

Inducing Self-Explanation: a Meta-Analysis

Kiran Bisra¹ · Qing Liu¹ · John C. Nesbit¹ ·
Farimah Salimi¹ · Philip H. Winne¹

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract Self-explanation is a process by which learners generate inferences about causal connections or conceptual relationships. A meta-analysis was conducted on research that investigated learning outcomes for participants who received self-explanation prompts while studying or solving problems. Our systematic search of relevant bibliographic databases identified 69 effect sizes (from 64 research reports) which met certain inclusion criteria. The overall weighted mean effect size using a random effects model was $g = .55$. We coded and analyzed 20 moderator variables including type of learning task (e.g., solving problems, studying worked problems, and studying text), subject area, level of education, type of inducement, and treatment duration. We found that self-explanation prompts are a potentially powerful intervention across a range of instructional conditions. Due to the limitations of relying on instructor-scripted prompts, we recommend that future research explore computer-generation of self-explanation prompts.

Keywords Self-explanation · Instructional explanation · Meta-analysis · Prompts

Self-explanation is a constructive cognitive activity learners can enact, at will or in response to a prompt, to comprehend new information or learn skills (Fonseca and Chi 2011). When self-explaining, it is theorized that learners generate inferences about causal connections and conceptual relationships that enhance understanding. The content of self-explanations ranges widely; for example, explanations can describe how a system functions, the effects of serial steps in a procedure, the motives of characters in a story, or concepts presented in expository text (Chi 2000; Siegler 2002).

It has often been observed that students learn steps in a procedure without understanding how each step relates to others or contributes to the goal of the procedure (Siegler 2002). Consequently, learners are less able to transfer the procedure to tasks with differing conditions. Similarly, students studying an expository text may read each sentence but neither connect new

✉ Kiran Bisra
kiran.bisra@sfu.ca

¹ Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada

information to prior knowledge nor consider implications arising from new information. Both situations can be characterized by absent or ineffective metacognition whereby a learner fails to recognize (metacognitive monitoring) and repair gaps (metacognitive control) in understanding.

Self-explanation is conceptualized as a tactic some students spontaneously use to fill in missing information, monitor understanding, and modify fusions of new information with prior knowledge when discrepancies or deficiencies are detected (Chi 2000). Chi considers self-explanation to differ from summarizing, explaining to another, or talking aloud. Self-explanation is directed toward one's self for the purpose of making new information personally meaningful. Because it is self-focused, the self-explanation process may be entirely covert or, if a self-explanation is expressed overtly, it may be intelligible only to the learner.

In early self-explanation research, stable individual differences were observed in the frequency and quality with which students spontaneously self-explained, and positive relationships were reported between self-explanation and learning (Chi et al. 1989; Renkl 1997). Renkl found more successful learners tended to self-explain either by predicting the next step in a problem solution (anticipative reasoning) or identifying the overall goal structure of the problem and the purpose of its subgoals (principle-based explaining).

Because self-explanation could be an effect rather than a cause of learning, observational data showing better learning outcomes for learners who self-explain is not evidence learners benefit from self-explanation training or prompts. Prompts and other exogenous inducements could lead learners to generate explanations less finely adapted to gaps in their knowledge than spontaneous self-explanations. Thus, prompted "self" explanations may be less effective.

Some research has found teaching or prompting self-explanation produces better achievement than comparison conditions (e.g., Lin and Atkinson 2013). These studies typically pose a primary learning activity for all participants and introduce a treatment in which some participants are prompted to self-explain during or after the activity. Lin and Atkinson had college students learn about the cardiovascular system from animated and static visualizations. After studying each visualization, participants in the self-explanation condition were given open-ended questions (e.g., "Could you explain how the blood vessels work?", p. 90) and an empty text box to provide their response to each question. Participants in the comparison condition were given an empty text box where they were told they could write notes. Participants prompted to self-explain performed better on a 20-item multiple-choice posttest than learners prompted to generate notes.

Although several reviews of relevant self-explanation studies support its use as a study skill (Rittle-Johnson and Loehr 2017; Chi and Wylie 2014; Dunlosky et al. 2013), we could find no published report that comprehensively synthesized research investigating the relationship between self-explanation and learning outcomes. Wittwer and Renkl (2010) conducted a meta-analysis exploring the influence of instructional explanations on example-based learning. One moderator variable coded whether participants in the control condition were prompted to self-explain. No effect for this condition was found among the six studies reviewed. A second meta-analysis was conducted on self-explanation in learning mathematics (Rittle-Johnson et al. 2017). The results showed prompting students to self-explain had a small to moderate effect on learning outcomes.

To assess the cognitive effects and instructional effectiveness of self-explanation interventions, we meta-analyzed experimental and quasi-experimental research which compared learning outcomes of learners induced to self-explain to those of learners who were not. In the meta-analysis, we examined moderating variables to investigate how learning outcomes

varied under a range of conditions and to explore their theoretical implications on applying self-explanations.

Self-Explanation Versus Instructional Explanation

One can speculate self-explanation is effective because it has the potential to add information beyond what is given to the learner. That is, learning may be enhanced by the product, not the process, of self-explanation. Hausmann and VanLehn (2010) labeled this the *coverage hypothesis*, describing that self-explanation works by generating “additional content that is not present in the instructional materials” (p. 303). The coverage hypothesis predicts cognitive outcomes of self-explanation are the same as those of a provided instructional explanation. An instructional explanation would presumably be superior in cases where the learner was unable to generate a correct or complete self-explanation.

Several studies examined various effects of self-explanation and instructional explanations. Cho and Jonassen (2012) observed instructional explanations were as effective as self-explanations but outcomes were even better when learners were prompted to compare their self-explanations to instructional explanations. de Koning et al. (2010) demonstrated equivalence between self- and instructional explanations on measures of retention and transfer but learners achieved higher inference scores when instructional explanations were provided. Gerjets et al. (2006) posited both self- and instructional explanations elevated germane load when students studied worked-out examples presented in a modular format. The coverage hypothesis was not supported but these researchers noted design issues with content learners studied. Kwon et al. (2011) contrasted self-explanations to “partially” instructional explanations for which participants filled in small bits of missing information. Self-explaining led to better outcomes. Owing to these diverse findings, our meta-analysis assessed the relative efficacy of self- vs. instructional explanations.

Does It Matter How a Self-Explanation Is Elicited?

Although self-explanation can be induced in various ways, theory is equivocal about their relative efficacy. For example, some researchers induced self-explanation by directions given before a learning activity (e.g., Ainsworth and Burcham 2007). Others provided self-explanation prompts during the learning activity (Hausmann and VanLehn 2010) or after it (Tenenbaum et al. 2008). Supplying prompts periodically during a learning activity may help learners engage in sufficient self-explanation by signaling when there is suitable content to explain. Alternatively, allowing learners to apply their own criteria to decide when and what to self-explain may support constructing explanations better adapted to idiosyncratic prior knowledge.

Several studies used unvarying prompts to avoid reference to specific content. For example, learners studying moves by a chess program were repeatedly asked to predict the next move and “explain why you think the computer will make that move” (de Bruin et al. 2007). In contrast, most research provided content-specific self-explanation prompts, such as “Write your explanation of the diagram in regard to how excessive water intake and lack of water intake influence the amount of urine in the human body” (Cho and Jonassen 2012). Providing more specific prompts might draw learners’ attention to difficult conceptual issues that would

otherwise be overlooked, but it may also interfere with or override learners' use of personal standards for metacognitively monitoring content meriting an explanation.

The format used for self-explanation prompts may affect features of learners' elaborative processing of content. Some researchers asked learners to generate open-ended responses while other researchers provided fill-in-the-blank or multiple-choice questions. As with other prompt characteristics, prompt format may modulate specificity provided by the prompt which, in turn, may affect qualities of elaboration generated in the self-explanation. By containing fewer cues than multiple-choice and fill-in-the-blank formats, an open-ended prompt format may invite elaborative processing better adapted to each learner's unique gaps in knowledge. However, in some contexts, stronger cues present in a multiple-choice question may signal more clearly a possible misconception the learner should address.

Prompts often convey the purpose or nature of the expected self-explanation. Researchers have prompted learners to (a) justify or give reasons for a decision or belief, (b) explain a concept or information in the content, (c) explain a prediction, or (d) make a metacognitive judgment about qualities of their understanding, reasoning, or explanation. When combined with the task context, self-explanation prompts may imply the learner should generate a description of causal relationships, conceptual relationships, or evidence supporting a claim. Of interest in our meta-analysis is the relative effectiveness of the types of self-explanation the inducement is intended to elicit.

Does It Matter that Self-Explanation Takes More Time?

Performing a learning task that invites self-explaining usually requires more time than performing the same learning task without self-explanation. From the perspective of a student interested in optimizing instructional efficacy, it is important to know whether additional time spent on self-explanation might have been better spent otherwise, e.g., re-studying a text or doing additional problems. From a theoretical perspective, the comparative efficiency of self-explaining may depend on the duration and timing of instances of self-explanation in an experiment. The theoretical significance of self-explanation duration and timing is apparent in research by Chi et al. (1994) in which 8th grade students studied an expository text about the human circulatory system. Participants prompted to self-explain after reading each line of text showed greater learning gains than those who read each line a second time. However, the self-explanation group spent almost twice as much time (approximately 1 h longer) studying the materials. Because students who took longer to study the materials were allowed to return for an additional session, those students in the self-explanation group may have benefited from the spaced-learning effect (Carpenter et al. 2012).

In research examining effects of self-explanation prompts on solving mathematics problem by elementary school students, McEldoon et al. (2013) included a comparison group who practiced additional problems to equate for additional time required by the self-explanation group. Over a variety of problem types and learning outcomes in the posttest, there was little difference between the time-matched self-explanation and additional-problem conditions. Unfortunately, most self-explanation research does not report the learning task completion time of each treatment condition and many studies do not even report the mean completion time of all treatment conditions. Nevertheless, where possible, we compared effect sizes of studies in which treatment duration was greater for self-explanation groups to those in which treatment duration was equivalent.

Another time-related question is whether the duration of learning activities is associated with the efficacy of self-explanation inducement. It is plausible that self-explanation helps learners maintain engagement throughout a lengthy task. Alternatively, prompts to self-explain that are initially effective and are internalized by learners may lose their potency over time if they interfere with learners' nascent ability to spontaneously self-explain.

Other Moderators of Interest

Several other variables potentially moderate the benefits of self-explanation inducement. Self-explanation is most commonly used to support either problem solving (including study of worked examples) or text comprehension tasks, and it seems likely that the cognitive effects of self-explanation inducement vary depending on the nature of the primary learning task. The type of knowledge being acquired through the learning activity (i.e., conceptual vs. procedural) is a related factor which we also anticipated may moderate the treatment effects. Because these two variables, task and knowledge type, are more directly related to cognitive operations, we hypothesized they are more likely to moderate treatment effects than subject matter (e.g., science, mathematics, social sciences), a variable we also coded.

Participants learned via several types of instructional media (i.e., digital, print, video), with digital media allowing three types of interactivity (computer-based instruction, intelligent tutoring system, simulation). Computer-based instruction (CBI) was defined as offering response feedback with no student modeling, and intelligent tutoring systems (ITS) were defined as modeling student knowledge to adapt feedback, task selection, or prompts (Ma et al. 2014). We investigated whether the interactivity afforded by digital media, especially the more adaptive forms of interaction provided by intelligent tutoring systems and simulations, would enhance self-explanation effects. Much of the research included diagrams in instructional materials and asked participants to explain or interpret them. We hypothesized explaining diagrams while problem solving may offer opportunities to translate between verbal and visual knowledge representations and thereby strengthen learners' ability to retrieve memories and generate inferences from them (Booth and Koedinger 2012; Chu et al. 2017). We also investigated whether studies that used visual pedagogical agents (i.e., images depicting fictional tutors) to prompt self-explanation might show enhanced treatment effects due to a "persona effect" (Dunsworth and Atkinson 2007; Schroeder and Gotch 2015).

During posttests, particular conditions and requirements of the assessment may allow participants to demonstrate self-explanation effects more fully. The beneficial effects of self-explanation may be evident more on transfer tests than recall tests, and more on long-form test items such as essays and problems than multiple-choice questions. Because interpreting diagrams often requires a more fully elaborated mental model to support translating between visual and verbal representations, tests that include diagrams may be more sensitive to self-explanation effects. To investigate the moderating effects of test characteristics, we coded for assessed learning outcomes (e.g., recall, transfer), test format (e.g., fill-in-the-blank, essay), and whether test items included diagrams.

Students' metacognitive and self-regulatory abilities increase throughout childhood and adolescence (Gestsdottir and Lerner 2008; Kopp 1982; Raffaelli et al. 2005; Schneider 2008; Weil et al. 2013). To investigate whether these developmental changes moderate the effect of

self-explanation inducement, we coded for level of schooling (e.g., elementary, secondary) which was the most reliable proxy for developmental level reported in the primary research.

There is an increasing awareness of how regional or cultural differences affect student learning (Frambach et al. 2012; Marambe et al. 2012). Differences have been reported between geographic regions in students' learning patterns (Marambe et al. 2012) and metacognition (Hartman 2001; Hartman et al. 1996), and these have the potential to moderate self-explanation effects.

Finally, we examined treatment fidelity (Smith et al. 2007), a methodological feature of the primary research that is a key marker of internal validity. Observing a relationship between effect size and methodological quality can help in interpreting meta-analytic results. For example, observing that studies of higher methodological quality tend to find lower effect sizes should decrease our confidence in the effect sizes found by lower quality studies.

Research Goals

We reviewed research that compared learning outcomes when participants were instructed or prompted to self-explain to a condition in which participants were not induced to self-explain. We coded 20 moderator variables to investigate the hypotheses and research questions discussed in the preceding section.

Method

Following the meta-analytic principles described by Borenstein et al. (2009), Lipsey and Wilson (2001), and Hedges and Olkin (1985), we (a) identified all relevant studies, (b) coded study characteristics, and (c) calculated weighted mean effect sizes and assessed their statistical significance.

Identifying Relevant Studies

Using the term *self-expla**, searches of ERIC, PsycINFO, Web of Science, Education Source, Academic Search Elite, and Dissertation Abstracts were carried out to identify and retrieve primary empirical studies of potential relevance. These searches retrieved 1306 studies. Reference sections of review articles (e.g., Atkinson et al. 2000; Chi 2000; VanLehn et al. 1992) were examined to identify candidate studies for inclusion in the meta-analysis. Inclusion criteria were first applied to abstracts. If details provided in abstracts were insufficient to accurately classify a study as appropriate for inclusion, the full text of the study was examined.

To be included, a study was required to:

- Include an experimental treatment in which learners were directed or prompted to self-explain during a learning task. The requirement for self-explanation was not considered satisfied by observing another person self-explain, explaining to another person (i.e., not a self-explanation), rehearsing information, making a prediction without an explanation, or choosing a rule or principle without providing an explanation (Atkinson et al. 2003).
- Provide a comparison treatment in which learners were not directed or prompted to self-explain.

- Avoid confounding the effect of self-explanation by combining it with another study tactic such as summarizing (Chung et al. 1999; Conati and Vanlehn 2000) or providing feedback (Siegler 1995).
- Measure a cognitive outcome such as problem solving or comprehension.
- Use a between-subjects research design.
- Provide sufficient statistical information to compute an effect size, Hedge's *g*.
- Be available in English.

A handful of studies claimed to use think-aloud protocols to elicit self-explanations, especially when participants were young children. However, as Chi (2000) observed, the expected outcomes of think-aloud and self-explanation inducements are quite different:

Think-aloud protocols, often collected in the context of problem-solving research, is a method of collecting verbal data that explicitly forbids reflection. Think-aloud protocols presumably ask the subject to merely state the objects and operators that s/he is thinking of at that moment of the solution process. It is supposed to be a dump of the content of working memory.... Self-explaining is much more analogous to reflecting and elaborating than to "thinking aloud." Talking aloud is simply making overt whatever is going through one's memory (see Ericsson & Simon, 1993), without necessarily exerting the effort of trying to understand. (p. 170)

Furthermore, there is evidence that think-aloud directions can improve performance on some types of problems (e.g., Fox and Charness 2010).

We decided to include studies identified by authors as using a think-aloud method only if the method matched our previously described operationalization of self-explanation inducement and did not provide any prompts to talk aloud. Specifically, we excluded talk aloud studies in which:

- The experimenter reminded participants to keep talking (McNamara 2004; Wong et al. 2002).
- Participants were provided with neutral prompts if they remained quiet (Gadgil et al. 2012).

Applying these criteria, we identified 64 studies reporting 69 independent effect sizes which are examined in the present meta-analysis.

Coding of Study Characteristics

The coding form included 43 fixed-choice items and 24 brief-comment items cataloging information about each study in 8 major categories: reference information (e.g., date of publication), sample information (e.g., participants' ages), materials (e.g., subject domain), treatment group (e.g., prompt type), comparison group (e.g., study tactic), research design (e.g., group assignment), dependent variable (e.g., outcome measure), and effect size computation. In an initial coding phase, three of the authors independently coded ten studies and obtained 86% agreement on the fixed-choice items and 78% agreement on free comment items. All disagreements in this initial phase were discussed and resolved. In the main coding phase, each of the three coders was assigned to code approximately a third of the remaining studies. Bi-weekly meetings were held throughout the main coding phase in

which the coders met with the rest of the research team. During these meetings, borderline cases, discrepancies, and unintended exceptions that arose while coding were resolved. As well, an online forum was used to facilitate discussion in weeks where a weekly meeting did not occur. Meetings and the online communications served four purposes: (1) resolving emerging issues, (2) sharpening category definitions, (3) disseminating definitional refinements to all coders, and (4) creating an archive showing how the group arrived at decisions.

After items in the coding form were used to code the studies, we decided several items should not be analyzed as moderator variables due to a lack of interpretable variance. For example, as a measure of methodological quality, we coded the method of assignment to treatment groups. However, we found 90% of studies randomly assigned individuals to groups by a method supporting high internal reliability (simple random assignment, block randomization, etc.). Almost all the remaining studies failed to report the method of assignment, and only one study reported a quasi-experimental method. Consequently, we decided not to present further analysis of this variable.

Effect Size Extraction

To avoid inflating the weight attributed to any particular study, it is crucial to ensure coded effect sizes are statistically independent (Lipsey and Wilson 2001). To this end, when a study reported more than a single comparison treatment (i.e., treatments in participants were not trained, directed, or prompted to self-explain), the weighted average of all comparison groups was used in calculating the effect size. When repeated measures of an outcome variable were reported, only the last measure was used to calculate the effect size. Lastly, when multiple outcome variables were used to measure learning, we used a combined score calculated using an approach outlined in Borenstein et al. (2009, p.27).

Data were analyzed using random effects models as operationalized in the software program Comprehensive Meta-Analysis (CMA) Version 3.3.070 (Borenstein et al. 2005). Hedge's g , an unbiased mean effect size, was generated for each contrast. Statistics reported are the number of participants (N) in each category, the number of studies (k), the weighted mean effect size (g) and its standard error (SE), the 95% confidence interval around the mean, and a test of heterogeneity (Q) with its associated level of type I error (p).

When significant heterogeneity among levels of a moderator was detected, post-hoc tests using the Z test method described by Borenstein et al. (2009, p.167-168) were further conducted to compare the effect size of each level with all others. The critical value was adjusted for the number of comparisons following the Holm-Bonferroni procedure (Holm 1979) to maintain $\alpha = .05$ for each moderator.

Results

Overall Effect Size for Self-Explanation

As shown in Table 1, a random effects analysis of 69 effect sizes (5917 participants) obtained an overall point estimate of $g = .55$ with a 95% confidence interval ranging from .45 to .65 favoring participants who were prompted or directed to self-explain ($p < .001$). There was statistically detectable heterogeneity, $Q = 196.63$ ($p < .001$, $df = 68$), a result which warrants

Table 1 Overall effect and weighted mean effect sizes for comparison treatments

	<i>N</i>	<i>k</i>	Effect size		95% CI		Test of heterogeneity		
			<i>g</i>	<i>SE</i>	Lower	Upper	<i>QB</i>	<i>df</i>	<i>p</i>
Overall: random-effects model	5917	69	.552*	.050	.454	.650	196.630	68	<.001
Comparison treatment							11.648	3	.009
No additional explanation	3141	41	.673*	.078	.519	.826			
Instructional explanation	573	6	.354*	.092	.175	.533			
Other strategy/technique	516	7	.304*	.091	.126	.483			
Mixed	1687	15	.462*	.083	.301	.624			

* $p < .05$

analyzing moderator variables. Factors other than random sampling accounted for 65% of variance in effect sizes ($I^2 = 65.42$).

Moderator Analysis

The coverage hypothesis can be evaluated by comparing the effects of self-explanation prompts to instructional explanation. Comparison treatments were coded as (a) *no additional explanation* if they provided no intervention beyond the primary instructional task, (b) *instructional explanation* if they provided scripted explanations corresponding to the self-explanation prompts presented to the self-explanation group, (c) *other strategy/technique* if they provided some other intervention beyond the primary instructional task, or (d) *mixed* if they provided two or more of the preceding three types of comparison treatments. Results, shown in Table 1, indicate that each type of comparison treatment was associated with a statistically detectable effect size ($p < .05$) and there was a statistically detectable difference among them. Post-hoc tests indicated studies that provided no additional explanation to the comparison group had effect sizes significantly greater than those that provided instructional explanation ($z = 2.696, p = .007$) and those that provided another strategy or technique ($z = 3.119, p = .002$).

Researchers used methods and formats to induce self-explanation that vary in specificity. These ranged from directive prompts, such as multiple-choice questions, to less directive inducements such as open-ended questions or a general instruction to explain. Researchers also attempted to engage learners in different types of cognitive or metacognitive processes. We used four moderators to characterize the self-explanation inducements: inducement timing, content-specificity, inducement format, and type of self-explanation elicited.

The optimal timing of self-explanation inducements is arguable and may depend on each learner's knowledge and self-explanation ability. Although prompts given during a learning activity model when to self-explain and may induce more frequent self-explanation, compared with instruction to self-explain given before the learning activity, prompts may inhibit learner's choices of content for additional processing. The timing of the self-explanation inducement was coded as (a) *concurrent* if prompts or reminders to self-explain were interspersed with the learning material, (b) *retrospective* if a prompt was provided after participants studied learning materials, and (c) *beginning* if the inducement was training or an instruction to self-explain was given before participants began the learning activity. If a beginning prompt specified the participant was to self-explain after each sentence, paragraph, or problem step and there were no explicit prompts during the learning activity, the inducement was coded as beginning.

The content specificity of inducements was coded as *specific* to named content (e.g., “please self-explain how the valve works”), *general* (e.g., “please self-explain the last paragraph”), or *both* if the treatment included both content-specific and content-general inducements.

Inducement format was coded as (a) *interrogative* if a prompt asked the participant a question, (b) *imperative* if a prompt directed the participant to self-explain, (c) *predirected* if the participant was told at the beginning of the activity to self-explain, (d) *fill-in-the-blank* if the participant was asked to complete a sentence with a self-explanation, (e) *multiple choice* if the participant was to choose a self-explanation from a list of explanations, or (f) *mixed* if participants were provided more than one inducement type. Although the overall comparison of inducement types returned $p = .048$, post-hoc tests found no significant differences among them.

Lastly, the type of self-explanation elicited by the inducement was coded as (a) *justify* if the participant was asked to provide reasoning for a choice or decision, (b) *conceptualize* if the inducement asked the participant to define or elaborate the meaning of a named concept presented in the material (e.g., iron oxide), (c) *explain* if the inducement asked the participant to explain a structural part of the material (e.g., the last paragraph), (d) *mixed* when the inducement elicited more than one type of response, (e) *justify for another* if the inducement asked the participant to provide reasoning for someone else’s choice or decision, (f) *metacognitive* if the inducement asked the participant to self-explain their planning or performance, (g) *anticipatory* if the prompt asked participants to predict the next step in a problem solution and self-explain why they believed this was the correct next step, and (h) *other* if the inducement asked participants to provide another type of self-explanation. Table 2 shows all categories of inducement were associated with a statistically detectable effect except multiple-choice inducements and metacognitive self-explanations. The overall comparison of levels returned $p < .001$, and post-hoc tests indicated that the prompts eliciting conceptual self-explanations ($z = 3.284$, $p = .001$) were more effective than those inducing metacognitive self-explanations. Post-hoc tests also suggested that prompts inducing anticipatory self-explanations were more effective than those inducing metacognitive self-explanations, but this outcome can be discounted because there was only one study inducing anticipatory self-explanations and its large effect size could be due to any of its particular characteristics.

The length of time spent utilizing a study strategy and engaging with learning materials may impact learning outcomes. Therefore, in examining the problem of differences in time on task between intervention and comparison groups discussed in the introduction, we found most studies did not report separate mean durations for each group’s study activities nor statistically compared durations across groups. For studies that did statistically compare duration, those in which the self-explanation treatment group took detectably longer were coded as *greater for SE group* and those in which there was no detectable difference in time were coded *equal for both groups*. Second, we coded studies according to the overall duration of the learning activity. Possibly, a brief learning phase might produce only transitory effects due to novelty of the intervention, or it might not allow enough time for participants to learn how to respond to inducements (Table 3).

To capture intended learning goals, three moderator variables were defined to characterize the learning task and the type of knowledge students were expected to acquire. Researchers have tended to investigate self-explanation in the context of problem solving (including worked examples) or text comprehension. Although these two learning activities are usually treated separately in the wider arena of educational research, and one might anticipate they

Table 2 Weighted mean effect sizes for self-explanation inducements

	<i>N</i>	<i>k</i>	Effect size		95% CI		Test of Heterogeneity		
			<i>g</i>	<i>SE</i>	Lower	Upper	<i>QB</i>	<i>df</i>	<i>p</i>
Inducement timing							2.685	2	.261
Concurrent	5187	57	.521*	.054	.415	.627			
Retrospective	533	8	.571*	.089	.397	.746			
Beginning	197	4	1.241*	.449	.361	2.121			
Content specificity							1.836	2	.399
Specific	4164	41	.510*	.058	.396	.625			
General	1419	22	.678*	.110	.461	.894			
Both	334	6	.511*	.183	.152	.869			
Inducement format							12.693	6	.048
Interrogative	3177	36	.559*	.070	.422	.696			
Imperative	720	12	.395*	.118	.163	.627			
Predirected	441	8	.700*	.104	.495	.905			
Fill-in-the-blank	279	2	.895*	.145	.612	1.179			
Multiple choice	204	2	.242	.250	-.247	.731			
Mixed	1043	8	.579*	.148	.290	.869			
Not reported	53	1	1.031*	.292	.458	1.604			
Self-explanation elicited							27.019	7	<.001
Justify	1700	20	.416*	.079	.261	.570			
Conceptualize	1678	13	.873*	.158	.563	1.183			
Explain	628	12	.678*	.111	.461	.896			
Mixed	677	8	.703*	.130	.449	.957			
Justify for another	530	8	.426*	.103	.224	.628			
Metacognitive	241	3	.192	.135	-.073	.456			
Anticipatory	62	1	1.372*	.328	.730	2.015			
Other	401	4	.322*	.114	.099	0.545			

* $p < .05$

would interact differently with self-explanation inducements, self-explanation theory deals with and thereby emphasizes their common elements. For example, both learning activities call for cognitive strategies that require metacognitive monitoring, a process which may detect gaps in understanding and productively trigger self-explanation. To determine if outcomes

Table 3 Weighted mean effect sizes for learning task duration

	<i>N</i>	<i>k</i>	Effect size		95% CI		Test of heterogeneity		
			<i>g</i>	<i>SE</i>	Lower	Upper	<i>QB</i>	<i>df</i>	<i>p</i>
Duration differences among groups							4.062	3	.255
Not reported	4230	47	.510*	.058	.396	.624			
Greater for SE group	1336	16	.722*	.120	.487	.957			
Equal for all groups	291	5	.411*	.124	.169	.653			
Greater for non-SE group	60	1	.352*	.260	-.158	.862			
Duration for all groups							4.309	4	.366
< 30 min	1842	17	.709*	.107	.500	.919			
30–60 min	880	13	.601*	.075	.455	.747			
1–2 h	341	6	.444*	.143	.164	.724			
> 2 h	436	6	.462*	.150	.168	.756			
Not reported	2418	27	.475*	.086	.308	.643			

* $p < .05$

were affected by type of learning task, the task common to all treatment groups was considered the primary learning activity and was coded as solve problem, study text, study worked example, study case, study simulation, or other. Learning activities that combined two or more of these types were coded as mixed. As shown in Table 4, each type of primary learning activity was associated with a statistically detectable mean effect size except studying simulation, for which there were only two studies. There was no statistically detectable difference among the different types of primary learning activity.

The knowledge type was coded as *conceptual*, *procedural*, or *both*. Often students study texts to develop conceptual knowledge, and they solve problems to develop a combination of conceptual and procedural knowledge. However, the type of learning task assigned does not reliably indicate the type of knowledge students are intended to acquire. One can productively study a text about how to execute a skill, and problem solving can be undertaken solely for conceptual insights it affords. Learning materials were coded according to their subject area (e.g., mathematics, social sciences). Associations between inducement to self-explain and learning outcome were statistically detected for each knowledge type and subject. No differences were statistically detected among the levels of each variable.

Features of the learning environment have potential to moderate effects of self-explanation inducement. Prompts may offer greater benefit in learning environments lacking interactive and adaptive features and, aside from the prompts, present only text to be studied or problems to be solved. More adaptive environments such as intelligent tutoring systems (Ma et al. 2014) are designed to detect gaps in individual learner's knowledge and provide remediating instruction or feedback, possibly obtaining outcomes similar to self-explanation via a different route. On the other hand, self-explanation prompts delivered by a visual pedagogical agent may have amplified effects due to greater salience or inducing greater compliance than prompts embedded in a studied text or problem. As shown in Table 5, learning environments

Table 4 Weighted mean effect sizes for learning task

	<i>N</i>	<i>k</i>	Effect size		95% CI		Test of heterogeneity		
			<i>g</i>	<i>SE</i>	Lower	Upper	<i>QB</i>	<i>df</i>	<i>p</i>
Task type							10.147	6	.119
Solve problems	1415	20	.476*	.094	.292	.660			
Study text	772	14	.787*	.154	.485	1.089			
Study worked examples	649	8	.362*	.129	.109	.615			
Study cases	324	4	.432*	.152	.134	.730			
Study simulation	122	2	.241*	.199	-.149	.630			
Other	1678	13	.552*	.088	.381	.724			
Mixed	957	8	.773*	.147	.486	1.061			
Knowledge type							.861	2	.650
Conceptual	2540	33	.598*	.075	.451	.745			
Procedural	204	3	.468*	.167	.142	.795			
Both	3173	33	.516*	.072	.374	.658			
Subject							3.855	5	.571
Science	1582	25	.565*	.070	.429	.701			
Math	1808	16	.441*	.097	.251	.632			
Social sciences	413	6	.555*	.191	.180	.929			
Computer science	818	9	.761*	.180	.407	1.114			
Mixed	93	1	.313*	.221	-.121	.746			
Other	1203	12	.591*	.139	.319	.863			

* $p < .05$

Table 5 Weighted mean effect sizes for learning environment characteristics

	<i>N</i>	<i>k</i>	Effect size		95% CI		Test of heterogeneity		
			<i>g</i>	<i>SE</i>	Lower	Upper	<i>QB</i>	<i>df</i>	<i>p</i>
Media type							2.677	4	.613
Digital	4255	42	.550*	.062	.428	.672			
Print	1223	21	.550*	.105	.344	.755			
Video	60	1	.833*	.284	.276	1.390			
Other	335	4	.630*	.129	.378	.883			
Not reported	44	1	.246	.263	-.270	.762			
Interactivity							2.269	3	.519
CBI	1130	7	.587*	.155	.282	.891			
ITS	692	7	.483*	.190	.110	.857			
Simulation	301	5	.379*	.122	.139	.618			
None	3794	50	.575*	.060	.458	.693			
Diagram in materials							.385	2	.825
Yes	1435	23	.588*	.102	.387	.788			
No	4297	43	.532*	.059	.416	.648			
Not reported	185	3	.661*	.283	.107	1.215			
Visual pedagogical agent							11.307	2	.004
Yes	199	3	.641	.336	-.018	1.300			
No	5613	65	.533*	.050	.435	.630			
Not reported	105	1	1.275*	.215	.853	1.697			

* $p < .05$

were coded and analyzed for four moderator variables: media type, interactivity, diagrams in materials, and visual pedagogical agent. The medium of the learning materials was coded as (a) *digital* when materials were presented on a computer, (b) *print* if materials were printed on paper, (c) *video* when the participants were asked to learn from a video, (d) *other* when the learning materials could not be placed in the previously stated categories. The type of interactivity was coded as computer-based instruction (CBI), intelligent tutoring system (ITS), simulation, or none. Learning materials were coded for whether or not diagrams were included. Providing diagrams in learning materials suggests visuospatial processing was a component of the intended learning goal. Associations between inducement to self-explain and learning outcome were statistically detected regardless of media type, interactivity, and whether diagrams were present or not. No differences were statistically detected between the levels of each variable.

A visual pedagogical agent was defined as a simulated person that communicated to participants and was represented by a static or animated face. Although the overall comparison among levels of the pedagogical agent moderator returned $p = .004$, post-hoc tests found only that a single study with non-reported pedagogical agent status outperformed studies that did not use a pedagogical agent to present self-explanation prompts—a difference that affords little meaningful interpretation.

Guided by the theory of self-explanation as explained in the introduction, we hypothesized that self-explanation would have a small beneficial effect on measures assessing comprehension and recall, and a larger beneficial effect on measures assessing transfer and inference. As shown in Table 6, the learning outcome was coded as *comprehension*, *inference*, *recall*, *problem solving*, *transfer*, *mixed* (if several outcomes were tested), or *other* depending on how it was described by the researchers. For example, the learning outcome in Ainsworth and Burcham (2007) was coded as inference because they described posttest questions as requiring

Table 6 Weighted mean effect sizes for learning outcome assessment

	<i>N</i>	<i>k</i>	Effect size		95% CI		Test of heterogeneity		
			<i>g</i>	<i>SE</i>	Lower	Upper	<i>QB</i>	<i>df</i>	<i>p</i>
Learning outcome							6.072	6	.415
Comprehension	187	3	.282	.151	-.014	.579			
Inference	78	2	1.794*	.854	.120	3.469			
Recall	398	6	.498*	.123	.256	.740			
Problem solving	1077	14	.445*	.111	.227	.664			
Transfer	1169	17	.534*	.094	.350	.718			
Mixed	999	15	.597*	.113	.377	.818			
Other	2009	12	.596*	.118	.365	.827			
Test format							5.156	6	.524
Essay questions	176	3	.624	.413	-.185	1.433			
Fill-in-the-blank questions	55	1	.624*	.293	.048	1.199			
Multiple-choice questions	579	7	.410*	.092	.229	.591			
Problem questions	949	10	.547*	.127	.297	.796			
Short-answer questions	1217	19	.671*	.122	.432	.910			
Mixed	1641	16	.436*	.100	.240	.632			
Other	1300	13	.650*	.113	.427	.872			
Diagram in test							.941	2	.625
Yes	1178	18	.657*	.128	.405	.908			
No	4613	49	.522*	.054	.417	.628			
Not reported	126	2	.512	.376	-.225	1.249			

* $p < .05$

“generation of new knowledge by inferring information from the text” (p. 292). The studies coded as assessing problem solving dealt with problems ranging from diagnosing medical cases to geometry problems. While not coded as transfer, these could be regarded as assessing problem solving transfer because they posed posttest problems that differed to some degree from problems solved or studied in the learning phase. The effect sizes for all types of learning outcomes, except comprehension, were statistically significant and no differences between them were detected.

Also shown in Table 6, learning assessments were coded according to test format and whether the test asked participants to complete or label a diagram. Test format was coded as *essay questions*, *fill-in-the-blank questions*, *multiple choice questions*, *problem questions*, *short-answer questions*, *mixed*, and *other*. The effect sizes for tests with or without diagrams and all test formats except essay questions were statistically significant. No differences among the categories of each moderator were statistically detected.

Because participants’ ages (in years) were not reliably reported, we assessed whether the effect of self-explanation inducements varied with level of schooling. As shown in Table 7, statistically detectable effects were found at all levels of schooling. No difference was detected among levels.

Social science research has been criticized for using data gathered only in Western societies to make claims about human psychology and behavior (Henrich et al. 2010). We were curious about the extent to which the research we reviewed was subject to similar sampling biases. We coded the region in which the research took place as North America, Europe, East Asia, and Australia/New Zealand. As shown in Table 7, most of the studies were conducted in North America, and most of the remainder was conducted in Europe. A comparison among regions returned $p = .011$. Post-hoc tests indicated that studies conducted in East Asia had effect sizes

Table 7 Weighted mean effect sizes for grade level and region

	<i>N</i>	<i>k</i>	Effect size		95% CI		Test of heterogeneity		
			<i>g</i>	<i>SE</i>	Lower	Upper	<i>QB</i>	<i>df</i>	<i>p</i>
Education level							3.051	3	.384
Elementary	849	10	.483*	.104	.279	.687			
High school	1415	13	.426*	.091	.247	.606			
Undergraduate	3349	42	.611*	.071	.472	.751			
Professional program	304	4	.680*	.291	.110	1.250			
Region							13.135	4	.011
North America	3456	42	.525*	.064	.399	.650			
Europe	1791	18	.481*	.071	.342	.620			
East Asia	456	6	1.191*	.204	.792	1.591			
Australia/New Zealand	121	2	.453	.490	-.508	1.414			
Mixed	93	1	.212	.219	-.217	.641			

* $p < .05$

significantly greater than those conducted in North America ($z = 3.105, p = .002$), Europe ($z = 3.275, p = .001$), and mixed ($z = 3.263, p = .001$).

Finally, we coded a treatment fidelity variable that indicated whether researchers ensured the group induced to self-explain actually engaged in self-explanation. Studies were coded as (a) *Yes, analyzed* if the researchers recorded or examined the self-explanations and verified that participants self-explained, (b) *No* if participants' responses were not verified as legitimate self-explanations, (c) *Yes, followed-up* if the participant was prompted again when the initial self-explanation was insufficient, and (d) *not reported* if it could not be determined whether the provided self-explanations were verified. As shown in Table 8, all four categories of effect sizes were statistically detectable and no differences among them were detected.

Publication Bias

Two statistical tests were computed with comprehensive meta-analysis to further examine the potential for publication bias. First, a classic fail-safe N test was computed to determine the number of null effect studies needed to raise the p value associated with the average effect above an arbitrary level of type I error (set at $\alpha = .05$). This test revealed that 6028 additional studies would be required to fail to confirm the overall effect found in this meta-analysis. This large fail-safe N implies our results are robust to publication bias when judged relative to a threshold for interpreting fail-safe N as $\geq 5k + 10$ (Rosenthal 1995), where k is the number of studies included

Table 8 Weighted mean effect sizes for methodology

	<i>N</i>	<i>k</i>	Effect size		95% CI		Test of heterogeneity		
			<i>g</i>	<i>SE</i>	Lower	Upper	<i>QB</i>	<i>df</i>	<i>p</i>
Treatment fidelity							4.735	3	.192
Yes, analyzed	2014	31	.613*	.085	.447	.779			
No	2868	21	.409*	.076	.259	.558			
Yes, followed-up	682	11	.587*	.089	.411	.762			
Not reported	353	6	.732*	.221	.299	1.165			

* $p < .05$

in the meta-analysis. Orwin's fail-safe N , a more stringent publication bias test, revealed 274 missing null studies would be required to bring the mean effect size found in this meta-analysis to under 0.1 (Orwin 1983). Although there is potential for publication bias in this meta-analysis, results of both tests suggest it does not pose a plausible threat to our findings.

Discussion

In a review of over 800 meta-analyses relating to educational achievement, Hattie (2009) identified some of the most efficacious instructional interventions as reciprocal teaching ($d = .74$), feedback ($d = .73$), and spaced learning ($d = .71$). Our results indicate inducement to self-explain, with a mean weighted effect size of $g = .55$, offers benefits similar in magnitude to interventions such as mastery learning ($d = .50$) and peer tutoring ($d = .55$). In almost all the categories we examined, inducements to self-explain were associated with statistically detectable benefits. The only exceptions were categories represented by a small number of studies. Only three moderator variables showed statistically detectable differences that have implications for theory or practice: (a) comparison treatments, (b) type of self-explanation elicited, and (c) region.

Self-Explanation Outperforms Instructional Explanation

Contrary to the coverage hypothesis (Hausmann and VanLehn 2010), which describes the effects of self-explanation as adding information that instead might be supplied by an instructor or instructional system, our results showed a statistically detectable advantage ($g = .35$) of self-explanation over instructional explanations. We attribute this result to cognitive processes learners use when generating an explanation and/or the idiosyncratic match to prior knowledge of the new knowledge generated by self-explaining. By retrieving relevant previously learned information from memory and elaborating it with relevant features in the new information, we hypothesize that meaningful associations are formed; constructing the explanation engages fundamental cognitive processes involved in understanding the explanation, recalling it later, and using it to form further inferences. At the same time, since self-explanation compared to no additional explanation yielded a detectably larger effect size ($g = .67$), a substantial portion of the benefit reported for self-explanation appears to be alternately available from instructor-provided explanations. Considering the difficulties learners often have in generating complete and correct self-explanations, Renkl (2002) proposed that optimal outcomes might be attained by first prompting learners to self-explain and then providing an instructional explanation upon request. However, Schworm and Renkl (2006) reported that participants who received only self-explanation prompts outperformed participants who received self-explanation prompts plus supplementary instructional explanations on demand. Schworm and Renkl argued making a relevant instructional explanation available undermines learners' incentive to self-explain effortfully, thus depriving them of cognitive benefits they would otherwise receive.

How Self-Explanation Is Elicited

Our introduction discussed how the cue strength associated with different prompt formats might affect learning outcomes. In the meta-analysis, multiple-choice prompts were associated

with an effect size ($g = .24$) not statistically different from zero (Table 2). Although the small number of studies using multiple-choice prompts ($k = 2$) warrants caution, it may be that the greater cue strength of multiple-choice prompts undermines self-explanation effects. To address the question of optimal cue strength more thoroughly, we recommend that further research directly compare the effects of multiple-choice prompts, fill-in-the-blank prompts, and open-ended prompts. Optimal cue strength likely depends on each student's ability to self-explain the concepts or procedures they are learning. When first introduced to a topic, students may benefit more from strongly cued self-explanation prompts, and as their knowledge increases they may benefit more from less strongly cued prompts.

We found all types of elicited self-explanations yielded a detectable effect size except metacognitive prompts (Table 2). A metacognitive prompt engages the learner in a self-explanation of planning or performance, and therefore we speculate that learners might not address content in these explanations. That is, content measured by the learning outcome may not be involved in responding to a metacognitive prompt. If so, it is not surprising learning outcomes are variable. Future research might explore different kinds of effects resulting from metacognitive prompts, e.g., increased planning when presented additional problems to solve.

Anticipatory self-explanations ($g = 1.37$) had substantially better learning outcomes than those in which participants justified their own decisions ($g = .42$) or the decisions of another ($g = .43$). Since only one study induced anticipatory self-explanations, the large effect size could be due to any of its particular characteristics. We suggest future research should investigate the efficacy of anticipatory self-explanations.

Findings of our meta-analysis regarding timing, content specificity, and format of inducement (Table 2) as well as the nature of the self-explanation elicited (Table 2) show variation in these factors rarely matters when learners self-explain. The same was the case for variations of the learning task (Table 4) and the learning environment (Table 5). As well, variation in learning outcome and test type rarely mattered. In almost all these cases, self-explaining benefited learning outcomes. We note that the absence of statistically detectable aggregate effect sizes in three cases: multiple-choice prompts, metacognitive self-explanations, and measures of learning outcomes formatted as essays. These invite considerations for future research.

Self-Explanation Usually Takes More Time. Does that Matter?

Although the effect size for studies in which self-explanation took more time than the comparison treatment ($g = .72$) was greater than for studies in which self-explanation took a similar amount of time ($g = .41$), the difference was not statistically significant. We interpret these results as showing that inducing self-explanation is a time-efficient instructional method but, in some research, its efficiency may be exaggerated because time-on-task was not controlled or reported. Most research on self-explanation has, regrettably, not fully reported time-on-task. We advocate reporting the mean and standard deviation of learning task duration for all treatment groups in self-explanation research.

Why Did Studies from East Asia Return High Effect Sizes?

Looking more closely at the six studies conducted in East Asia, we found that almost all had participants study texts as the learning task (Table 4) and gave comparison treatments providing no additional explanation or alternate learning strategies (Table 1). These two

conditions were associated with relatively large effect sizes, and we speculate their confluence in the East Asian studies led to a significantly larger effect size for that region. An alternative interpretation rests on evidence Asian students' learning strategies, shaped by cultural and educational contexts, tend toward reproduction-oriented studying aiming for verbatim recall on tests (Biemans and Van Mil 2008; Marambe et al. 2012; Vermunt 1996; Vermunt and Donche 2017). If, as a result of this orientation, learners in East Asia are less likely to engage spontaneously in self-explanation, they may receive greater benefit from prompts to self-explain than learners who are more likely to self-explain without being prompted.

Automatic Generation of Self-Explanation Prompts

In all the research we reviewed, self-explanation prompts or pre-study instructions were scripted by researchers or instructors. In most cases, the prompts were derived from the content of the learning task and were not generic statements that could be re-used with other content. Pre-scripted, instructor-generated prompts may be suitable for instructional settings in which the same readings or problems are assigned to many students over repeated courses or modules, but they cannot be delivered at scale when the content is highly differentiated or personalized. In resource-inquiry models of instruction (e.g., problem-based learning conducted in an environment with access to the internet and resource databases), students collaborate to identify information needs that are satisfied by online searches (Nesbit and Winne 2003). In such settings, computer-generation of content-specific prompts may be the most feasible method for supporting self-explanation.

Two of the authors have collaborated with computer scientists in developing and evaluating algorithms that automatically generate questions for expository texts (Lindberg et al. 2013; Odilinye et al. 2015). The premise underlying this work is that a computer-generated question can serve as a self-explanation prompt even when the computer has no way of assessing the correctness of a student's explanation or when the answer to the question is not explicitly represented in the source text. Systems like this could be used to scaffold students' skills in self-explaining, even if students are engaged in academic tasks where the materials they study are not controlled, such as researching material for writing an essay. If a learner model is part of such a system, the prompts could be adapted to match each learner's subject knowledge, reading vocabulary, and self-regulatory ability.

Conclusions

Our findings have significant practical implications. The foremost is that having learners generate an explanation is often more effective than presenting them with an explanation. Another major implication for teaching and learning is that beneficial effects of inducing self-explanation seem to be available for most subject areas studied in school, and for both conceptual (declarative) and procedural knowledge. The most powerful application of self-explanation may arise after learners have made an initial explanation and then are prompted to revise it when new information highlights gaps or errors.

Research on self-explanation has arrived at a stage where the efficacy of the learning strategy is established across a range of situations. There is now a need for clearer mapping of the unique cognitive benefits self-explanation may promote, the specific effects of different types of self-explanations and prompts, and how self-explanation might be optimally

combined and sequenced with other instructional features such as extended practice, explanation modeling, and inquiry-based learning. Future research that investigates these questions should be designed to directly compare different self-explanation conditions. Effective as self-explanation prompts may be, the primary goal of future research should be to identify strategies whereby prompts are faded so that self-explanation becomes fully self-regulated.

Funding This study was funded by Social Sciences and Humanities Research Council of Canada (grant number 435–2012-0723).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Articles included in the meta-analyses are marked with an asterisk

- *Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self-explanation. *Learning and Instruction, 17*(3), 286–303.
- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181–214.
- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774–783.
- *Berthold, K., & Renkl, A. (2009). Instructional aides to support a conceptual understanding of multiple representations. *Journal of Educational Psychology, 101*(1), 70–87.
- *Berthold, K., Nueckles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction, 17*(5), 564–577.
- *Berthold, K., Eysink, T. H. S., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science: An International Journal of the Learning Sciences, 37*(4), 345–363.
- Biemans, H., & Van Mil, M. (2008). Learning styles of Chinese and Dutch students compared within the context of Dutch higher education in life sciences. *Journal of Agricultural Education and Extension, 14*(3), 265–278.
- *Bodvarsson, M. C. (2005). *Prompting students to justify their response while problem solving: A nested, mixed-methods study* (unpublished doctoral dissertation). University of Nebraska-Lincoln, USA.
- Booth, J. L., & Koedinger, K. R. (2012). Are diagrams always helpful tools? Developmental and individual differences in the effect of presentation format on student problem solving. *British Journal of Educational Psychology, 82*(3), 492–511.
- *Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction, 25*, 24–34.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2005). *Comprehensive meta-analysis, version 3.3.070 [computer software]*. Englewood, NJ: Biostat Inc (Englewood, NJ).
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.
- Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: review of recent research and implications for instruction. *Educational Psychology Review, 24*(3), 369–378.
- *Chamberland, M., St-Onge, C., Setrakian, J., Lanthier, L., Bergeron, L., Bourget, A., et al. (2011). The influence of medical students' self-explanations on diagnostic performance. *Medical Education, 45*(7), 688–695.
- Chi, M. T. H. (2000). Self-explaining expository texts: the dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in Instructional Psychology* (pp. 161–238). Hillsdale: Lawrence Erlbaum Associates.

- Chi, M., & Wylie, R. (2014). The ICAP framework: linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145–182.
- *Chi, M. T. H., DeLeeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477.
- *Cho, Y. H., & Jonassen, D. H. (2012). Learning by self-explaining causal diagrams in high-school biology. *Asia Pacific Education Review, 13*(1), 171–184.
- *Chou, C., & Liang, H. (2009). Content-free computer supports for self-explaining: Modifiable typing interface and prompting. *Educational Technology & Society, 12*(1), 121–133.
- Chu, J., Rittle-Johnson, B., & Fyfe, E. R. (2017). Diagrams benefit symbolic problem-solving. *British Journal of Educational Psychology, 87*(2), 273–287.
- Chung, S., Chung, M.J., & Severance, C. (1999). Design of support tools and knowledge building in a virtual university course: effect of reflection and self-explanation prompts. Paper presented at the WebNet 99 World Conference on the WWW and Internet Proceedings, Honolulu, Hawaii. (ERIC Document Reproduction Service No. ED448706).
- *Chung, S., Severance, C., & Chung, M. J. (2003). Design of support tools for knowledge building in a virtual university course. *Interactive Learning Environments, 11*(1), 41–57.
- *Coleman, E. B., Brown, A. L., & Rivkin, I. D. (1997). The effect of instructional explanations on learning from scientific texts. *The Journal of the Learning Sciences, 6*(4), 347–365.
- Conati, C., & Vanlehn, K. (2000). Toward computer-based support of meta-cognitive skills: a computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education (IJAIED), 11*, 389–415.
- *Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level-reasoning-question effect: the role of dialogue and deep-level-reasoning questions during vicarious learning. *Cognition and Instruction, 24*(4), 565–591.
- *Crippen, K. J., & Earl, B. L. (2007). The impact of web-based worked examples and self-explanation on performance, problem solving, and self-efficacy. *Computers and Education, 49*(3), 809–821. <https://doi.org/10.1016/j.compedu.2005.11.018>.
- *de Bruin, A. B., Rikers, R. M., & Schmidt, H. G. (2007). The effect of self-explanation and prediction on the development of principled understanding of chess in novices. *Contemporary Educational Psychology, 32*(2), 188–205.
- *de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Learning by generating vs. receiving instructional explanations: two approaches to enhance attention cueing in animations. *Computers & Education, 55*(2), 681–691.
- *DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology, 113*(4), 552–568.
- Dunlosky, J., Rawson, K., Marsh, E., Nathan, M., & Willingham, D. (2013). Improving students' learning with effective learning techniques. *Psychological Science in the Public Interest, 14*(1), 4–58.
- Dunsworth, Q., & Atkinson, R. K. (2007). Fostering multimedia learning of science: exploring the role of an animated agent's image. *Computers & Education, 49*(3), 677–690.
- *Earley, C. E. (1998). *Expertise acquisition in auditing: Training novice auditors to recognize cue relationships in real estate valuation*. (9906361 Ph.D.), Ann Arbor: University of Pittsburgh.
- *Earley, C. E. (2001). Knowledge acquisition in auditing: training novice auditors to recognize cue relationships in real estate valuation. *The Accounting Review, 76*(1), 81–97.
- *Eckhardt, M., Urhahne, D., Conrad, O., & Harms, U. (2013). How effective is instructional support for learning with computer simulations? *Instructional Science, 41*(1), 105–124.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis* (2nd ed.). Cambridge: Bradford/MT Press.
- *Eysink, T. H. S., de Jong, T., Berthold, K., Kolloffel, B., Opfermann, M., & Wouters, P. (2009). Learner performance in multimedia learning arrangements: an analysis across instructional approaches. *American Educational Research Journal, 46*(4), 1107–1149.
- Fonseca, B. A., & Chi, M. T. (2011). The self-explanation effect: A constructive learning activity. In R. Mayer & P. Alexander (Eds.), *Handbook of Research on Learning and Instruction* (pp. 270–321). Routledge Press.
- Fox, M. C., & Charness, N. (2010). How to gain eleven IQ points in ten minutes: thinking aloud improves Raven's matrices performance in older adults. *Aging, Neuropsychology, and Cognition, 17*(2), 191–204. <https://doi.org/10.1080/13825580903042668>.
- Frambach, J. M., Driessen, E. W., Chan, L. C., & van der Vleuten, C. P. (2012). Rethinking the globalisation of problem-based learning: how culture challenges self-directed learning. *Medical Education, 46*(8), 738–747.
- *Fukaya, T. (2013). Explanation generation, not explanation expectancy, improves metacomprehension accuracy. *Metacognition and Learning, 8*(1), 1–18.

- Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. H. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction, 22*(1), 47–61.
- *Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction, 16*(2), 104–121.
- Gestsdottir, S., & Lerner, R. M. (2008). Positive development in adolescence: the development and role of intentional self-regulation. *Human Development, 51*(3), 202–224.
- *Graf, E. (2000). *Designing a computer tutorial to correct a common student misconception in mathematics* (unpublished doctoral dissertation). University of Washington, USA.
- *Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition, 36*(1), 93–103.
- *Große, C. S., & Renkl, A. (2006). Effects of multiple solution methods in mathematics learning. *Learning and Instruction, 16*(2), 122–138.
- Hartman, H. J. (2001). Developing students' metacognitive knowledge and skills. In *Metacognition in learning and instruction* (pp. 33–68). Netherlands: Springer.
- Hartman, H. J., Everson, H. T., Toblas, S., & Gourgey, A. F. (1996). Self-concept and metacognition in ethnic minorities: predictions from the BACEIS model. *Urban Education, 31*(2), 222–238.
- Hattie, J. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Abingdon. Abingdon: Routledge.
- *Hausmann, R. G., & VanLehn, K. (2007). Explaining self-explaining: a contrast between content and generation. *Frontiers in Artificial Intelligence and Applications, 158*, 417–424.
- *Hausmann, R. G., & VanLehn, K. (2010). The effect of self-explaining on robust learning. *International Journal of Artificial Intelligence in Education, 20*(4), 303–332.
- Hedges, L., & Olkin, I. (1985). *Statistical models for meta-analysis* (Vol. 6). New York: Academic Press.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>.
- *Hilbert, T. S., & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Computers in Human Behavior, 25*(2), 267–274.
- *Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: learning from heuristic examples and how it can be supported. *Learning and Instruction, 18*(1), 54–65.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.
- *Honomichl, R. D., & Chen, Z. (2006). Learning to align relations: the effects of feedback and self-explanation. *Journal of Cognition and Development, 7*(4), 527–550.
- *Huk, T., & Ludwigs, S. (2009). Combining cognitive and affective support in order to promote learning. *Learning and Instruction, 19*(6), 495–505.
- *Jonas, I. G., Cernusca, D., & Collier, H. L. (2012). Prior knowledge influence on self-explanation effectiveness when solving problems: an exploratory study in science learning. *International Journal of Teaching and Learning in Higher Education, 24*(3), 349–358.
- *Kapli, N. V. (2010). *The effects of segmented multimedia worked examples and self-explanations on acquisition of conceptual knowledge and problem-solving performance in an undergraduate engineering course*. (3442880 Ph.D.), The Pennsylvania State University, Ann Arbor.
- *Kastens, K. A., & Liben, L. S. (2007). Eliciting self-explanations improves children's performance on a field-based map skills task. *Cognition and Instruction, 25*(1), 45–74.
- Kopp, C. B. (1982). Antecedents of self-regulation: a developmental perspective. *Developmental Psychology, 18*(2), 199–214.
- *Kramarski, B., & Dudai, V. (2009). Group-metacognitive support for online inquiry in mathematics with differential self-questioning. *Journal of Educational Computing Research, 40*(4), 377–404.
- *Kwon, K., Kumalasari, C. D., & Howland, J. L. (2011). Self-explanation prompts on problem-solving performance in an interactive learning environment. *Journal of Interactive Online Learning, 10*(2), 96–112.
- *Lee, A. Y., & Hutchison, L. (1998). Improving learning from examples through reflection. *Journal of Experimental Psychology: Applied, 4*(3), 187–210.
- Lin, L., & Atkinson, R. K. (2013). Enhancing learning from different visualizations by self-explanation prompts. *Journal of Educational Computing Research, 49*(1), 83–110.
- Lindberg, D., Popowich, F., Nesbit, J., & Winne, P. (2013). Generating natural language questions to support learning on-line. In *Proceedings of the 14th European Workshop on Natural Language Generation* (pp. 105–114). Sofia, Bulgaria.
- Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis* (Vol. 49). Thousand Oaks: Sage publications.

- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: a meta-analysis. *Journal of Educational Psychology, 106*(4), 901–918.
- Marambe, K. N., Vermunt, J. D., & Boshuizen, H. P. (2012). A cross-cultural comparison of student learning patterns in higher education. *Higher Education, 64*(3), 299–316.
- *Mayer, R. E., & Johnson, C. I. (2010). Adding instructional features that promote learning in a game-like environment. *Journal of Educational Computing Research, 42*(3), 241–265.
- *Mayer, R. E., Dow, G. T., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: what works in the design of agent-based microworlds? *Journal of Educational Psychology, 95*(4), 806–812.
- McEldoon, K. L., Durkin, K. L., & Rittle-Johnson, B. (2013). Is self-explanation worth the time? A comparison to additional practice. *British Journal of Educational Psychology, 83*(4), 615–632.
- *McNamara, D. S. (2004). SERT: self-explanation reading training. *Discourse Processes, 38*(1), 1–30.
- *Molesworth, B. R. C., Bennett, L., & Kehoe, E. J. (2011). Promoting learning, memory, and transfer in a time-constrained, high hazard environment. *Accident Analysis & Prevention, 43*(3), 932–938.
- *Moreno, K. K., Bhattacharjee, S., & Brandon, D. M. (2007). The effectiveness of alternative training techniques on analytical procedures performance. *Contemporary Accounting Research, 24*(3), 983–1014.
- *Nathan, M.J. Mertz, K., & Ryan, B. (1993). *Learning through self-explanation of mathematical examples: Effects of cognitive load*. Paper presented at the 1994 Annual Meeting of the American Educational Research Association.
- Nesbit, J. C., & Winne, P. H. (2003). Self-regulated inquiry with networked resources. *Canadian Journal of Learning and Technology, 29*, 71–92.
- Odilinye, L., Popowich, F., Zhang, E., Nesbit, J., & Winne, P. H. (2015). *Aligning automatically generated questions to instructor goals and learner behaviour*. Paper presented at the Semantic Computing (ICSC), 2015 I.E. International Conference on Semantic Computing.
- *O'Reilly, T., Symons, S., & MacLachy-Gaudet, H. (1998). A comparison of self-explanation and elaborative interrogation. *Contemporary Educational Psychology, 23*(4), 434–445.
- Orwin, R. G. (1983). A fail-safe N for effect size. *Journal of Educational Statistics, 8*, 147–159.
- *Pillow, B. H., Mash, C., Aloian, S., & Hill, V. (2002). Facilitating children's understanding of misinterpretation: explanatory efforts and improvements in perspective taking. *The Journal of Genetic Psychology, 163*(2), 133–148.
- *Pine, K. J., & Messer, D. J. (2000). The effect of explaining another's actions on children's implicit theories of balance. *Cognition and Instruction, 18*(1), 35–51.
- Raffaelli, M., Crockett, L. J., & Shen, Y. L. (2005). Developmental stability and change in self-regulation from childhood to adolescence. *The Journal of Genetic Psychology, 166*(1), 54–76.
- Renkl, A. (1997). Learning from worked-out examples: a study on individual differences. *Cognitive Science, 21*(1), 1–29.
- Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction, 12*(5), 529–556.
- *Rittle-Johnson, B. (2004). Promoting flexible problem solving: the effects of direct instruction and self-explaining. *Proceedings of the Cognitive Science Society, 26*(26).
- *Rittle-Johnson, B. (2006). Promoting transfer: effects of self-explanation and direct instruction. *Child Development, 77*(1), 1–15.
- Rittle-Johnson, B., & Loehr, A. (2017). Eliciting explanations: constraints on when self-explanation aids learning. *Psychonomic Bulletin & Review, 24*(5), 1501–1510.
- Rittle-Johnson, B., Loehr, A., & Durkin, M. (2017). Promoting self-explanation to improve mathematics learning: a meta-analysis and instructional design principles. *ZDM, 49*(4), 599–611.
- Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*(2), 183–192.
- Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: major trends and implications for education. *Mind, Brain, and Education, 2*(3), 114–121.
- Schroeder, N. L., & Gotch, C. M. (2015). Persisting issues in pedagogical agent research. *Journal of Educational Computing Research, 53*(2), 183–204 <https://doi-org.proxy.lib.sfu.ca/10.1177/0735633115597625>.
- *Schwonke, R., Ertelt, A., Otieno, C., Renkl, A., Alevin, V., & Salden, R. J. C. M. (2013). Metacognitive support promotes an effective use of instructional resources in intelligent tutoring. *Learning and Instruction, 23*, 136–150.
- *Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: when instructional explanations reduce self-explanations. *Computers and Education, 46*(4), 426–445.
- *Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology, 99*(2), 285–296.
- Siegler, R. (1995). How does change occur: a microgenetic study of number conservation. *Cognitive Psychology, 25*, 225–273.

- Siegler, R. (2002). Microgenetic studies of self-explanation. In N. Garnott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31–58). Cambridge: Cambridge University Press.
- *Siegler, R., & Chen, Z. (2008). Differentiation and integration: guiding principles for analyzing cognitive change. *Developmental Science*, *11*(4), 433–448.
- Smith, S. W., Daunic, A. P., & Taylor, G. G. (2007). Treatment fidelity in applied educational research: expanding the adoption and application of measures to ensure evidence-based practice. *Education and Treatment of Children*, *30*(4), 121–134.
- *Tajika, H., Nakatsu, N., Nozaki, H., Neumann, E., & Marumo, S. (2007). Effects of self-explanation as a metacognitive strategy for solving mathematical word programs. *Japanese Psychological Research*, *49*(3), 222–233.
- *Tenenbaum, H. R., Alfieri, L., Brooks, P. J., & Dunne, G. (2008). The effects of explanatory conversations on children's emotion understanding. *British Journal of Developmental Psychology*, *26*(2), 249–263.
- *Vandervelde, C. K. (2009). *The importance of causal antecedent emotional state inferences to narrative reading comprehension*. (3374081 Ph.D.), The University of Iowa, Ann Arbor.
- VanLehn, K., Jones, R. M., & Chi, M. T. (1992). A model of the self-explanation effect. *The Journal of the Learning Sciences*, *2*(1), 1–59.
- Vermunt, J. D. (1996). Metacognitive, cognitive and affective aspects of learning styles and strategies: a phenomenographic analysis. *Higher Education*, *31*(1), 25–50.
- Vermunt, J. D., & Donche, V. (2017). A learning patterns perspective on student learning in higher education: state of the art and moving forward. *Educational Psychology Review*, *29*(2), 269–299.
- *Weerasinge, A., & Mitrovic, A. (2005). Supporting deep learning in an open-ended domain. In M. Gh. Negoita & B. Reusch (Eds.), *Real World Applications of Computational Intelligence, studies in fuzziness and soft computing* (Vol. 179, pp. 105–152). Berlin: Springer.
- Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., Dolan, R., & Blakemore, S. J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition*, *22*(1), 264–271.
- *Wichmann, A. (2010). Multi-level support with respect to inquiry, explanations and regulation during an inquiry cycle. Dissertation. Universität Duisburg-Essen, Fachbereich Bildungswissenschaften, Essen.
- *Wichmann, A., & Leutner, D. (2009). Inquiry learning multilevel support with respect to inquiry, explanations and regulation during an inquiry cycle. *Zeitschrift Für Pädagogische Psychologie*, *23*(2), 117–127.
- *Williams, J. J., & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology*, *66*(1), 55–84.
- Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review*, *22*(4), 393–409.
- Wong, R. M. F., Lawson, M. J., & Keeves, J. (2002). The effects of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction*, *12*(2), 233–262.
- *Wylie, R., Sheng, M., Mitamura, T., & Koedinger, K. R. (2011). Effects of adaptive prompted self-explanation on robust learning of second language grammar. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Proceedings of the 15th International Conference on Artificial Intelligence in Education (AIED'11)* (pp. 588–590). Berlin: Springer.
- *Yeh, Y. F., Chen, M. C., Hung, P. H., & Hwang, G. J. (2010). Optimal self-explanation prompt design in dynamic multi-representational learning environments. *Computers & Education*, *54*(4), 1089–1100.
- *Yerushalmi, E., Cohen, E., Mason, A., & Singh, C. (2012). What do students do when asked to diagnose their mistakes? Does it help them? I. An atypical quiz context. *Physical Review Special Topics-Physics Education Research*, *8*(2), 020109–1–020109-19.
- *Yuasa, M. (1993). *The effects of active learning exercises on the acquisition of SQL query writing procedures* (unpublished doctoral dissertation). Georgia Institute of Technology, USA.