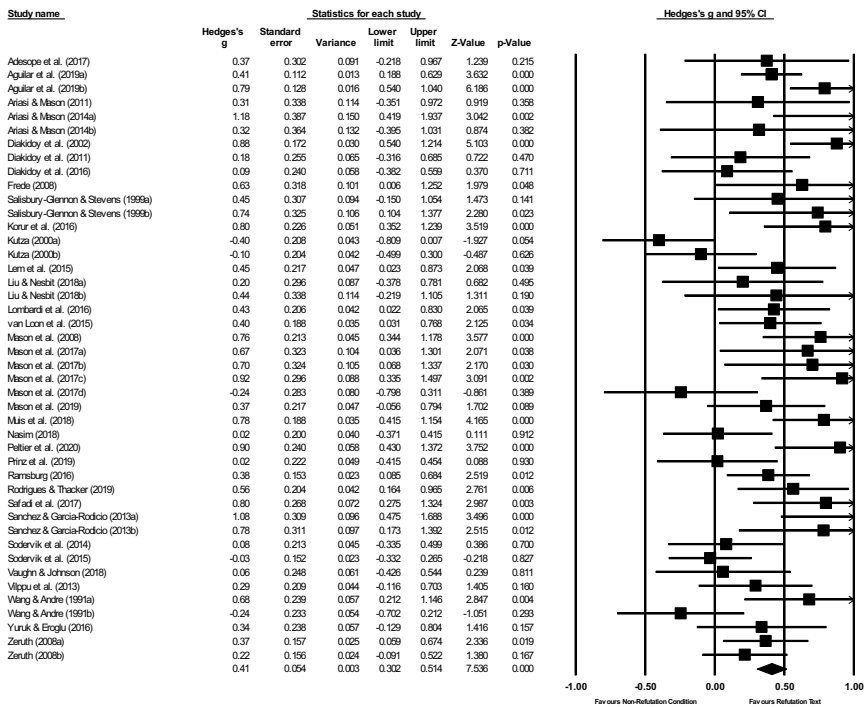


Fig. 2 Forest plot showing the random-effects meta-analysis results

Results

RQ1: How Effective Is Refutation Text in Promoting Learning, and Is There Any Evidence of Publication Bias in the Analysis?

After broadly searching the medical science, social science, psychological science, and science education literature, 33 studies that included 44 independent comparisons met the inclusion criteria. A random-effects meta-analysis of the 44 independent comparisons found a moderate overall effect size ($g=0.41$, $p<0.001$, 95% CI [0.30, 0.51], $n=3,869$) with significant heterogeneity ($Q(43)=109.59$, $p<0.001$, $I^2=60.76$). The forest plot is in Fig. 2.

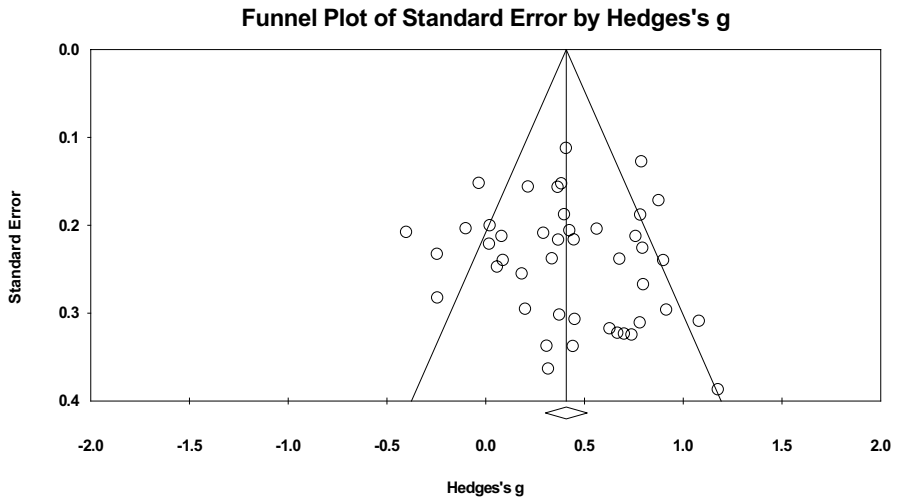


Fig. 3 Funnel plot of the studies included in the meta-analysis

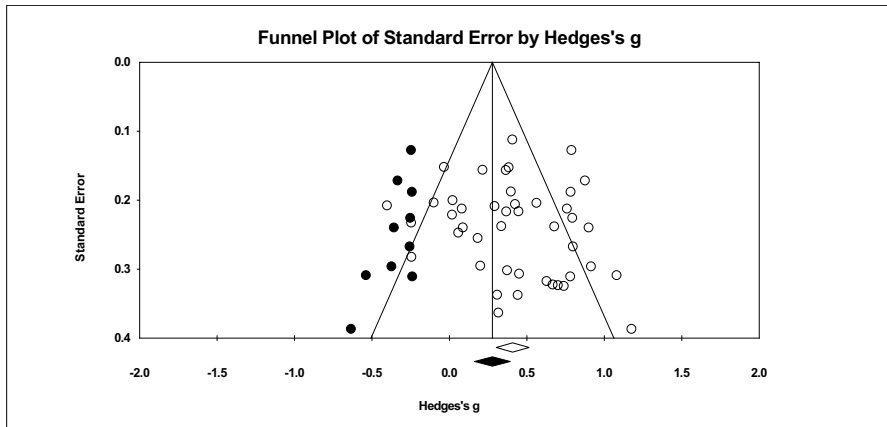


Fig. 4 Funnel plot of the studies adjusted to the left of mean in the trim and fill analysis. The solid marks indicate adjusted studies and overall effect

Publication Bias

Publication bias is a consistent concern with meta-analyses, so we took steps to examine whether this was an issue within our data set. We examined the funnel plot for asymmetry and calculated a number of indicators of publication bias. The funnel plot appeared relatively symmetrical (Fig. 3). Egger's regression intercept (Egger

et al., 1997) showed no significant evidence of funnel plot asymmetry ($t(42)=0.55$, one-tailed $p=0.29$).

Despite the result of the Egger’s regression intercept test, we also utilized Duval and Tweedie’s (2000) trim and fill analysis. The analysis found no missing studies to the right of the mean. However, 10 studies were adjusted to the left of the mean (Fig. 4), which would make the overall effect of refutation text compared to other conditions $g=0.28$, 95% CI [0.16, 0.39]. Borenstein et al. (2011) noted that when evaluating the impact of bias, Duval and Tweedie’s (2000) trim and fill may indicate either a trivial change, a modest change that does not change the overall major finding, or a substantial change that would change the interpretation of the study. In this case, the adjusted effect size is still considered a

Leave One Out Analysis

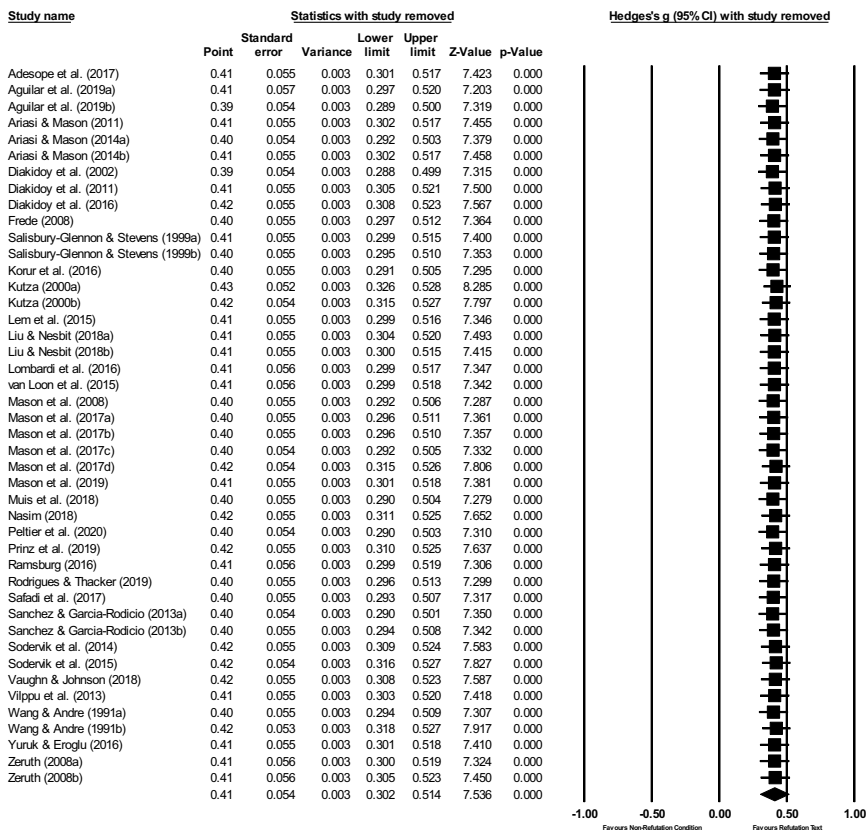


Fig. 5 Forest plot of the leave one out meta-analysis. Note that there is no dramatic effect on the overall effect size when any individual study is omitted

small to moderate effect showing that refutation text benefits learning and that the effect size found from the trim and fill analysis is within two hundredths of being within the 95% confidence interval of the meta-analytic result. Moreover, the 95% confidence interval of the adjusted effect size overlaps with the 95% confidence interval of the overall meta-analytic result. Accordingly, we see indications of a modest change that does not change the overall interpretation of the key conclusion.

Table 2 Potentially moderating variables related to text features

	$n_{refutation}$	$n_{control}$	k	g	SE	95% CI		Q
						Lower	Higher	
Domain								
Math	301	292	5	0.32*	0.15	0.02	0.62	
Science	1,208	1,214	33	0.45*	0.07	0.32	0.57	
Social science	431	423	6	0.31*	0.14	0.03	0.59	
$Q_{between}$								1.09
Additional materials								
Images	261	267	9	0.36*	0.12	0.12	0.60	
Mix of items	142	147	3	0.83*	0.19	0.45	1.21	
None	1,537	1,515	32	0.38*	0.06	0.26	0.49	
$Q_{between}$								5.23
Media format								
Electronic	757	736	15	0.55*	0.09	0.38	0.73	
Paper	626	636	16	0.44*	0.09	0.27	0.60	
Mixed	82	84	2	0.24	0.23	-0.22	0.69	
Not reported	475	473	11	0.21*	0.10	0.02	0.41	
$Q_{between}$								7.03
Text length difference								
< 10%	594	610	18	0.45*	0.09	0.27	0.63	
10–20%	600	581	9	0.20	0.11	-0.02	0.41	
20–50%	229	226	5	0.50*	0.15	0.20	0.80	
> 50%	57	43	1	0.56	0.34	-0.09	1.22	
Not reported	460	469	11	0.49*	0.11	0.28	0.70	
$Q_{between}$								5.23
Text reading level								
Middle level	40	40	2	0.31	0.30	-0.27	0.89	
High school level	566	541	6	0.46*	0.13	0.21	0.72	
Post-secondary level	167	177	5	0.26	0.17	-0.07	0.59	
Not reported	1,167	1,171	31	0.42*	0.07	0.29	0.55	
$Q_{between}$								1.12

* $p < .05$

Finally, we used the classic fail-safe n test (Rosenthal, 1979) to examine how many studies would be needed to bring $p > 0.05$. The test showed that 1,625 studies would be needed to bring $p > 0.05$.

Since the majority of test results indicated that publication bias is not a notable issue in this sample, and the trim and fill analysis found an adjusted effect size that does not change the key finding of the study, publication bias was determined not to be a substantial concern in this sample.

Sensitivity Analysis

We used Comprehensive Meta-Analysis to run the random-effects meta-analyses with each individual study omitted. The results were very similar to the results of the overall meta-analysis (Fig. 5), with the overall effect sizes ranging from $g = 0.39$ to $g = 0.43$ depending on the individual study omitted. From this analysis, it appears that no individual study had noticeably strong influence on the results of the meta-analysis.

RQ2: To What Extent Do Features of the Refutation Text Moderate Its Effectiveness?

To examine this research question, we examined potentially moderating variables in relation to the features of the text (Table 2). Each potentially moderating variable is addressed below.

We first examined whether refutation text was moderated by the field in which it was conducted, categorizing studies as either science, mathematics, or social sciences. We found that the effects of refutation text were robust across content domains and did not significantly vary based on the domain of the materials ($Q_b(2) = 1.09, p = 0.58$).

Next, we examined how having additional materials with the refutation text, such as images, influenced the effectiveness of the refutation text. The effects of refutation text did not significantly vary based on the inclusion of additional instructional materials ($Q_b(2) = 5.23, p = 0.07$). However, the majority of studies were text-only interventions that did not include these features.

Due to the vast amount of media being consumed in electronic format, it was important to examine whether the effectiveness of refutation text was moderated by media format (e.g., print, electronic). We found that the effectiveness of refutation text was robust across media formats, as the effects did not significantly vary based on the format of the media ($Q_b(3) = 7.03, p = 0.07$).

One common feature of refutation text is that it tends to be longer than an expository text. Accordingly, we examined how the effectiveness of refutation text was moderated by length differences between the refutation text and the comparison condition. We found that the effectiveness of refutation text did not significantly vary by text length differences ($Q_b(4) = 5.23, p = 0.26$). Therefore, the effect of refutation text is robust even when the texts are of similar length to the control text.

Table 3 Potentially moderating variables related to methodological features

	$n_{refutation}$	$n_{control}$	k	g	SE	95% CI		Q
						Lower	Higher	
Comparison								
Expository/scientific text	1,410	1,394	30	0.36*	0.07	0.23	0.49	
Mixed control condition	42	41	1	0.80*	0.36	0.10	1.49	
Non-specified text	298	298	10	0.52*	0.12	0.28	0.76	
Problem solving	190	196	3	0.42*	0.19	0.04	0.80	
$Q_{between}$								2.56
Learner Age								
Primary (e.g., K-5)	94	85	2	0.57*	0.24	0.10	1.03	
Middle (e.g., 6–8)	199	204	4	0.71*	0.17	0.39	1.04	
Secondary (e.g., 9–12)	85	91	4	0.49*	0.20	0.10	0.88	
Post-secondary	1,169	1,193	30	0.33*	0.06	0.20	0.45	
Mixed adult (post-secondary and after)	91	73	2	0.33	0.24	-0.14	0.80	
Mixed ages	302	283	2	0.59*	0.20	0.20	0.98	
$Q_{between}$								6.60
Prior knowledge								
Low	186	193	6	0.40	0.15	0.11	0.70	
Mixed	87	84	1	-0.03	0.30	-0.63	0.56	
Not reported	1,667	1,652	37	0.42	0.06	0.31	0.54	
$Q_{between}$								2.18
Participant assignment								
Random	1,690	1,677	39	0.37*	0.06	0.26	0.48	
Convenience (classroom random assignment)	72	75	1	0.88*	0.31	0.27	1.49	
Not reported	178	177	4	0.63*	0.17	0.28	0.97	
$Q_{between}$								4.33
Format of test								
Likert	260	271	5	0.23	0.15	-0.07	0.53	
Mixed item types	396	402	13	0.43*	0.11	0.22	0.64	
Multiple choice	416	419	9	0.34*	0.12	0.10	0.57	
Open-ended	386	384	10	0.40*	0.12	0.17	0.63	
True false	342	324	4	0.59*	0.17	0.25	0.93	
Not reported	140	129	3	0.60*	0.20	0.20	1.00	
$Q_{between}$								3.76
Timing of test								
Same day	654	644	17	0.39*	0.09	0.21	0.57	
2 days to 1 week	632	624	11	0.38*	0.11	0.17	0.60	
8 days to 1 month	527	538	13	0.43*	0.10	0.22	0.64	
More than 1 month	86	78	2	0.56*	0.26	0.05	1.07	
Not reported	41	45	1	0.45	0.36	-0.26	1.16	
$Q_{between}$								0.48

Table 3 (continued)

	$n_{refutation}$	$n_{control}$	k	g	SE	95% CI		Q
						Lower	Higher	
Publication type								
Proceeding	129	118	2	0.73*	0.21	0.31	1.15	
Dissertation/thesis	388	398	6	0.11	0.12	-0.13	0.34	
Journal article	1,423	1,413	36	0.45*	0.06	0.34	0.56	
$Q_{between}$								9.01*
Study location								
Europe	770	775	21	0.41*	0.08	0.25	0.57	
North America	951	948	18	0.38*	0.08	0.21	0.54	
Not reported	219	206	5	0.52*	0.16	0.21	0.83	
$Q_{between}$								0.63

* $p < .05$

It was also important to examine whether the effectiveness of refutation text varied with the reading level of the text. While we found that reading level did not significantly influence the effectiveness of the refutation text ($Q_b(3) = 1.12$, $p = 0.77$), most study authors did not report the reading level of the text. Therefore, these results should be interpreted with caution.

RQ3: To What Extent Do Methodological Features of the Research Moderate the Effectiveness of Refutation Text?

To examine this research question, we examined potentially moderating variables in relation to the methodological features of the research (Table 3).

We sought to examine how the effects of refutation text vary when compared to other conditions such as narrative or expository texts. Importantly, this moderator analysis showed that refutation text was significantly more effective than all comparison conditions, and these positive effects did not vary significantly depending on the comparison condition ($Q_b(3) = 2.56$, $p = 0.47$).

We also questioned whether the age of the learner moderated the effectiveness of refutation text. We found that the effects of refutation text were robust and consistent regardless of the age of the learners ($Q_b(5) = 6.60$, $p = 0.25$). While most of the comparisons in our sample used post-secondary samples, the results revealed that refutation text was associated with positive effects across all other age groups examined.

The learners' prior knowledge could also potentially moderate the effectiveness of refutation text. However, we found that most studies did not categorize the prior knowledge of participants ($k = 37$), and the moderator analysis revealed that prior knowledge did not have a significant moderating effect ($Q(2) = 2.18$, $p = 0.34$).

We next examined whether the method used to assign participants to conditions moderated the effectiveness of refutation text. The results showed that the type of participant assignment did not significantly influence the effectiveness of refutation text ($Q_b(2)=4.33$, $p=0.11$). However, most comparisons assigned participants to conditions randomly, and those that did not were associated with notably larger effect sizes.

Another key element to investigate was potential moderation depending on type of outcome test. We found that the type of outcome test did not significantly influence the effectiveness of refutation text ($Q_b(5)=3.76$, $p=0.58$).

Does the effect of refutation text persist over time? This question prompted our inclusion of the timing of the outcome test as a potential moderator. We found that the effectiveness of refutation text did not significantly differ depending on when the outcome test occurred ($Q_b(4)=0.48$, $p=0.98$). While most studies used a test on the same day as the refutation text was read, two studies found that the positive effects of refutation text persist more than a month after reading the refutation text.

We did find that the effectiveness of refutation text significantly varied by the type of publication the comparison appeared in ($Q_b(2)=9.01$, $p=0.01$). Conference proceedings were associated with the highest effect size ($g=0.73$, $p=0.001$), followed by journal articles ($g=0.45$, $p<0.001$) and dissertations or theses ($g=0.11$, $p=0.38$).

Finally, we examined whether the location the study took place in had any significant role in moderating the effects of refutation text, and it did not ($Q_b(2)=0.63$, $p=0.73$).

Discussion

Inspired by a desire to help move learners towards scientifically accepted understandings, we conducted this meta-analysis to assess the effectiveness of refutation text on learning outcomes. In this section, we discuss the results in relation to each research question.

RQ1: How Effective Is Refutation Text in Promoting Learning, and Is There Any Evidence of Publication Bias in the Analysis?

Previous research has indicated that refutation text can be an effective intervention for facilitating learning (Guzzetti et al., 1993). However, that work was completed nearly 2 decades ago, and there has been a plethora of research around refutation text since. Examining research around a similar construct, conceptual change texts, Armağan et al. (2017) found very strong effects. Our findings are largely consistent with these prior syntheses of the literature. Across 44 independent comparisons spanning 3,869 participants, our results revealed that refutation text was broadly effective for all learners, irrespective of the content domain, test type, or timing of test administration. Across all of our moderator analyses, we found only one negative effect size, and it was not statistically significant. Of the individual comparisons

included in the meta-analysis, only five of the 44 showed an advantage for the non-refutation text condition. We looked for publication bias using four different techniques or analyses, and no notable evidence of publication bias was found that would dramatically change the interpretation of the results. Our sensitivity analysis also revealed that no individual study greatly influenced the overall meta-analytic result. Together, the results presented here provide compelling evidence that refutation text can be a broadly effective strategy for facilitating learning.

RQ2: To What Extent Do Features of the Refutation Text Moderate Its Effectiveness?

Through moderator analyses, we sought to examine how various design features of the text may moderate the effectiveness of the refutation text. We first examined the domain of the learning materials. We found the effects of refutation text did not differ across high-level domain classifications (science, social science, mathematics). However, Zengilowski et al. (2021) observed that many refutation texts covered a relatively small knowledge domain, and our findings largely confirm this as most studies were about scientific concepts. We therefore reiterate their call for refutation text research around new topics to see if the effects vary in different knowledge domains. This is potentially important because the learners' prior knowledge and beliefs can influence their learning (Van Loon et al., 2015; Zengilowski et al., 2021). While many subjects, like photosynthesis, may not have strong personal beliefs involved, other topics, like vaccination, may have strong personal beliefs involved. Thus, it seems plausible that the effects of refutation texts may vary in these different subject areas.

Based on the multimedia learning literature (Mayer, 2017, 2021), we had expected that refutation texts containing images would be more effective than refutation texts without images. Although the multimedia principle has been found to be robust, we did not find any evidence of it in this meta-analysis. In fact, the effect size found for studies that used images was very similar to the effect size found for text-only interventions. This is an interesting finding that is difficult to explain; however, we question if the images used simply did not add value compared to the text and were therefore ignored by the learners. Since relatively few studies used images in our sample ($k=9$), more research is needed to better understand why graphics may, or may not, aid learning with refutation texts.

We also examined whether the refutation text being presented on a screen or on paper moderated the effect. Based on recent meta-analyses (Clinton, 2019; Kong et al., 2018), we expected that reading on paper would be superior. However, this was not replicated in our results. Clinton (2019) found that expository texts were associated with lower (i.e., worse) effects than narrative texts when reading on an electronic device. We wonder if refutation texts would lead to similar results if they had been included in their analysis. Primary research is warranted to investigate how well refutation texts maintain their efficacy when presented on a computer screen as compared to paper.

The length of refutation texts is also a concern, as in many cases the nature of a refutation text demands additional text as compared to an expository text on the same subject. Our analysis indicated that the length of the refutation text did not make a significant difference on learning. In other words, it did not matter whether the refutation texts were 25% longer than expository texts or < 10% longer than the expository texts; there were no statistically significant differences between levels of the moderator. Recall that a refutation text states a common misconception about a subject, refutes it, and provides an explanation of the correct conception (Hynd, 2001; Mason et al., 2019; Tippett, 2010). In order for a refutation text to be the same length as or only marginally longer than a comparable expository text, it is likely that some text must be removed from the explanation of the accepted understanding. In other words, it appears that the refutation text structure *potentially* allows for a shorter explanation of the subject and yet produces superior learning outcomes than expository texts.

We also sought to understand whether the reading level of the refutation text moderated the effect based on the hypothesis that texts at lower reading levels may be easier to understand, and texts at higher reading levels may be more complex or use more advanced terminology. However, we did not find any evidence to support this hypothesis in the moderator analysis as the learning effect did not significantly vary with experimental text reading level.

In summary, we found no evidence that any of the refutational text design features we analyzed significantly moderated the effects of refutation text. Rather, the effect was robust across a variety of different design components. However, research is certainly warranted around the domain of the learning materials, and we echo Zengilowski et al.'s (2021) call for research in new topic areas.

RQ3: To What Extent Do Methodological Features of the Research Moderate the Effectiveness of Refutation Text?

Our third research question was focused on examining whether specific aspects of the primary research significantly moderated the effectiveness of refutation text. Based on prior work, we knew that learners preferred refutation texts (Tippett, 2010), but we questioned whether they would be more effective compared to some conditions as opposed to others. For example, we questioned whether a refutation text had differential effects when compared to expository texts or narrative texts. Interestingly, we did not find significant differences by control condition, implying that refutation texts consistently facilitated learning. The question, then, is why this occurred? Presumably there are cognitive mechanisms that can explain these benefits. Through the lens of KReC, it seems plausible that, as Butterfuss and Kendeou (2020) argued, the format of the refutation text helps the learner move towards activating the correct conception rather than the incorrect one. Researchers have been investigating why refutation text may be effective through alternative measurements such as think-aloud protocols (Kendeou et al., 2019) and eye tracking (Ariasi & Mason, 2011; Ariasi et al., 2017), which could be a productive way forward for understanding how learners process refutation texts compared to other conditions.

We also questioned whether the learners' age played any role in moderating the effects of refutation text. Based on Tippet's (2010) review, we expected to see greater benefits for the middle grades rather than elementary, secondary, or post-secondary education. We did not find any significant differences depending on the age of the learner. However, we speculate that this difference between our work and Tippet's may be due to a few factors. First, we categorized a learner's age by the grade range in schooling (i.e., primary, secondary), whereas Tippet examined by specific grade level. Second, there are comparably few studies in our sample that investigated the effects of refutation texts with students not in post-secondary education. In addition, our inclusion criteria were different than Tippet's. While we did not find statistically significant differences between levels of the moderator for learner age, we note that while there were fewer studies investigating students in K-12 settings, these were associated with larger effect sizes even if they were not statistically different than the effect derived from post-secondary learners. Therefore, future work may be warranted to investigate this phenomenon. Perhaps younger learners are more open to changing their conceptions about things they do not understand well, or perhaps the novelty of refutation texts has more influence on their learning. It also seems plausible that the topics of study investigated with these younger learners have fewer personal beliefs attached, which may explain the strong effect sizes. For instance, Mason et al. (2008) investigated refutation texts with fifth grade students learning about light. This is opposed to some studies with older learners that may have investigated topics that the learners felt more invested in or had stronger beliefs around, such as influenza (Vaughn & Johnson, 2018).

When we examined the learners' prior knowledge as a potentially moderating variable, we had hoped to be able to decipher whether refutation text was more or less effective for learners of differing levels of preparation. This could be important for two reasons. First, we know that learners' misconceptions can influence their conceptual change and learning (Dole & Niederhauser, 1990). Second, there has been evidence for expertise-reversal effects in other areas of research, where learners with high prior knowledge do not find the same benefits as learners with low prior knowledge (Chen et al., 2017; Kalyuga & Renkl, 2010). However, the reporting of the primary studies largely inhibited our ability to examine how refutation text differed depending on learners' prior knowledge, in that 37 comparisons did not categorize their learners' level of prior knowledge. This could be due in part to our coding scheme that set to classify learners as low, high, or mixed prior knowledge. We note that in many studies, authors did report the scores of the learning outcome pre-test. However, we feel it is important for future researchers to not only provide the scores, but also classify whether they consider the learners to have low or high prior knowledge so phenomenon such as the expertise-reversal effect can be addressed in this area.

As a methodological note in relation to meta-analysis more generally, our results show the importance of broadly searching the literature. We found that the largest effects were located in conference proceedings, whereas non-significant small effects were found in dissertations. Meanwhile, moderate effects were extracted from journal articles. Together, these results show the benefits of a comprehensive literature search that is not restricted to only peer-reviewed journal articles.

Finally, we note that specific aspects of the research design, such as how participants were assigned to condition and the type of outcome test used, can also have significant implications for the interpretation of meta-analyses. For instance, if significant effects of the intervention were found in studies that used convenience sampling but not for studies that used random assignment, one may question the efficacy of the intervention. In addition, if studies found benefits for multiple-choice tests but not for applied tasks, transfer tasks, or written essay questions, one may question whether the information was learned in a generalizable and transferable way or only through rote memorization. In this case, we did not find any significant moderation by the way participants were assigned to condition, nor due to the type of outcome test. This adds strength and some level of trustworthiness to the overall meta-analytic result that refutation text aids learning.

These results are further strengthened by the finding that the effects of refutation text did not vary depending on when the learning test occurred. By coding the most delayed test, we were able to examine the effects of refutation text across a variety of time periods, from the same day of the intervention through more than a month later. We found that the effects of refutation text persisted over time, although we note that the number of studies that examined the effects more than a month later was minimal ($k=2$). The findings of the meta-analysis are also strengthened by the finding that the effects of refutation text did not vary by the continent on which the study took place. Overall, we found relatively robust evidence that refutation text can be an effective learning intervention across multiple contexts.

Implications for Theory

Recall that the purpose of this paper is not to validate one theory of conceptual change as opposed to another. Therefore, we explain our results in relation to the cognition-oriented KReC framework for similar reasons as explained in the literature review.

As noted, the KReC framework suggests that knowledge can be revised through activating relevant prior knowledge in relation to the new information, which facilitates the development of an integrated mental model which can then be used for activation in place of previous incorrect or incomplete mental models (Kendeou et al., 2019; Lassoende et al., 2017; Trevors et al., 2017; Will et al., 2019). In line with this theoretical perspective, it would appear that refutation texts can facilitate these knowledge revision processes, resulting in the increased activation needed for the new mental models to outcompete the older, incorrect, or incomplete mental models. However, research has shown that these incorrect or incomplete models do not just disappear; rather, they are maintained but can be suppressed (Mason & Zaccoletti, 2021; Treagust & Duit, 2008). This supports the idea that conceptual change is an ongoing process, rather than an instantaneous and permanent shift in conceptions (Treagust & Duit, 2008). Researchers have begun examining the role of inhibition in this process and have theorized it to be important for conceptual change, but more research in the area is needed (for a review, see Mason & Zaccoletti, 2021).

It should be noted that while KReC is a cognition-oriented theory, it can potentially account for the affective and non-cognitive factors (e.g., message

characteristics) posited to underlie conceptual change by other researchers (e.g., Pintrich et al., 1993; Sinatra, 2005). For example, Trevors et al. (2017) investigated that the role emotion can play in knowledge revision through the lens of KReC. Moreover, it seems challenging to argue that the multitude of factors discussed by scholars (e.g., Pintrich et al., 1993; Sinatra, 2005; Treagust & Duit, 2008), such as learners' prior conceptions, motivation, and characteristics of the learning materials, would not influence learners' activation of knowledge. Moving forward, it seems plausible that these views on knowledge revision or conceptual change could be generally complimentary.

Due to the reporting of the primary studies, one theoretical issue that this study could not meaningfully address is the question of whether learners' prior knowledge or prior misconceptions (or preconceptions) influences the effectiveness of refutation text. However, prior work has indicated that these beliefs, and the strength of these beliefs, can be an important factor in moderating the effects of refutation text (Van Loon et al., 2015; Zengilowski et al., 2021). In our sample, few studies reported the learners' level of prior knowledge as low or high. Furthermore, we had planned to examine how learners' prior misconceptions (or inaccurate preconceptions) moderated the effectiveness of refutation text. Unfortunately, very few studies specified the level of misconceptions held by learners and those that did took varied approaches. For example, Mason et al. (2017) stated that all their participants held a common misconception, Ariasi and Mason (2011) used a cutoff score on the pre-test as evidence of what they termed alternative conceptions, and Liu and Nesbit (2018) stated that participants demonstrated misconceptions at the pre-test. As of a result of these varied reporting methods and the considerable number of studies which did not mention inaccurate preconceptions, we did not examine this variable in the analysis. Thus, how and to what degree learners' preconceptions or misconceptions impact learning outcomes when using refutation text is still an open question. In addition, Chi (2008) differentiated between beliefs and mental models. Beliefs make up mental models and therefore require a finer grain of measurement than we were able to pursue with a meta-analysis such as this. This may become a particularly important question into the future as refutation text is used with concepts that do not have a clear scientifically correct answer or are more tied to political or social issues (Zengilowski et al., 2021).

Implications for Practice

While the implications of this meta-analysis encourage reconciliation among theoretical perspectives, the practical implications are quite clear. Our results show that refutation text is effective for facilitating learning in all situations that were examined through the moderator analyses. Moreover, we did not find differential effects based on learner age, content domain, or test timing, which shows that refutation texts have been consistently effective across many contexts. In short, our results show that refutation texts can be used when there are commonly held misconceptions in the subject area (but recall that a common misconception must be known to write a refutation text), at least for one-time interventions involving shorter texts.

One particularly notable finding of this meta-analysis is that the benefits of refutation text were maintained over time. Zengilowski et al. (2021) recently qualitatively reviewed and critiqued the refutation text literature. Their analysis revealed that the persistent benefits of refutation text over time tended to occur among studies that dealt with scientific topics rather than those addressing political or social issues. This is consistent with our findings. In our analysis, we had two studies that examined learning more than a month after the intervention, one around light (Mason et al., 2008), and one around heat and temperature (Yürük, & Eroğlu, 2016). The question of how the effects of refutation text vary over time depending on the knowledge domain highlights an important question worthy of future research. Similarly, we question whether learners broadly generalize what they learned from refutation text outside of the original context, as conceptual change is viewed as domain specific (Treagust & Duit, 2008).

It is important to note that while we did examine some features of refutation text design as potentially moderating variables, this study was not intended to examine how to specifically design refutation texts. To do so in depth would require comparisons of design elements between multiple refutation texts, and this type of study would not have met our inclusion criteria. We note that not all forms of refutation text that were included in this study followed an identical format. For example, some used explicit refutation (Rodrigues & Thacker, 2019), examined the credibility of the explanation (Van Boekel et al., 2017), or used more implicit refutation (Adesope et al., 2017). In addition, some studies used questions rather than statements of misconceptions (Ariasi & Mason, 2011, 2014), and some aligned more with the conceptual change text format but still contained the components of refutation text (Wang & Andre, 1991). These different designs could potentially influence the social, emotional, or motivational beliefs of an individual learner and thereby impact the conceptual change process (Pintrich et al., 1993; Sinatra, 2005). As such, research synthesis is needed to disentangle these design elements to better understand which types of refutation text are more effective, if any difference exists at all.

Limitations

Like all meta-analyses, this study is limited by the primary studies it is based upon. Studies meeting the inclusion criteria generally focused on scientific or mathematical concepts rather than social sciences, public policy, or public health issues, and we were unable to address how prior knowledge and prior misconception (and pre-conception) differences influenced the effectiveness of refutation text due a lack of explicit reporting in primary studies. In addition, the studies we evaluated were primarily one-time interventions with relatively short text passages (the longest passage reported was 1,542 words). The effectiveness of refutation text in these short passages is supportive of its use in common communication tools, such as news media and websites specifically designed to counter myths or misconceptions. However, it is unclear whether the effect would persist in longer text readings, such as textbooks, or other contexts such as repeated viewings over time. This type of research is needed to better understand if refutation text, as a mainstream, commonly used teaching strategy, maintains its effectiveness.

Another limitation was the reporting of primary studies. We had a substantial number of studies that were excluded from the analysis because they did not provide enough information for effect size extraction ($k=15$) or did not provide very basic information about their experiment, such as the number of participants in each group ($k=15$). Had this basic information been provided, our sample of studies analyzed would have nearly doubled in size. We encourage researchers in future studies to clearly and fully report their experimental methods and results.

Conclusion

We broadly and comprehensively systematically reviewed the literature and conducted a meta-analysis examining the extent to which learning is facilitated by refutation text compared to other learning conditions. Across 44 independent comparisons, we found compelling evidence to support the use of refutation text as a low-to-zero cost strategy to facilitate learning. Transitioning textual materials, particularly short pre-existing texts, to refutation text format would likely require little effort and minimal time investment. Therefore, we suggest that when knowledge revision is critical, particularly about scientific topics, refutation text may be an effective way to promote scientific understanding. However, we concur with Zengilowski et al. (2021) that refutation text should likely not be viewed as the panacea to combating all misinformation and misconceptions. Working from the foundational understanding that refutation text can aid learning (as shown in this meta-analysis), but may need to be designed differently depending on the context (see Zengilowski et al., 2021), researchers and practitioners alike can use refutation text as a tool for combating misconceptions.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10648-021-09656-z>.

Acknowledgements We are grateful to Dr. William Romine for his comments on a previous version of this manuscript.

Declarations

Conflict of Interest The authors declare no competing interests.

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.
- *Aguilar, S. J., Polikoff, M. S., & Sinatra, G. M. (2019). Refutation texts: A new approach to changing public misconceptions about education policy. *Educational Researcher*, 48(5), 263–272.
- *Ariasi, N., & Mason, L. (2011). Uncovering the effect of text structure in learning from a science text: An eye-tracking study. *Instructional Science*, 39(5), 581–601.
- *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 & Mathematics Education*, 12(3), 493–523.
- Ariasi, N., Hyönä, J., Kaakinen, J. K., & Mason, L. (2017). An eye-movement analysis of the refutation effect in reading science text. *Journal of Computer Assisted Learning*, 33(3), 202–221.
- Armağan, F. Ö., Keskin, M. Ö., & Akin, B. S. (2017). Effectiveness of conceptual change texts: A meta analysis. *European Journal of Science and Mathematics Education*, 5(3), 343–354.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97–111.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to Meta-Analysis*. John Wiley & Sons.
- Butterfuss, R., & Kendeou, P. (2020). Reducing interference from misconceptions: The role of inhibition in knowledge revision. *Journal of Educational Psychology*, 112(4), 782–794.
- Çalik, M., Ayas, A., & Coll, R. K. (2007). Enhancing pre-service elementary teachers' conceptual understanding of solution chemistry with conceptual change text. *International Journal of Science and Mathematics Education*, 5(1), 1–28.
- 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: The Official Journal of the National Association for Research in Science Teaching*, 34(2), 107–123.
- Chen, O., Kalyuga, S., & Sweller, J. (2017). The expertise reversal effect is a variant of the more general element interactivity effect. *Educational Psychology Review*, 29(2), 393–405.
- Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of Research on Conceptual Change* (pp. 61–82). Erlbaum.
- Clinton, V. (2019). Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading*, 42(2), 288–325.
- *Diakidoy, I.-A. N., Kendeou, P., & Ioannides, C. (2002). *Reading about energy: The effects of text structure in science learning and conceptual change*. Paper presented at the Annual Meeting of the American Educational Research Association.
- Diakidoy, I.-A.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.
- *Diakidoy, I.-A.N., Mouskounti, T., & Ioannides, C. (2011). Comprehension and learning from refutation and expository texts. *Reading Research Quarterly*, 46(1), 22–38.
- *Diakidoy, I.-A.N., Mouskounti, T., Fella, A., & Ioannides, C. (2016). Comprehension processes and outcomes with refutation and expository texts and their contribution to learning. *Learning & Instruction*, 41, 60–69.
- Dole, J. A., & Niederhauser, D. S. (1990). Students' level of commitment to their naive conceptions and their conceptual change learning from texts. *National Reading Conference Yearbook*, 39, 303–310.
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463.

- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634.
- *Frède, V. (2008). Teaching astronomy for pre-service elementary teachers: A comparison of methods. *Advances in Space Research*, *42*(11), 1819–1830.
- Guzzetti, B. J., Snyder, T., & Glass, G. (1992). Promoting conceptual change in science: Can texts be used effectively? *Journal of Reading*, *35*(8), 642–649.
- 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*, 117–159.
- Hattie, J. (2015). The applicability of visible learning to higher education. *Scholarship of Teaching and Learning in Psychology*, *1*(1), 79.
- Heddy, B. C., Danielson, R. W., Sinatra, G. M., & Graham, J. (2017). Modifying knowledge, emotions, and attitudes regarding genetically modified foods. *Journal of Experimental Education*, *85*(3), 513–533.
- Hynd, C. R. (2001). Refutational texts and the change process. *International Journal of Educational Research*, *35*(7–8), 699–714.
- Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issue. *Instructional Science*, *38*(3), 209–215.
- 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.
- 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.
- Kendeou, P., Butterfuss, R., Kim, J., & Van Boekel, M. (2019). Knowledge revision through the lenses of the three-pronged approach. *Memory & Cognition*, *47*(1), 33–46.
- Kong, Y., Seo, Y. S., & Zhai, L. (2018). Comparison of reading performance on screen and on paper: A meta-analysis. *Computers & Education*, *123*, 138–149.
- *Korur, F., Enil, G., & Göçer, G. (2016). Effects of two combined methods on the teaching of basic astronomy concepts. *Journal of Educational Research*, *109*(2), 205–217.
- *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* (2001–95003–130; Issues 8-A) [ProQuest Information & Learning].
- Lassonde, K. A., Kolquist, M., & Vergin, M. (2017). Revising psychology misconceptions by integrating a refutation-style text framework into poster presentations. *Teaching of Psychology*, *44*(3), 255–262.
- *Lem, S., Kempen, G., Ceulemans, E., Onghena, P., Verschaffel, L., & Dooren, W. (2015). Combining multiple external representations and refutational text: An intervention on learning to interpret box plots. *International Journal of Science & Mathematics Education*, *13*(4), 909–926.
- *Liu, Q., & Nesbit, J. C. (2018). Conceptual change with refutational maps. *International Journal of Science Education*, *40*(16), 1980–1998.
- *Lombardi, D., Danielson, R. W., & Young, N. (2016). A plausible connection: Models examining the relations between evaluation, plausibility, and the refutation text effect. *Learning & Instruction*, *44*, 74–86.
- Mason, L., & Zaccoletti, S. (2021). Inhibition and conceptual learning in science: A review of studies. *Educational Psychology Review*, *33*, 181–212.
- *Mason, L., Gava, M., & Boldrin, A. (2008). On warm conceptual change: The interplay of text, epistemological beliefs, and topic interest. *Journal of Educational Psychology*, *100*(2), 291–309.
- *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.
- *Mason, L., Zaccoletti, S., Carretti, B., Scrimin, S., & Diakidoy, I.-A.N. (2019). The role of inhibition in conceptual learning from refutation and standard expository texts. *International Journal of Science & Mathematics Education*, *17*(3), 483–501.
- Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, *33*(5), 403–423.
- Mayer, R. E. (2021). Evidence-based principles for how to design effective instructional videos. *Journal of Applied Research in Memory and Cognition*, *10*(2), 229–240.
- Mayo Clinic Staff. (2020). *Debunking COVID-19 (coronavirus) myths*. Mayo Clinic. Retrieved June 10, 2020, from <https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/coronavirus-myths/art-20485720>

- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, 6(7), e1000097.
- *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.
- *Nasim, A. M. (2018). *Using Refutation Texts to Change Attitudes and Knowledge Concerning Auditory Verbal Hallucinations*
- *Peltier, T. K., Heddy, B. C., & Peltier, C. (2020). Using conceptual change theory to help preservice teachers understand dyslexia. *Annals of Dyslexia*. MEDLINE. <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.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227.
- *Prinz, A., Golke, S., & Wittwer, J. (2019). Refutation texts compensate for detrimental effects of misconceptions on comprehension and metacomprehension accuracy and support transfer. *Journal of Educational Psychology*, 111(6), 957–981.
- *Ramsburg, J. T. (2016). *The lesson and the learner: The effect of individual differences and task scaffolding on category learning* (2017–19723–005; Issues 7-B(E)) [ProQuest Information & Learning].
- *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.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638.
- Roth, K. J. (1985). Conceptual change learning and student processing of science texts. *Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL*. Retrieved from <https://eric.ed.gov/?id=ED267980>
- Rusanen, A.-M. (2014). Towards an explanation for conceptual change: A mechanistic alternative. *Science & Education*, 23(7), 1413–1425.
- *Safadi, R., Safadi, E., & Meidav, M. (2017). Reflection on solutions in the form of refutation texts versus problem solving: The case of 8th graders studying simple electric circuits. *Physics Education*, 52(1).
- *Salisbury-Glennon, J. D., & Stevens, R. J. (1999). Addressing Preservice Teachers' Conceptions of Motivation. *Teaching and Teacher Education*, 15(7), 741–752.
- *Sanchez, E., & Garcia-Rodicio, H. (2013). Using online measures to determine how learners process instructional explanations. In *LEARNING AND INSTRUCTION* (Vol. 26, pp. 1–11). PERGAMON-ELSEVIER SCIENCE LTD. <https://doi.org/10.1016/j.learninstruc.2012.12.003>
- Sinatra, G. M. (2005). The "warming trend" in conceptual change research: The legacy of Paul R. Pintrich. *Educational Psychologist*, 40(2), 107–115.
- 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.
- Sinatra, G. M., Broughton, S. H., Diakidoy, I.-A.N., Kendeou, P., & van den Broek, P. (2011). Bridging reading comprehension and conceptual change in science education: The promise of refutation text. *Reading Research Quarterly*, 46(4), 374–393.
- *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.
- *Södervik, I., Virtanen, V., & Mikkilä-Erdmann, M. (2015). Challenges in understanding photosynthesis in a university introductory biosciences class. *International Journal of Science & Mathematics Education*, 13(4), 733–750.
- Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science & Mathematics Education*, 8(6), 951–970.
- Treagust, D. F., & 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.
- Trevors, G. J., Kendeou, P., & Butterfuss, R. (2017). Emotion processes in knowledge revision. *Discourse Processes*, 54(5/6), 406–426.
- Van Boekel, M., Lassonde, K., O'Brien, E., & Kendeou, P. (2017). Source credibility and the processing of refutation texts. *Memory & Cognition*, 45(1), 168–181.

- *van Loon, M. H., Dunlosky, J., van Gog, T., van Merriënboer, J. J. G., & de Bruin, A. B. H. (2015). Refutations in science texts lead to hypercorrection of misconceptions held with high confidence. *Contemporary Educational Psychology*, 42, 39–48.
- *Vaughn, A. R., & Johnson, M. L. (2018). Communicating and enhancing teachers' attitudes and understanding of influenza using refutational text. *Vaccine*, 36(48), 7306–7315.
- *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.
- Vosniadou, S. (Ed.). (2008). *International handbook of research on conceptual change*. Routledge.
- *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.
- Will, K. K., Masad, A., Vlach, H. A., & Kendeou, P. (2019). The effects of refutation texts on generating explanations. *Learning & Individual Differences*, 69, 108–115.
- World Health Organization. (n.d.). *COVID-19 Mythbusters – World Health Organization*. World Health Organization. Retrieved July 26, 2020, from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters>
- *Yürük, N., & Eroğlu, P. (2016). The effect of conceptual change texts enriched with metaconceptual processes on pre-service science teachers' conceptual understandings of heat and temperature. *Journal of Baltic Science Education*, 15(6), 693–705.
- 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*. <https://doi.org/10.1080/00461520.2020.1861948>
- *Zeruth, J. A. (2008). *Promoting probabilistic reasoning: The interplay of learner, task, and text* (2009–99090–593; Issues 11-A) [ProQuest Information & Learning].
- Zielinski, S. F. (2017). *From No to Yes: The Impact of an Intervention on The Persistence of Algebraic Misconceptions among Secondary School Algebra Students*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.