Inducing self-explanation: A meta-analysis and experiment

by

Kiran Bisra

M.A., Simon Fraser University, 2010

B.Sc., University of British Columbia, 2007

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Educational Psychology Program
Faculty of Education

SIMON FRASER UNIVERSITY

Fall 2020

Declaration of Committee

Name: **Kiran Bisra**

Degree: **Doctor of Philosophy**

Thesis Title: **Inducing self-explanation: A meta-analysis and experiment**

Committee:

Chair: **Lucy Le Mare**
Professor, Education

John Nesbit
Supervisor
Professor, Education

Philip Winne
Committee Member
Professor, Education

David Kaufman
Examiner
Professor, Education

Patricia Ann Alexander
External Examiner
Professor
Human Development and Quantitative Methodology,
College of Education
University of Maryland

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Abstract

Self-explanation is a process by which learners generate inferences about causal connections or conceptual associations. This dissertation seeks to contribute to the literature on inducing self-explanations, by way of prompting, to facilitate learning. More specifically, this research seeks to understand the effects on learning gains when learners are prompted to self-explain in various contexts and with various prompts. As such, one goal of the dissertation is to provide a comprehensive review of prior research on self-explanation.

A meta-analysis was conducted on research that investigated learning outcomes of 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 (SEPs) are a potentially powerful intervention across a range of instructional conditions.

To further investigate the effect of various prompts on studying expository text, I conducted an online experiment employing a 2 x 2 x 3 factorial design, in which one factor was within subject. One hundred and twenty-six participants were randomly assigned to one of three self-explanation prompt conditions (content-free (generic), content-specific (specific), and no SEP). The results support the utilization of generic self-explanation prompts in comparison to specific self-explanation prompts and receiving no prompt. Specifically, the generic self-explanation group outperformed the other two groups on the reading comprehension outcome in the short-answer question format.

Keywords: Self-explanation; Instructional explanation; Meta-analysis; Prompts; Online experiment

Dedication

I dedicated this dissertation to:

My mom, who has always:

- supported all of my academic endeavours, even when they didn't made sense to her;
- pushed me to be more, even when she didn't know what more could be; and
- showed me what it meant to be of generous spirit, even when life becomes difficult.

Acknowledgements

The challenging aspects of completing this dissertation were not the countless hours staring at free recall responses, or advertising to potential participants, or even self-explaining self-explanation. It was trying to reach the goal that I had set over a decade ago. This goal was to conduct research that attempted to use innovative research and analysis techniques, while maintaining school-work-life balance. I would like to express my sincerest gratitude to the people who have inspired, supported, and guided me to my goal.

First, I would like to thank my senior supervisor Dr. John Nesbit, who has been my supervisor for the better part of my adult life. I have greatly benefited from his mentorship, high standards of academic excellence, and confidence in my ability to “figure it out”. I have gone through several life changes and stages during the past decade, and John has been provided me with sound advice, perspective, and at times, much needed encouragement. I have grown as a researcher, a writer, and a thinker under his guidance, and will forever use his supportive mentorship style as a model for when I have opportunities to mentor others.

I would also like to thank Dr. Phillip Winne for being an inspiration for originality, innovation and creativity. His critical evaluation of my methods and analysis greatly improved quality and added depth. Phil changed how I conceptualize statistical analysis and this shift has had a tremendously beneficial influence on my professional work.

To my husband, Hardeep and dearest daughter Sahara, thank you for enduring the times when I was away so I could complete this dissertation. Hardeep, even when I was not sure I should continue with this journey, you never let me give up because deep down you knew it was what I wanted. I thought that once I had a baby, I would become too busy to complete this dissertation. However, Sahara you inspired me to be a better role model and reinvigorated my energy.

To my mom, thank you for not only encouraging me to complete what I started, but also for the countless hours of babysitting. I would like to also thank my dad who has asked me, “Are you done yet?” each time I saw him during the past few years. In its own way, I found his interest in my work motivating. And, I would like to thank my sister who has come through for me in ways I could never have imagined.

Finally, I'd like to offer a heartfelt thanks to my best friend – Lisa. I am genuinely impressed by your ability to seamlessly switch from one supporting role to another - from editor, to life coach, to therapist, to enabler, and much more. You are my confidence, my resolve, and my persistence in the pursuit of my life goals.

Table of Contents

List of Tables

List of Figures

List of Acronyms

CBI	Computer-based instruction
ITS	Intelligent tutoring systems
SE	Self-explanation
SEP	Self-explanation prompt
WE	Worked-out example

Chapter 1. Overview

Self-explanation (SE) is the activity of explaining to oneself new information presented in text, an animation, diagram, or another medium (Chi, 2000; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Roy & Chi, 2005). It is a learner-generated explanation and is different from explanation provided by an external source (i.e., an instructor or a textbook). It may occur spontaneously (Chi et al., 1989; Pirolli & Recker, 1994) or in response to instruction (Bielaczyc, Pirolli, & Brown, 1995). Engaging in self-explanation allows the learner to integrate current learning with prior knowledge (Chi, 2000; Chi, De Leeuw, Chiu, & LaVancher, 1994).

The literature on self-explanation has shown effects across domains and learning tasks (Dunlosky et al., 2013). Self-explanation has been used to promote learning in biology (Chi et al., 1994), second-language grammar (Wylie, Koedinger, & Mitamura, 2009), engineering (Chung, Severance, & Moon-Jung, 2003), geometry (Aleven & Koedinger, 2002), statistics (Renkl, Stark, Gruber, & Mandl, 1998) and in many other disciplinary contexts (Chi, 2000).

Without guidance or prompting, many learners do not engage in self-explanation (Renkl, 1997). Self-explanation prompts (SEPs) can be provided by teacher (Chi, De Leeuw, et al., 1994), generated by a computer-based tutor (Aleven & Koedinger, 2002; McNamara, 2004) or embedded into the learning materials studied by the learner (Renkl et al., 1998). As well, metacognitive SEPs have been shown to encourage learners to engage in metacognitive monitoring and control activities (Nokes, Hausmann, VanLehn, & Gershman, 2011; Pintrich, 2000; Winne, 2000; Winne & Hadwin, 1998; Zimmerman, 2000). Unlike prompts used in other learning strategies (e.g., elaborative-interrogation), SEPs have been quite varied across studies (Dunlosky et al., 2013).

Although one of the strengths of self-explanation research is the range of learning conditions, domains, learning outcomes, and prompt types it covers; this extensive range makes it difficult to reduce the findings into a general summative statement about the efficacy and underlying mechanism of self-explanation (Dunlosky et al., 2013). The seemingly broad applicability of self-explanation can be misleading and create issues for implementation. An instructional designer or instructor still must make decisions about how to induce the self-explanation (e.g., the exact wording of the SEP, placement within a learning activity, number of SE inducements).

This dissertation contributes to the literature on self-explanation by quantifying and integrating results from self-explanation research to provide some implementation guidance and identify gaps in the literature. As such, one goal of this work is to comprehensively review prior research and produce a meta-analysis. Unlike narrative or classic systematic literature reviews, meta-analyses are research reviews with the same requirements as empirical studies: replicability, objectivity, and systematization (Sánchez-Meca & Marín-Martínez, 2010) The meta-analysis focuses on cognitive effects of learning while self-explaining under a variety of conditions and investigates how a range of conditions, or moderating variables, impact learning outcomes. From this investigation, implications for theorizing about and applying self-explanations emerged.

The meta-analysis supports the notion that the self-explanation effect is dependent on the type of SEP used (Dunlosky et. al., 2013). One key aspect in which SEPs differ is the extent to which the SEP contains content specific to the learning material. Reviewing gaps in the literature revealed that the effects of content-specific and content-free SEPs had not been compared. Therefore, a second goal of this dissertation is to investigate how content embedded into the SEP impacts reading comprehension and recall. The results from the investigation inform the theoretical mechanism underlying self-explanation.

This chapter will (a) develop the research questions of this dissertation within the context of the research literature on self-explanation; and (b) summarize the methods, results, and contributions of the dissertation.

1.1. Theoretical Framework & Research Findings

This section presents the theoretical framework and research findings of the cognitive effects of self-explanation. Cognition and reasoning are described in terms of mental models, and self-explanation is explained as a process by which learners fill gaps and revise their mental models.

1.1.1. Mental Models & Self-explanation

Mental models are internal or mental representations used in thinking (Vosniadou & Brewer, 1992). They are represented in memory as interconnected pieces of information organized to correspond in some way to an external structure, such as the solar system (Gentner & Stevens, 1983; Johnson-Laird, 1983). Through experience, people create internal mental models of situations that help them make predictions or explain phenomena. Mental models are akin to mental animated simulations used to generate predictions under various conditions (Taylor, Pham, Rivkin, & Armor, 1998). From this perspective the difference between a domain-expert and a domain-novice is that a domain-expert's mental model is more elaborate and contains more connections in comparison to a domain-novice's mental model (Sterman & Ford, 1998). Furthermore, mental models may evolve over time, and can, at times, be incorrect.

A student learning a concept often comes to the learning situation with a mental model of the to-be-learned concept constructed from previous schooling or direct experience. This mental model could have be incorrect or contain conflicting prior knowledge (Chi, 2008). As they engage in instructional activities, students with missing, little, or incomplete prior knowledge add new knowledge to enrich their mental model (Carey, 1999). They will essentially compare their mental model with the presented information, and fill knowledge gaps. Students whose prior knowledge conflicts with the to-be-learned concept may undergo conceptual revision and change their mental models (Vosniadou, 2007).

It is hypothesized that self-explanation aids in filling knowledge gaps in mental models (Chi & Bassok, 1989; Lin & Lehman, 1999; VanLehn, Jones, & Chi, 1992) and in detecting conflict detection and revising mental models (Chi et al., 1994; DeLeeuw & Chi, 2003). The following sections provide more detail about the two main cognitive effects of SE: inference generation and conceptual revision (Chi, 2000).

1.1.2. Inference Generation

An inference is a conclusion based on premises, evidence or reasoning (Johnson-Laird, 1995). Inferences are necessary to link up ideas and fill in details that are not explicitly stated (e.g., Garnham & Oakhill, 1987; Graesser, Singer, & Trabasso, 1994; Johnson-Laird, 1983). Self-explanation affords opportunity for learners to notice gaps between their current mental model and new information, and to make inferences to bridge these gaps (Chi, Slotta, & De Leeuw, 1994).

Research findings have shown that readers who self-explain the text either spontaneously or when prompted to do so, understand more from the text and therefore construct better mental models of the content (Chi & Bassok, 1989; Chi et al., 1994; Magliano, Trabasso, & Graesser, 1999). In Chi et al. (1994), one group of participants was asked to read an expository text and self-explain after reading each line of the passage, while a second group was asked to read the expository text twice. The prompted group outperformed the unprompted group, and participants who generated a larger number of self-explanations had a greater understanding than participants who provided a smaller number of self-explanations. It has been hypothesized that self-explanations are inference products that 1) fill-in missing information from the text or 2) fill-in missing information from a learner's mental model.

The incomplete text hypothesis assumes that text is incomplete in some way. One possibility is that the text is poorly written and missing connective links and is therefore difficult to comprehend (Kintsch & Vipond, 1979). Or, the text is explanatorily incoherent (Kintsch & van Dijk, 1978) and missing crucial pieces of information. Reading comprehension research has shown that high-domain knowledge students learn just as well from explanatorily incoherent texts, but low-domain knowledge students learn better from coherent texts where background information is supplied (McNamara, Kintsch, Songer, & Kintsch, 1996). Self-explanation allows a learner to turn incoherent text into coherent text by identifing gaps and ommissions in the text and inducing linkages. High domain learners may naturally be able to fill in these gaps by generating inferences from their own knowledge to fill in the missing pieces of an incoherent text.

The imperfect mental model view assumes that text could be structurally and explanatorily coherent, but still contain omissions from a learner's perspective. All text assumes the reader has the prior knowledge necessary to make sense of what is written. However, each person has a unique prior schema, and may be missing the necessary information needed to understand the text (for example, a definition). The imperfect mental model view assumes the learner has come to the learning situation with different deficient pre-existing mental model. Therefore, if the text is incomplete in some way, the goal of self-explaining is to generate inferences to fill the omissions in one's own mental model. Through prompting, learners are encouraged to deduce the omitted information and gaps in their knowledge (Chi, 2000, p. 198).

In either case, whether the text is incomplete in some way or the learner has an imperfect mental model, self-explanation improves reading comprehension by supporting inference generation.

Experimental studies have demonstrated that embedding SEPs into problem solving improves learning (Chi et al., 1989; Hmelo-Silver, 2004). For the sake of brevity, only the research on the educational impact of embedding SEPs into worked-out examples will be presented.

Worked-out examples (WEs) consist of a problem, solution steps and a final solution. Research has shown that learning from WEs in well-structured domains, such as mathematics or programming, can be very beneficial for the initial acquisition of a skill (VanLehn, 1996). Further, this learning activity can be more effective than problem solving (Sweller & Cooper, 1985; Sweller, Van Merrienboer, & Paas, 1998) or explicit instruction (Zhu & Simon, 1987). Although WEs consist of a series of solution steps, they do not generally include conditions under which a principle or concept is applied. Therefore, WEs inherently contain completeness gaps that invite a learner to generate inferences to add coherence to the example (Catrambone, 1996; Chi et al., 1989). If the student fills the gaps and is able to provide the correct links between the steps and the sub-goals of the problem, then the student is able to encode the problem type and develop a robust problem-solving mental model. SEPs can be deployed to encourage the learner to make these inferences. However, the manner in which a learner self-explains impacts learning effectiveness.

Chi et al. (1989) conducted a study in which they concluded that the extent learners profit from studying WEs depends on how well the learners self-explain the solution of the example. More specifically, successful learners relate domain-specific principles to problem solving operators and more frequently elaborate the conditions and goals of the operator. Subsequent studies have found similar results (Ferguson-Hessler & de Jong, 1990; Pirolli & Recker, 1994; Renkl, 1997). As most learners only superficially explain to themselves (Renkl, 1997), they need to be encouraged to actively self-explain worked-out examples (Renkl, 2005).

1.1.3. Conceptual Revision

Conceptual revision occurs when new information is in conflict with naïve prior knowledge, and the learner takes intentional steps to resolve this dissonance (Chi, 2000). Naïve knowledge is incorrect and often impedes the learner from gaining a deeper conceptual understanding. Some types of naïve knowledge are readily revised and can be improved with instruction (Chi & Roscoe, 2002, p.3), and these are referred to as "preconceptions" (Clement, 1993). However, other types of naïve knowledge are highly resistant to change. They persist despite contradictory instruction and are referred to as "misconceptions" (Eaton, Anderson, & Smith, 1984).

Although the term "conceptual change" could refer to revising or removing any ill-conceived prior knowledge, it is often reserved for correcting or repairing misconceptions (Posner, Strike, Hewson, & Gertzog, 1982). Cognitive processes associated with conceptual change occur at the level of propositions (belief revision), the level of interrelated propositions (mental model revision), or the level of underlying categories (ontological revision) (Chi, 2000, 2008). This dissertation focuses on conceptual change at the mental model level.

Conceptual revision requires metaconceptual awareness of misunderstandings (Vosniadou, Ioannides, Dimitrakopoulou, & Papademetriou, 2001) and metacognitive regulation to set revision of knowledge and reconciliation of contradictions as goals (Chi, 2000; Winne & Hadwin, 1998). However, students are generally unaware of their misconceptions (Roy & Chi, 2005) because their pre-existing schemas do not allow them to recognize when new information contradicts their prior understanding (Vosniadou et al., 2001). This is because new information can only be assimilated into existing schemas when it is compatible with them (Piaget, 1953; Sweller & Chandler, 1994). When it is not compatible, experience and learning are interpreted through an erroneous lens, and prior incorrect knowledge interferes with understanding new information. Research has shown that learners will misinterpret or distort new evidence when it contradicts existing mental models (Eaton, Anderson, & Smith, 1983; Stepans, 1991).

Furthermore, students may be unaware of their misconception because their flawed mental model may be coherent and share a number of prepositions with the correct mental model (Chi & Roscoe, 2002). When a student has an incomplete or fragmented mental model, they often are aware of the gaps in their understanding. However, coherent models are organized and allow students to generate explanations that are meaningful to them (Vosniadou & Brewer, 1992, 1994), even if the explanations are incorrect. When data collected by Chi et al. (1994) was reviewed, it was found that over half of the participants had coherent but flawed mental models of the circulatory system (Chi, 2000).

SEPs have been shown to promote and facilitate conceptual change by urging students to reconcile their existing knowledge with the correct knowledge (Gadgil, Nokes-Malach, & Chi, 2012). For example, metacognitive SEPs (Nokes et al., 2011) encourage learners to identify discrepancies among their current state of learning, the desired state of learning (Butler & Winne, 1995), and the presented material (Nokes et al., 2011). Once these discrepancies are identified, the learner will generate inferences and make repairs to her mental model (Chi, 2000).

One main research finding is that self-explanation encourages conceptual change in scientific knowledge (Chi et al., 1994), a domain in which thousands of studies have documented numerous misconceptions (Chi, 2005). Three decades of research have consistently shown that misconceptions are persistent and robust (Treagust & Duit, 2008) and cannot be corrected easily by explicit instruction (Eaton et al., 1983; Fisher, 1985)

Two teams studied how participants responded to SEPs while studying from text about the circulatory system. Chi et al. (1994) found that approximately two-thirds successfully transformed their flawed model to the correct model. And Gadgil et al. (2012) discovered evidence that participants were more likely to acquire a correct mental model of the circulatory system when asked to self-explain.

1.1.4. SEPs

SEPs can be verbal cues from a person (Chi, De Leeuw, et al., 1994), messages generated by computer tutors (Aleven & Koedinger, 2002; McNamara, 2004) or questions embedded in reading materials (Renkl et al., 1998). Two main considerations when designing a learning activity with SEPs are: 1) when to prompt and 2) type of SEP.

SEPs can be provided to learners at the beginning, concurrently with, or at the end of the learning activity. According the meta-analysis reported in this dissertation, most studies investigating the efficacy of self-explanation provide SEPs concurrently with the learning activity. This includes studies which ask participants to self-explain each sentence or paragraph (Ainsworth & Burcham, 2007; Griffin, Wiley, & Thiede, 2008), or each problem or step in a problem set (Hausmann & VanLehn, 2007; Nathan, 1994).

One issue with providing SEPs during a learning activity is that self-explaining may require substantive cognitive resources and impose a heavy cognitive load on a learner (Chandler & Sweller, 1991; Mayer, 1997; Mayer, Heiser, & Lonn, 2001). For example, learners with low prior knowledge may have to split their cognitive resources between the learning task and providing SEs, which may interfere with learning. A second issue with injecting SEPs into a learning activity is that requiring participants to self-explain each sentence or each problem can be tedious and time-consuming.

SEPs can be designed to induce particular cognitive activities in a learner. For example, prompts have asked the learner to justify the correct step (Berthold, Eysink, & Renkl, 2009; Rittle-Johnson, 2006), engage in reflection (Chi et al., 1994), and reference terms in a glossary (Aleven & Koedinger, 2002). Furthermore, prompts can be specific to the domain content, or more general in nature. Research has shown that less able students learn better with specific prompts and more able students learn better with generic prompts in well-structured domains (Kramarski, Weiss, & Sharon, 2013).

1.2 Self-explanation Meta-analysis

One issue that challenges all research fields is how to synthesize and make inferences about a phenomenon when it has been studied in multiple contexts across numerous projects. A meta-analysis is a research approach or mechanism that applies statistical techniques to examine, standardize, and combine the results of different empirical studies in which the treatment is considered consistent by investigators across the set (Borenstein, Hedges, Higgins, & Rothstein, 2009; Hedges & Olkin, 1985; Wasserman, Hedges, & Olkin, 1988).

Although SEPs and their effects on learning have been documented for years (Chi, 2000; Roy & Chi, 2005), a meta-analysis on the subject has not been performed. Researchers, teachers, and students have limited empirical guidance as to which conditions will boost or inhibit learning while self-explaining. Synthesis of the research may provide valuable insight into the types of cognitive activities, prompts, and learner characteristics best situated to using SEPs for teaching and learning. I led a research project that completed a meta-analysis to answer several research questions about SEPs (Bisra, Liu, Nesbit, Salimi, & Winne, 2018).

The meta-analysis, reported in Chapter 2, assessed the cognitive effects and instructional effectiveness of self-explanation interventions. Experimental and quasi-experimental research, in which at least one group was induced to self-explain, was meta-analyzed. A set of primary studies was identified in which the participants were induced to self-explain in response to prompts during a learning task or were directed to self-explain prior to the learning task. In the meta-analysis, results are reported about how learning outcomes varied under a range of conditions and implications for theorizing about and applying self-explanation.

After an extensive search, 69 studies (5,917 participants) met a specified inclusion criterion see (p. 32), and the extracted effect sizes produced an overall point estimate of g = .55. To put this into context, self-explanation offers benefits similar in magnitude to those for interventions such as mastery learning (d = .50) and peer tutoring (d = .55) (Hattie, 2009). Furthermore, the meta-analysis showed a statistically detectable advantage (g = .35) of self-explanation over instructional explanations. This result supports the hypothesis that self-explanation engages learners to apply cognitive processes that generate inferences and/or aid in linking prior knowledge with the new knowledge generated by self-explaining.

The learning and teaching community may use these results to inform pedagogical practise in the classroom and study strategies outside of the classroom. The results from this meta-analysis may also help researchers identify gaps in the self-explanation literature and spur fresh lines of inquiry into the topic.

1.3 Experiment with SEPs

This dissertation also extends previous research by examining the effect of inducing self-explanations using prompts that contain content related to the text passage. SEPs have been quite varied in the literature (Dunlosky et al., 2013) and, although prior work supports the efficacy of self-explanation in different contexts (Bisra et al., 2018; Chi, 2000), the effect of using one type of SEP rather than another is unclear.

The format of elicited inducement and timing of the prompt are just a few ways in which SEPs can differ from one another. For example, learners have been provided with SEPs in a myriad of formats including as a fill-in-the-blank task (Berthold et al., 2009), a multiple-choice question (Booth, Lange, Koedinger, & Newton, 2013; Johnson & Mayer, 2010), or an open-ended question (Graf, 2000; Kwon & Jonassen, 2011; Schworm & Renkl, 2007).

Instructional designers must make decisions about the type of SEP to use because students may find multiple SEP types in the same activity cumbersome and even confusing. As well, one type of SEP may be better suited for a particular learner characteristic. Therefore, investigating the impact of different types of SEPs can not only lead to prescriptive recommendations, but can also shed light on the cognitive processes that are engaged when self-explaining.

One issue that arises when designing and embedding a SEP is whether the prompt should make specific reference to the content of the material. More content-specific language may guide the participant's attention and direct self-explanation. However, allowing learners to apply their own criteria to decide when and what to self-explain may support constructing explanations that better complement prior knowledge.

The effect of receiving content-free (generic) and content-specific (specific) prompts to self-explain during a reading task was investigated. One hundred and twenty-six participants were randomly assigned in a three-way factorial design experiment. Participant recruitment and experimentation occurred online, and data were collected over the course of three sessions. Participants completed a reading task during both Session 1 and 3 which required reading a short passage, completing a free recall task about the passage, and answering short-answer and multiple-choice questions about the passage. The test scores collected in Session 1 and 3 constituted a within-subject (repeated measures) factor.

There were two reading tasks (A and B), and one of the between-subject factors was the order of the reading task. As such, about half of the participants received reading task A during Session 1, and then reading task B during Session 3, while the other half received the reading tasks in the order of BA.

The second between-subject factor for this experiment was the SEP treatment participants received during Session 3: no SEP, generic SEP, or specific SEP. Groups who received SEPs received self-explanation training in Session 2. Meanwhile, the no SEP group spent Session 2 watching a video unrelated to any of the content in reading task A and B.

The results support the hypothesis that generic SEPs are superior to specific SEPs and receiving no SEPs on the reading comprehension outcome in the short-answer question format. However, no differences in scores were detected on the free recall measure.

1.4 Overview of the Chapters

Chapter 2 reports on the self-explanation meta-analysis and presents the corresponding research questions, inclusion criteria, methods, and results.

Chapter 3 reports the online experiment that compared the effects of content-free (generic) and content-specific (specific) SEPs while reading text. The chapter presents the context for the experiment and provides a detailed account of the study procedures, learning materials, methods, results, and a discussion of the implications of the results.

Chapter 4 presents a general discussion of the results of the two prior chapters in the context of the self-explanation literature. The chapter concludes the dissertation with theoretical and practical implications of the two studies and discusses future avenues of research on how to advance the field and address the limitations of my work.

In conclusion, my dissertation generates new understandings and knowledge about self-explanation as a study strategy and has direct implications for the education community.

Chapter 2. Inducing Self-explanation: A Meta-Analysis

This Chapter presents a meta-analysis about the cognitive effects and instructional effectiveness of self-explanation interventions. Significant portions of this chapter were adapted from an article published in *Educational Psychology Review* (Bisra et al., 2018). It is comprised of five sections. The first two sections present the purpose and the research questions addressed by the work. Next, the methodological approach is described, including inclusion criterion for research studies. Lastly, the meta-analysis results are presented and discussed relative to this dissertation.

2.1 Purpose of the Meta-Analysis

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 & 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 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 use spontaneously 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 & 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 post-test than learners prompted to generate notes.

Although several reviews of relevant self-explanation studies support its use as a study skill (Chi & Wylie, 2014; Dunlosky et al., 2013; Rittle-Johnson, Loehr, & Durkin, 2017), we could find no published report that comprehensively synthesized research investigating the relationship between self-explanation and learning outcomes. Wittwer & 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.

2.1.1 Self-Explanation versus Instructional Explanation

One can speculate self-explanation is effective because it has 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, Tabbers, Rikers and Pass (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, Schieter and Catrambone (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, Kumalasari, and Howland (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.

2.1.2 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 particular text or solving 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, de Leeuw, Chiu, and LaVancher (1994) in which 8^{th} 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 one hour 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, Cepeda, Rohrer, Kang, & Pashler, 2012).

In research examining effects of self-explanation prompts on solving mathematics problem by elementary school students, McEldoon, Durkin and Rittle-Johnson (2012) 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 post-test, 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.

2.1.3 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, Adesope, Nesbit, & Liu, 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 that 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 & Koedinger, 2012; Chu, Rittle-Johnson, & Fyfe, 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 & Atkinson, 2007; Schroeder & Gotch, 2015).

During post-tests, 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 & Lerner, 2008; Kopp, 1982; Raffaelli, Crockett, & Shen, 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, Driessen, Chan, Cees, & Van Der Vleuten, 2012; Marambe, Vermunt, & Boshuizen, 2012). Differences have been reported between geographic regions in students' learning patterns (Marambe et al., 2012) and metacognition (Hartman, 2001; Hartman, Everson, Toblas, & Gourgey, 1996), and these have the potential to moderate self-explanation effects.

Finally, we examined treatment fidelity (Smith, Daunic, & Taylor, 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.

2.2 Research Questions

The meta-analysis addressed the following research questions:

1) What are the effects of learning with SEPs in comparison with no additional explanation, instructional explanation, or another strategy/technique?

2) How are the effects of learning with various self-explanation treatments moderated by timing, prompt type, content specificity, and types of elicited self-explanations?

3) How are the effects of learning with SEPs moderated by learning task and environment?

4) How are the effects of learning while self-explaining moderated by the goal materials (i.e. subject domains, knowledge types, images) and assessment type?

5) How are these effect sizes influenced by methodological features of the research?

6) How do these effects vary while self-explaining for different education levels and regions?

2.3 Method

For this meta-analysis, we followed meta-analytic methods described by Borenstein, Hedges, Higgins and Rothstein (2009), Lipsey and Wilson (2001) and Hedges and Olkin (1985). The major phases were: (1) identifying all relevant studies, (2) coding the study characteristics, and (3) calculating effect sizes across study characteristics.

2.3.1 Selection Criteria

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 1,306 studies. Reference sections of review articles (e.g., Atkinson, Derry, Renkl, & Wortham, 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, Renkl, & Merrill, 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, Chung, & Severance, 1999; Conati & 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 as a means of eliciting self-explanations, especially when the 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 think-aloud directions can improve performance on some types of problems (e.g. Fox & 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, Lawson, & Keeves, 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.

2.3.2 Coding of Study Characteristics

The coding form included 43 fixed-choice items and 24 brief-comment items that cataloged information about each study in eight 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 us independently coded 10 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 put forth and resolved. As well, an online forum was used to facilitate discussion during the weeks in which these weekly meetings did not occur. Meetings and the online communications served four purposes: 1) to resolve emerging issues, 2) to sharpen category definitions, 3) to disseminate definitional refinements to all coders and 4) to create 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.

2.3.3 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 & Wilson, 2001). To this end, when a study reported more than a single comparison treatment (i.e., treatments in which 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, Hedges, Higgins, & Rothstein, 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.

2.4 Results

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

Table 1: Overall Effect and Weighted Mean Effect Sizes for Comparison Treatments

			Effect Size		95% CI		Test of Heterogeneity		
	N	*k*	*g*	*SE*	*Lower*	*Upper*	Q_B	*df*	*p*
Overall: Random-Effects Model	5917	69	.552*	.050	.454	.650	196.630	68	< .001
			Effect Size		**95% CI**		**Test of Heterogeneity**		
	N	k	g	SE	Lower	Upper	QB	Df	p
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

2.4.2 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 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.

Table 2: Weighted Mean Effect Sizes for Self-Explanation Inducements

			Effect Size		95% CI		Test of Heterogeneity		
	N	*k*	*g*	*SE*	Lower	Upper	Q_B	*df*	*p*
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							12.69	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									
Justify	1700	20	.416*	.079	.261	.570	27.019	7	< .001
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

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: Weighted Mean Effect Sizes for Learning Task Duration

			Effect Size		95% CI		Test of Heterogeneity		
	N	*k*	*g*	*SE*	*Lower*	*Upper*	Q_B	*df*	*p*
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 mins	1842	17	.709*	.107	.500	.919			
30-60 mins	880	13	.601*	.075	.455	.747			
1-2 hrs	341	6	.444*	.143	.164	.724			
> 2 hrs	436	6	.462*	.150	.168	.756			
Not reported	2418	27	.475*	.086	.308	.643			

**p* < .05

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 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 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.

Table 4: Weighted Mean Effect Sizes for Learning Task

			Effect Size		95% CI		Test of Heterogeneity		
	N	*k*	*g*	*SE*	*Lower*	*Upper*	Q_B	*df*	*p*
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	0.745			
Procedural	204	3	.468*	.167	.142	.795			
Both	3173	33	.516*	.072	.374	.658			
Subject							3.855	5	.571
Science	1158	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*	0.18	.407	1.114			
Mixed	93	1	.313	.221	-.121	.746			
Other	1203	12	.591*	.139	.319	.863			

*p < .05

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 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.

Table 5: Weighted Mean Effect Sizes for Learning Environment Characteristics

			Effect Size		95% CI		Test of Heterogeneity		
	N	*k*	*g*	*SE*	*Lower*	*Upper*	Q_B	*df*	*p*
Media Type							2.677	5	.503
Digital	4255	42	.550*	.062	.428	.672			
Print	1223	21	.550*	0.11	0.34	.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*	0.06	0.46	0.69			
Diagram in Materials							.385	2	.825
Yes	1435	23	.588*	0.12	0.387	0.788			
No	4291	43	.532*	.059	.416	.648			
Not reported	185	3	.661*	.283	.107	1.215			
Visual Pedagogical Agent							11.31	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			

*$p < .05$

Guided by the theory of self-explanation as explained in the introduction, we hypothesized 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 post-test questions as requiring “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 post-test 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.

Table 6: Weighted Mean Effect Sizes for Learning Outcome Assessment

			Effect Size		95% CI		Test of Heterogeneity		
	N	*k*	*g*	*SE*	*Lower*	*Upper*	Q_B	*df*	*p*
Learning Outcome							6.072	6	.415
Comprehension	187	3	.282	.151	-.014	.579			
Inference	78	2	1.794*	.854	.120	3.469			
Knowledge	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 Type							5.156	6	.524
Essay	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			
Problems 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	.512	-.225	1.249			

*$p < .05$

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, Heine, & Norenzayan, 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 were conducted in Europe. A comparison among regions returned *p* = .011. Post-hoc tests indicated studies conducted in East Asia had effect sizes 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).

Table 7: Weighted Mean Effect Sizes for Grade Level and Region

			Effect Size		95% CI		Test of Heterogeneity		
	N	*k*	*g*	*SE*	*Lower*	*Upper*	Q_B	*df*	*p*
Education Level							3.051	3	.384
Elementary	849	10	.483*	.104	.276	.687			
High School	1415	13	.426*	.091	.247	.606			
Undergraduate	3349	42	.611*	.071	.472	.751			
Professional Program	304	4	.680*	0.291	.110	1.250			
Region							13.135	4	0.01
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

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.

Table 8: Weighted Mean Effect for Methodology

			Effect Size		95% CI		Test of Heterogeneity		
	N	*k*	*g*	*SE*	*Lower*	*Upper*	Q_B	*df*	*p*
Treatment Fidelity							4.735	3	.192
Yes, analyzed	2014	31	.613*	.085	.447	.779			
No	2869	21	.409*	.076	.259	.558			
Yes, followed-up	682	11	.587*	.089	.411	.762			
Not reported	353	6	.732*	.221	.299	1.165			

*$p < .05$

2.4.3 Publication Bias

Two statistical tests were computed with Comprehensive MetaAnalysis 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 α = .05). This test revealed 6,028 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 ≥ 5*k* + 10 (Rosenthal, 1995), where *k* is the number of studies included 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.

2.5 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 other effective 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. This section includes a focus on the impact of region on self-explanation and automatic generation of SEPs. For a more general discussion on this meta-analysis, please refer to Chapter 4.

2.5.1 Why Did Studies from East Asia Return High Effect Sizes?

Looking more closely at the six studies conducted in East Asia, we found 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 & Van Mil, 2008; Marambe et al., 2012; Vermunt, 1996; Vermunt & 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.

2.5.2 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 used 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 & Winne, 2003). In such settings, computer-generation of content-specific prompts may be a more 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, Popowich, Nesbit, & Winne, 2013; Odilinye, Popowich, Zhang, Nesbit, & Winne, 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 for which 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.

Chapter 3. An Experiment on SEPs

This chapter reports an experiment that investigated the learning effects of two types of SEPs. First, an overview of research on learning with SEPs is presented. After which, the research questions and related hypotheses based on previous research on self-explanation are included. The chapter then focuses on the learning task, the self-explanation training, the dependent variables, research design, procedures through which data were collected and analyzed. Finally, the results of the study are presented and discussed in relation to previous empirical studies on the impact of SEPs.

3.1 Learning with Self-explanation

Self-explanation is the activity of explaining to oneself new information presented in text, animation, a diagram or another medium (Chi, 2000; Chi et al., 1989; Roy & Chi, 2005). It is a learner-generated explanation as opposed to explanation received from an external source (i.e., an instructor or a textbook), and it may occur spontaneously (Chi et al., 1989; Pirolli & Recker, 1994) or in response to instruction (Bielaczyc et al., 1995). Engaging in self-explanation allows the learner to integrate current learning with prior knowledge (Chi, 2000; Chi, De Leeuw, et al., 1994). Research has shown that when encouraged to self-explain, learners perform better on problem-solving tasks, repair flawed mental models, and generate inferences (Chi, 2000; Chi, De Leeuw, et al., 1994).

A search on PsycINFO database conducted in October 2019 showed that over 450 peer-reviewed articles have discussed the use of self-explanation for various purposes. In the meta-analysis by Bisra, Liu, Nesbit, Salimi and Winne (2018), we found that, under different prompt conditions, settings, and experimental features, the use of self-explanation when compared with no strategy use yielded an overall effect size g = .55.

3.1.1 Does it Matter How Self-Explanation is Elicited?

SEPs can be used strategically trigger a learner’s awareness of a prior mental model and highlight discrepancies between current understanding and the presented information. Although self-explanation can be induced in various ways, it is unclear which type of SEP is most useful to a learner in a particular context. For example, some researchers induced self-explanation by directions given before a learning activity (e.g., Ainsworth & Burcham, 2007). Others provided SEPs during the learning activity (Hausmann & VanLehn, 2010b) or after it (Tenenbaum, Alfieri, Brooks, & Dunne, 2008). Supplying prompts during a learning activity may help learners engage in more effective self-explanation as it signals key content in need of explanation. On the other hand, allowing learners to apply their own criteria to decide when and what to self-explain may support constructing explanations that are better adapted to prior knowledge.

The format used for SEPs may affect features of learners' elaborative processing of content. Some researchers asked learners to give open-ended responses while others used fill-in-the-blank or multiple-choice questions. As with other prompt characteristics, the prompt format may contribute to the prompt's specificity which, in turn, may affect degree and qualities of cognitive elaboration in the self-explanation. An open-ended prompt format offers fewer cues than multiple-choice and fill-in-the-blank formats and might lead to elaborative processing that is better adapted to unique gaps in knowledge of the individual learner. However, in some contexts stronger cues provided by a multiple-choice question may signal more clearly the possibility that the learner holds a misconception that should be confronted.

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 an identified concept or element of the content, (c) explain a prediction, or (d) make a metacognitive judgment about the quality of their understanding, reasoning or explanation. When combined with the task context, SEPs may imply that the learner is to give a description of either causal relationships, conceptual relationships or evidence supporting a claim; but these distinctions are rarely explicit in the prompt and are difficult for a meta-analyst to code reliably.

Dunlosky et al. (2013) examined the self-explanation research and noted that the effects of this strategy demand nontrivial time. The time demands associated with self-explanation may be conflated by increased time on task. As an example, one study prompted learners to self-explain after reading each sentence of a text, which led to the self-explanation group spending double the time reading than the no prompt group (Chi, De Leeuw, et al., 1994). As such, investigating efficient SEPs, ones which provide "bang for the buck" (Dunlosky et al., 2013) is an important line of inquiry. Students may find multiple SEP types a bit cumbersome, confusing, and time consuming when they are in a real-life studying situation.

3.1.2 Content-specific (specific) and Content-free (generic) SEPs

Several studies have used unvarying prompts that 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, Rikers, & Schmidt, 2007). In contrast, most research has given content-specific SEPs 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 & Jonassen, 2012). Providing more specific prompts might draw learners’ attention to difficult conceptual issues that would otherwise be overlooked, but this may also hinder learners’ search for contradictions between the content and the particular understanding each of them brings to the topic.

Bisra et al. (2018) in their review identified only 6 papers that included both content-specific (specific) and content-free (generic) SEPs. In four of these studies, participants were given both types of prompts which makes it difficult to ascertain the effect of receiving generic vs. specific SEPs. Two studies have investigated the effects of generic vs. specific SEPs; results were different.

Wichmann (2010) studied the effect of participants receiving generic SEPs, generic and specific SEPs with hints, and no SEPs when learning about photosynthesis. Students who received both types of prompts as well as hints did better than both other groups on declarative multiple-choice knowledge tests. The time-on-task did not vary across treatment groups in this investigation. In a similar study, participants solving probability questions provided generic and, what the researcher termed, assistance-giving self explanation prompts in fill-in-the-blank format outperformed participants not provided with a prompt or provided only generic prompts (Berthold et al., 2009). More specifically, they outperformed them on a conceptual knowledge task. Time-on-task differences were not reported. In other studies, participants provided with both types of prompts outperformed participants who did not receive a SEP (Berthold & Renkl, 2009; Wong et al., 2002). In the study by Wong et al. (2002), SEPs did increase the length of time students spent studying.

Bodvarsson (2005) had participants read economics case studies and answer conceptual questions grounded in the theories and principles discussed in class or in the textbook. A third of the students were provided with no SEPs, another third with generic SEPs, and the remaining students received specific SEPs. No statistically detectable difference was noted among the groups and the researcher cited participant non-compliance as one possible reason for the results. In another study, participants were instructed to read a text about the building process of a red-black tree in Computer Science (Chou & Liang, 2009). After which, participants were provided with trees and asked to judge whether the tree was binary. There were three groups in this study, one which received no prompts, one which received generic prompts, and one which received specific prompts. The results revealed that the two self-explanation groups performed better in applying target procedural knowledge, but the group that received the specific prompts had more negative self-monitoring, a self-explanation in which the participant was uncertain or questioning of the text, than the generic prompt group. The authors noted that this might mean that specific prompts made the participants more aware of their ignorance of the text.

3.2 Participant Recruitment from Crowdworking Platforms

The number of online experiments in the social sciences has swelled in recent years (Palan & Schitter, 2018; Stewart, Chandler, & Paolacci, 2017). A recent article in *Science* reported that experiments using participants sourced from Amazon Mechanical Turk (*MTurk*), a web platform that connects individuals with virtual tasks, grew from less than 70 in 2011 to more than 1,200 by 2015 (Bohannon, 2016). There are several reasons for this upsurge.

More and more researchers are gravitating towards online experiments in part because study results have demonstrated successful replication with results from in-person experiments. These replication studies include of a wide range of well-known experiments from economics and psychology (Amir, Rand, & Gal, 2012; Crump, McDonnell, & Gureckis, 2013; Horton, Rand, & Zeckhauser, 2011; Palan & Schitter, 2018; Suri & Watts, 2011). There is a growing body of evidence demonstrating that the data collected from online experiments are at least as reliable as data obtained from the laboratory or other in-person methods (Buhrmester, Kwang, & Gosling, 2011; Palan & Schitter, 2018). Researchers have found no difference in results between in person and online experiments in prisoner’s dilemma tasks (Horton et al., 2011), gender differences in risk taking (Eriksson & Simpson, 2010) and memory tasks (Finley & Penningroth, 2015).

For example, one study examined the behaviour of participants recruited online on four economics games and found the results to be comparable to those run in the laboratory (Amir et al., 2012). These games involved strategic interactions between individuals and were based on game theory. To demonstrate, Player 1 (the dictator) chose an amount equal or less than $1 to transfer to Player 2, resulting in Player 1 receiving the remainder of the money. The results showed an average transfer of $0.33, which is similar to the average of reported in a meta-analysis as 28.4% (Engel, 2011). Another example of a direct replication using similar economics games found agreement between contribution levels among online participants and in the laboratory participants (Suri & Watts, 2011). Based on a public goods game, Fehr and Gächter (2000) had participants play a 10-round game in which they received information after each round about the decisions of the other group members. They found evidence that indeed there is willingness to cooperate and punish those viewed as free-riders.

There is also evidence that participants respond consistently across experiments (Rand, 2012). When one experimenter examined the data from two different studies, conducted some time apart, he found that a small percentage of participants had responded to both experiments and 96% reported the same gender in both studies and 93% reported the same age.

Researchers are also turning to online platforms for participant recruitment and experimentation because they offer access to a broader, possibly more representative population. Although, the suitability of a diverse sample would depend on the nature of the question asked, results could be more broadly applicable if participant recruitment methods changed.

In an article published in *Nature*, Henrich, Heine, and Norenzayan (2010) argued that the vast majority of studies published in top psychology journals use participants from WEIRD (Western, Educated, Industrialized, Rich and Democratic) societies. A survey of the top psychology journals found that 96% of participants were from WEIRD societies, which contribute only 12% of the world’s population (Arnett, 2008).

Not only are most studies conducted with WEIRD participants, they recruit overwhelmingly white college students participating for class credit (Hanel & Vione, 2016; Peterson, 2001; Peterson & Merunka, 2013). Unease about the practise of recruiting and experimenting with such a narrow samples of convenience has been expressed by researchers for decades (Gosling, Rentfrow, & Swann, 2003; Henry, 2008; Sears, 1986; Smart, 1966). As McNemar (1942) wrote, "The existing science of human behavior is largely the science of the behavior of sophomores" (p. 333). Recruiting participants online allows researchers to tap into a wide pool of demographic diversity (i.e., ranges in age, geography, ethnic and economic backgrounds) (Casler, Bickel, & Hackett, 2013).

Gosling, Vazire, Srivastava, & John (2004) investigated this systemic issue by pooling participants characteristics of all studies published in 2002 of the premier journal in personality-social psychology (*Journal of Personality and Social Psychology*) and found that the sample consisted of 71% females, 80% average of white participants, and had a mean age of 22.9. This mostly demographically homogenous sample in which participants are in their later teens or early adulthood may have limited applications beyond their demographic group. The brain changes throughout a person's lifetime, and patterns in behaviour and cognition differ between this age group and older adults, children, babies and seniors (Lebel & Beaulieu, 2011). For example, undergraduate students have been shown to be more sensitive to reward when compared with older adults (Cohen et al., 2010). On the other hand, Paolacci, Chandler, and Ipeirotis (2010) found the average age of workers on *MTurk* to be 36 years old, which is younger than the average Canadian age of 40.8 (Canada, 2019), but significantly older than the average university student.

Traditional samples collected from the university environment also are over represented by highly educated individuals from high-income families (Autumn & Backhaus, 2009; *The Condition of Education - Letter From the Commissioner*, 2019). According to a report from the US Department of Education's National Centre for Education Statistics, students from the lowest quintile were more likely to pursue an associate degree (42%), while those in the wealthiest quintile were more likely to seek a four-year degree (78%) (National Center for Education Statistics, 2019).

Some have questioned whether participants recruited online are unmotivated (Azar, 2000; Buchanan, 2000). One of the benefits of laboratory or in-person studies is that the researcher is able to clinically screen participants for motivation, cognitive engagement, and non-investment (Oppenheimer, Meyvis, & Davidenko, 2009). However, studies have consistently reported that online participants are motivated by the extra source of income (Paolacci et al., 2010) and lower amounts of pay result in longer recruit times but equal data quality (Buhrmester et al., 2011). Online studies are similar to in-person studies in that safeguards and checks need to be put into place to ensure participants are paying attention, following directions, and sufficiently motivated. This includes pre-screenings, recording time on page and engagement metrics (i.e., clicks on page), and restricting access to only those in the pool who meet certain requirements (Casler et al., 2013).

However, these lines of inquiry into participant motivation lead one to wonder whether creating a highly motivating environment is not in of itself adding bias to the results. Or, put another way, how much external validity can exist in a laboratory experiment (Reips, 2000)? As Levitt and List write, “the critical assumption underlying the interpretation of data from lab experiments is that the insights gained can be extrapolated to the world beyond” (2007). This assumption has been shown to be tenuous in the literature. Participants have acted differently in laboratory settings than in a real life situation for a myriad of reasons including (a) participants knowing they are being monitored, recorded, or otherwise watched (List, 2006; McCambridge, de Bruin, & Witton, 2012; McCarney et al., 2007), (b) the context in which decisions are made (Goldberg & Gorn, 1987; Meyers-Levy, Zhu, & Jiang, 2010), and (c) the stakes or incentives involved (Bonner, Hastie, Sprinkle, & Young, 2000; Singer, 2002; Singer, Van Hoewyk, Gebler, Raghunathan, & Mcgonagle, 1999).

The Pygmalion effect (Brophy & Good, 1974; Harrison & List, 2004; Rosenthal & Jacobson, 1968) is one example of a validity concern with conducting traditional in-person experiments. Researchers have documented that performance expectations can positively or negatively influence actual performance, essentially creating a self-fulfilling prophecy (Bezuijen, van den Berg, van Dam, & Thierry, 2009). As such, it can be hypothesized that participants will exert discretionary effort on cognitive tasks in person because they believe the researcher has expectations of them. In-person experiments are also more likely to suffer from the observer-expectancy effect (Rosenthal, 1976), according to which the researcher's own cognitive bias causes them to subconsciously influence the participant, and the Hawthorne effect, according to which participants modify their behaviour in response to their awareness of being observed (Adair, 1984; Wickstrom & Bendix, 2000).

While evaluating external validity, one must consider mundane realism (Aronson & Carlsmith, 1968; Bauman, Peter McGraw, Bartels, & Warren, 2014), or "the extent to which events occurring in the research setting are likely to occur in the normal course of the subjects' lives, that is, in the 'real world'" (Aronson, Brewer, & Carlsmith, 1985, p. 485). Laboratory experiments are highly controlled situations in which students are often in an unfamiliar place behaving in possibly artificial ways (Martin, 1996; Orne, 1962; Reips, 2000; Winkler & Murphy, 1973). In contrast, participants in an online experiment remain in familiar situations on their own computer or phone (Reips, 2000), and this increased the ecological validity of the experiment (Christensen, 2001)

Another source of unease with using participants recruited online has focused on whether this sample is somewhat more depressed or maladjusted than the greater population (Kraut et al., 1998). This fear has largely been debunked (Gosling et al., 2004) and has subsided with the proliferation of internet access.

On the other hand, particular methodological challenges with online experiments have recently gained attention (Palan & Schitter, 2018). For one, it is possible that a professional participant community may be evolving in which the same group is volunteering for many experiments (Benndorf, Moellers, & Normann, 2017). It is unclear if this is happening but the types of experiments that suffer from practice effects may experience the most potential bias (Chandler, Mueller, & Paolacci, 2014; Palan & Schitter, 2018).

In conclusion, using online methodology to recruit participants and collect data is increasingly widespread and trusted (Casler et al., 2013; Davidov & Depner, 2011; Smyth & Pearson, 2011). With sensible safeguards and manipulation checks in place, data integrity in online experimentation is not a greater concern than traditional methods of data collection. In fact, there are some unique benefits of utilizing this approach that could broaden the applicability of results and more accurately reflect the intervention effect size. Although the distribution of diversity of participants in online experimentation is not perfect or 100% representative, it is still likely far superior to the traditional samples collected conveniently across university campuses.

3.3 Research Questions & Hypotheses

This goal of this study is to investigate the effect of receiving generic, specific, and no SEPs on recall and reading comprehension outcomes when the learning material is text. In previous work, the impact of generic vs. specific self-explanation on the outcome measure problem solving was studied (Bodvarsson, 2005; Chou & Liang, 2009).

Research question: When learners are asked to read brief text passages, do instructional conditions providing generic SEPs, specific SEPs, and no prompts have differing effects on learning outcomes?

1) Hypothesis: Participants who receive the generic SEPs will outperform participants who receive the specific SEPs and no SEPs on reading comprehension measures. Generic SEPs allow participants to fill their own knowledge gaps.

2) Hypothesis: Participants who receive the specific SEPs will outperform participants who receive the generic SEPs and no SEPs on the recall measure. Specific SEPs cue participants to focus on text elements related to the prompt, which will lead them to remember more on the free recall task.

3.4 Participant Recruitment

Participants were recruited using several online platforms, including “Find Partipants.com”, “Craigslist” and “Kijiji”. “FindParticipants.com” was chosen because it specifically caters to online experiments, whereas “Craigslist” and “Kijiji” were chosen because advertisement could be directed at specific locations. The ads requested English speakers. A sample ad is shown in Appendix A. Those who volunteered to participate in the experiment signed a consent form and were paid a $10 participation fee upon completing Session 1 and 2, and then $15 for completing Session 3 (see Appendix B).

3.5 Participants

The first session was completed by 190 participants, and 126 participants completed all three sessions. A portion of the participants dropped out after the first session, others after the second session, and a few participants were removed due to lack of engagement (i.e., time on passages was shorter than 5 seconds) or other technical issues (i.e. participants returned to previous page or an incorrect link was sent to the participant). Participants were randomly assigned to one of 6 groups factorially varying by order of text passage and type of prompt (*AB-no prompt, AB-generic, AB-specific, BA-no prompt, BA=generic, and BA-specific*).

The average age of the 126 participants was 25.67 (*SD* = 6.14) years. Overall, there were 86 females (68.3%), 39 males (31.0%), and 1 person did not provide their gender. Nineteen participants (15.1%) consider themselves as someone for whom English is an additional language or a secondary language. As shown in Table 9, almost all participants were living in Canada, the United States of America, or the United Kingdom.

Table 9: Current residence of participants

Country		n		Percentage	
Canada		46		36.5%	
	British Columbia		31		24.6%
	Ontario		15		11.9%
	Nunavut		1		.8%
USA		45		35.7%	
United Kingdom		31		24.3%	
Other		4		3.2%	

Table 10 summarizes the highest level of education completed by the participants with 55.5% having completed at least one post-secondary degree and 48% currently enrolled in a higher education program. Of those who graduated with a post-secondary degree, the majority have a degree in Arts and Social Sciences (34%), Business (19%), and Science (19%).

Table 10: Highest level of completed education

Education level	n	Percentage
High School	56	44.4%

Bachelor	56	44.4%
Masters	12	9.5%
Doctorate	2	1.6%

3.6 Materials & Instruments

All three sessions of the experiment were administered using an online survey software (*Fluidsurveys*). This software allows for questions to be asked in the form of radio-buttons, textboxes, and drop-down menus.

3.6.1 Purpose of Study and Consent Form

A downloadable pdf with information (see Appendix C) detailed the purpose and goals of the study, risks, payment, and a statement of confidentiality as required by the university's Office of Research Ethics. A consent form was provided along with the research study information (see Appendix A) for the participant to read, review, and digitally sign-off.

3.6.2 Questionnaire

The "Participant Questionnaire" (see Appendix E) was created and administered digitally via *Fluidsurveys*. The questionnaire requested information about age, gender, current country of residence, English as an additional language or secondary language status, level of education completed, graduating CGPA, and current faculty/program of study.

3.6.3 Need for Cognition

Cacioppo, Petty, and Kao (1984) authored the short form of the Need for Cognition Scale (NCS) that consisted of 18 items and measures "an individual's tendency to seek, engage in and enjoy effortful cognitive activities" (Zhou, 2008, p.48). The responses to each of the items are answered on a 7-point scale ranging from 1 (not at all true of me) to 7 (very true of me). Participant responses were combined over all items as need for cognition (NFC) is conceptualized as a one-dimensional construct (Cacioppo, Petty, Feinstein, & Jarvis, 1996). This instrument was chosen as another measure of individual differences among the participants.

3.6.4 Reading Passages

Three short reading passages were written to be approximately the same length and reading difficulty (see Table 11). The rationale for writing them instead of copying previous passages was so that participants could not find the answers to the multiple choice and short answer questions online. The texts were expository and written in such a manner as to not situate themselves easily into a single discipline, nor contain high-level vocabulary. The questions were based on the information provided in the passage or information/ideas that could be inferred.

Corresponding short answer and multiple-choice questions were developed for each. These reading passage and questions were piloted with 10 undergraduate and graduate volunteers for clarity, difficulty, and tonal comparativeness. After this process, the third passage titled, “Common Property” was dropped from the experiment and minor edits were made to the other two passages and question sets. See Appendix D for all three passages and related multiple-choice questions.

Table 11: Passage Descriptives

Reading Task	Title	Words	Sentences	Paragraphs	Flesch-Kincaid Grade level
A	Overpopulation	535	27	7	12.5
B	Nudge policy	573	29	6	12.6
C	Common Property	569	25	7	12.8

3.7 Dependent Measures

3.7.1 Free Recall

Participants received the following free recall prompt, “Please type out as much as you can remember about the passage you just read.” A conventional proposition scoring method was used to assess the free recall test (Nesbit & Adesope, 2011; Rewey, Dansereau, Skaggs, Hall, & Pitre, 1989). The “Overpopulation” passage and the “Nudge policy” passage were decomposed into 56 and 60 propositional statements respectively that were equivalent in meaning to the original passages (see Appendix J). For example, the sentence, “The global population began to rise quickly as antibiotics, vaccines, and technology increased the expected lifespan of humans.” was converted into four propositions: (1) the global population began to rise quickly, (2) global populations increased as the expected lifespan of humans increased (3) antibiotics led to increased expected lifespan, (4) vaccines led to increased expected lifespan, and (5) technology led to increased expected lifespan. A rater, unaware of condition assignment, scored the free recall responses awarding a score of 0 (absent or inaccurate proposition), 1 (partially accurate proposition), or 2 (accurate proposition). To boost reliability, an iterative approach was adopted with a second rater. A trained second rater would score 10 randomly selected participants, and then the two raters would meet to discuss differences. This cycle occurred twice for each passage, and then interrater reliability was assessed on 20 randomly chosen participant responses by calculating an intraclass correlation coefficient (0.95 with a 95% confidence interval 0.94-0.95). At this point, the remaining free recall participant responses were coded with Rater 1 scoring almost 75% of them and rater 2 scoring the rest.

Appendix J is a sample of the scored free recall test. Each participant's free recall score was the sum of the proposition scores.

A second analysis investigated whether group assignment had an impact on the recall of propositions directly related to the specific SEPs. For example, the following two propositions were identified as being related to the SEP, "Type out your self-explanations about stagflation below."

- World oil shocks contributed to "stagflation".
- Stagflation is a combination of rising unemployment with higher prices

For the reading passage titled "Overpopulation", there were 13 propositions that were identified (possible score of 26) and for the second passage titled "Nudge Policy" there were 15 propositions that were identified (possible total score of 30) as being related to the content of specific SEPs. Each score was turned into a percentage and a two-way ANOVA was computed to examine the effects of group assignment and reading task on the recall of these propositions.

3.7.2 Short-answer Questions

Three short-answer questions related to each passage and were designed to measure comprehension. The complete answer for each question was split into propositional statements (see Section 3.7.1). The three “Overpopulation” short answer questions had 9 propositions in total, whereas the “Nudge policy” passage had 10 propositions in total. A rater scored each question response by assigning each proposition a 0 (absent or inaccurate), .5 (incomplete) or a 1 (complete). Two raters scored 20 randomly chosen participant responses and then to assess inter-rater reliability, an intraclass correlation coefficient (ICC) was run. The ICC was 0.96 with a 95% confidence interval of 0.93-0.98. At this point, the remaining short-answer question participant responses were coded with Rater 1 scoring almost 85% of them and rater 2 scoring the rest. Each participant’s short answer score was the sum of the proposition scores

3.7.3 Multiple-choice Questions

Each passage was accompanied by five multiple-choice questions designed to measure reading comprehension. These questions were scored either 0 (incorrect) or 1 (correct), and the sum was the participant’s comprehension score for the passage.

3.7.4 Time-On-Task

In this study, time-on-task is operationalized as how long participants spent reading the text during Session 3. For the two self-explanation groups, this also included the time it took them to type-out their self-explanations; and for the no SEPs group, this included the time it took them to write-out any notes in the textboxes. As such, an ANOVA was performed with group assignment as the independent variable and the outcome measure as time-on-task. A separate analysis was conducted for each reading task (A and B) because there was no reason to believe that the time it would take a participant to go through the two passage would be the same.

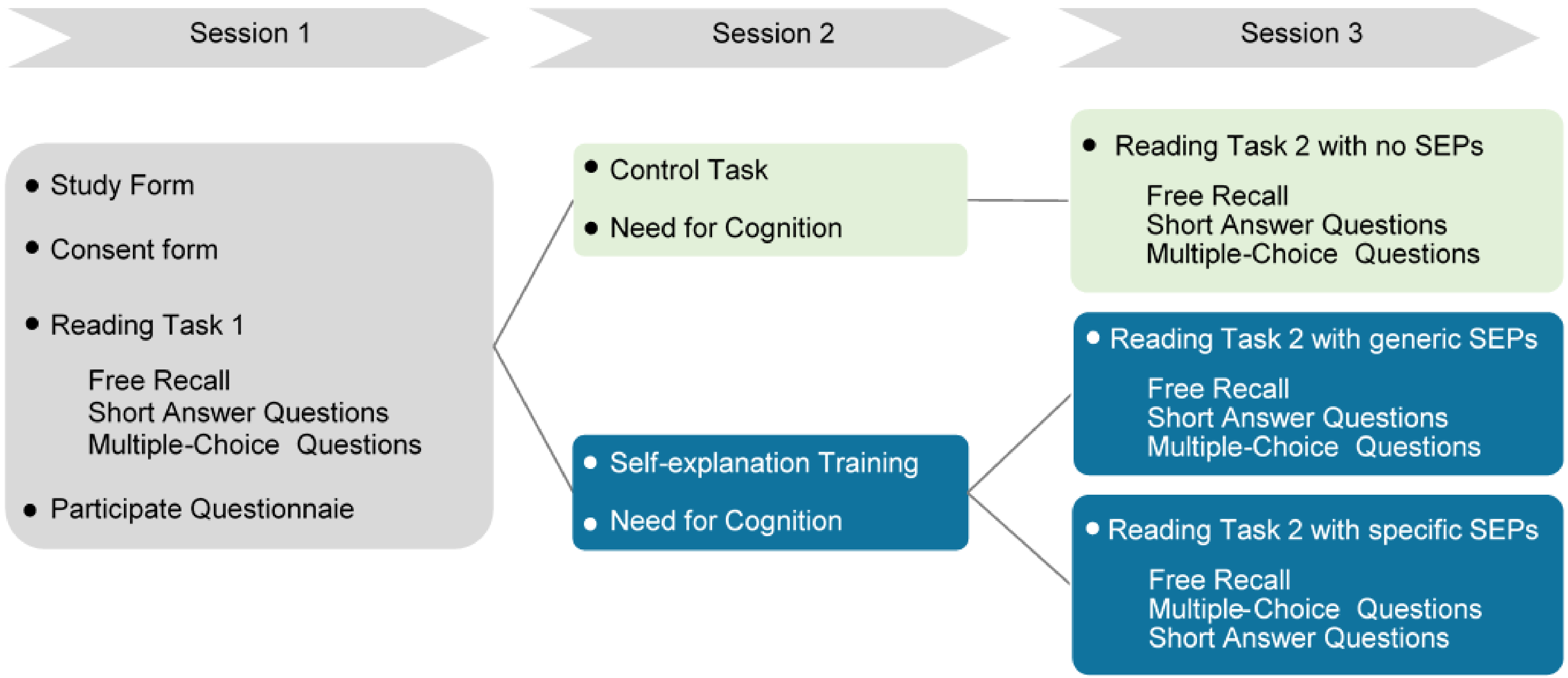
3.8 Procedure

The reading task consisted of 1) reading a short passage, 2) completing a free recall about the passage, 3) answering three short answer questions, and 4) answering five multiple-choice questions. The reading passages that were used in this experiment are described above as “Overpopulation” (A) and “Nudge Policy” (B).

There were also three-levels in the prompt treatment, which was only presented in Session 3 (see below). Participants in the “no SEPs” condition were given the following instructions, “Textboxes have been provided for you if you wish you take notes. You are NOT required to you use them.” and presented with three text boxes throughout the reading passage. The participants in the two self-explanation conditions were provided with three prompts at locations in each reading passage. The generic SEP was “Type out your self-explanation below.” Participants in the specific SEPs condition received prompts containing content about the passage, for example, “Please self-explain the Malthusian view.” See Appendix D for a complete listing of all SEPs.

Figure 1 presents the procedure for each of the three sessions. At the beginning of Session 1, participants were provided information about the study and the consent form before they were presented with a Reading Task. Once they completed Reading Task 1 (A or B), the last part of Session 1 was to complete the participation questionnaire. During Session 2, participants either watched a video and answered questions about it (control task) or underwent self-explanation training. Participants also responded to the “Need for Cognition” questionnaire at the end of Session 2. Participants were asked to complete Reading Task 2 during Session 3. Depending on prompt treatment assignment, participants either received no SEPs, generic SEPs, or specific SEPs during their reading.

Figure 1: **Procedure showing phases within each session**



3.8.1 Experiment Factor Design

This study used a 2 x 2 x 3 factorial design in which the between-subjects factors were (1) order of reading tasks and (2) prompt type and the within-subject factor (session) was whether the score was captured during Session 1 or Session 3. Table 12 summarizes the experimental design conditions.

Table 12: Experiment design conditions

		Prompt treatment			Session	
		No SEPs	Generic SEPs	Specific SEPs	1	3
Reading Task Order	Group AB	AB – No SEPs	AB – Generic SEPs	AB – Specific SEPs		
	Group BA	BA – No SEPs	BA – Generic SEPs	BA – Specific SEPs		

3.8.2 Self-explanation Training Instructions

Participants were provided with a set of instructions on self-explanation (Appendix H). The self-explanation training consisted of a series of slides adapted from earlier materials used by Bielaczyc, Pirolli, and Brown (1995), Ainsworth and Burcham (2007) and Hodds, Alcock, and Inglis (2014). The slides explained the benefits of self-explanation training and clarified the principles of SE: (a) identifying key ideas in each line of a text and (b) explaining each line in terms of previous ideas presented in the proof or in terms of previous knowledge. The slides then demonstrated the self-explanation strategy via an example.

3.8.3 No SEPs Task

Participants in the no SEPs group watched a TedTalk video by Jonathan Drori about various ways flowering plants have evolved to attract insects to spread their pollen (Drori, 2011). After watching, they answered three short answer questions about the video. This session for the no SEPs group was added as a safeguard to limit unwanted or potentially selective attrition differences between prompt treatment groups (Horton et al., 2011; Shank, 2016).

3.9 Results

All variables were examined for outliers and normality through SPSS features (i.e., frequencies, descriptive statistics). No outliers were detected and distributions were mostly normal, falling within acceptable levels of skewness and kurtosis (Tabachnick & Fidell, 2007, p. 79).

Subsequently, the free-recall and short-answer scores were transformed onto the same scale to allow for comparisons between the two reading passage tasks. For example, the free recall score was calculated to be out of 60 propositions even though the "Overpopulation" passage had 60 propositions, and the "Nudge policy" passage had 56 propositions.

3.9.1 Session 1 Scores

Table 13 contains a summary of the Session 1 scores for the three dependent variables, Cronbach's alpha coefficients, and test information to determine if there were differences in scores between Reading Task A and B.

A Welch *t*-test was computed to investigate differences in free recall scores between Reading Task A and B due to the assumption of homogeneity of variances being violated, as assessed by Levene's test for equality of variances (p = .007). Welch's *t*-test was chosen because it is robust to this violation when sample sizes are approximately equal. Mean free recall score for Reading Task A (12.17 ± 8.61) was lower than Reading Task B (17.39 ± 11.16), a statistically detectable difference of 5.21 (95% *CI*, 1.69 to 8.74), $t(114.7) = 2.93$, $p = .004$.

Independent-samples *t*-tests were computed to examine whether there was a difference between Reading Task A and B short-answer and multiple-choice scores. There was not a statistically detectable difference between the mean scores ($p > .05$).

Table 13: Session 1 Reading Task scores

	Reading Task A				Reading Task B						
	N	M	SD	*Cronbach's* ***α***	*N*	*M*	*SD*	*Cronbach's* ***α***	*t*	*df*	*Sig (2-tailed)*
Free Recall	64	12.17	8.61		62	17.39	11.16		2.93	114.7	0.004
Short Answer Questions	63	3.44	2.05	0.65 (3)	61	3.44	2.82	0.59 (3)	0.00	122	0.997
Multiple Choice Questions	64	2.73	1.212	0.43 (5)	62	3.02	1.43	0.61 (5)	0.64	124	0.235

As shown in Table 14, there were statistically detectable, moderate positive correlations among the three dependent variables per reading task. The one exception was that there was the correlation between free recall and short answer questions scores for Reading Task B was slight, *r*(61) = .24, *p* =.061.

Table 14: Relationship amoung dependent variables from Session 1

		Short Answer Questions	Multiple Choice Questions
Reading Task A	Free Recall	.592**	.496**
	Short Answer Questions	-	.362**
Reading Task B	Free Recall	0.241	.485**
	Short Answer Questions	-	.442**

**Correlation is statistically detectable at *p* ≤ .01 (2-tailed)

3.9.2 Session 3 Scores

Table 15 summarizes the Session 3 scores and Table 16 displays the correlations among the three dependent variables per reading task.

Table 15: Descriptive Statistics for Session 3 scores

		Control			Generic			Specific		
		N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SE*
Group AB	Free Recall	24	12.41	9.27	21	12.14	8.29	19	11.90	8.53
	Short Answer Questions	24	3.25	2.19	20	3.63	2.06	19	3.50	1.95
	Multiple Choice Questions	24	2.63	1.35	21	2.86	1.15	19	2.74	1.15
Group BA	Free Recall	21	12.3	9.71	22	14.27	8.35	19	14.66	8.08
	Short Answer Questions	20	2.75	2.42	22	3.98	2.14	19	3.37	2.50
	Multiple Choice Questions	21	2.76	1.14	22	2.91	0.971	19	2.50	1.20

Table 16: Relationship amoung dependent variables for Session 3

		Control		Generic		Specific	
		Short Answer Questions	Multiple Choice Questions	Short Answer Questions	Multiple Choice Questions	Short Answer Questions	Multiple Choice Questions
Group AB	Free Recall	.592**	.433*	0.034	0.281	0.19	.504*
	Short Answer Questions		.549**		0.218		0.37
Group BA	Free Recall	.800*	.458*	0.241	0.149	.512*	.625**
	Short Answer Questions		.538*		0.183		.552*

* Correlation is statistically detectable at $p \leq .05$ (2-tailed)
**Correlation is statistically detectable at $p \leq .01$ (2-tailed)

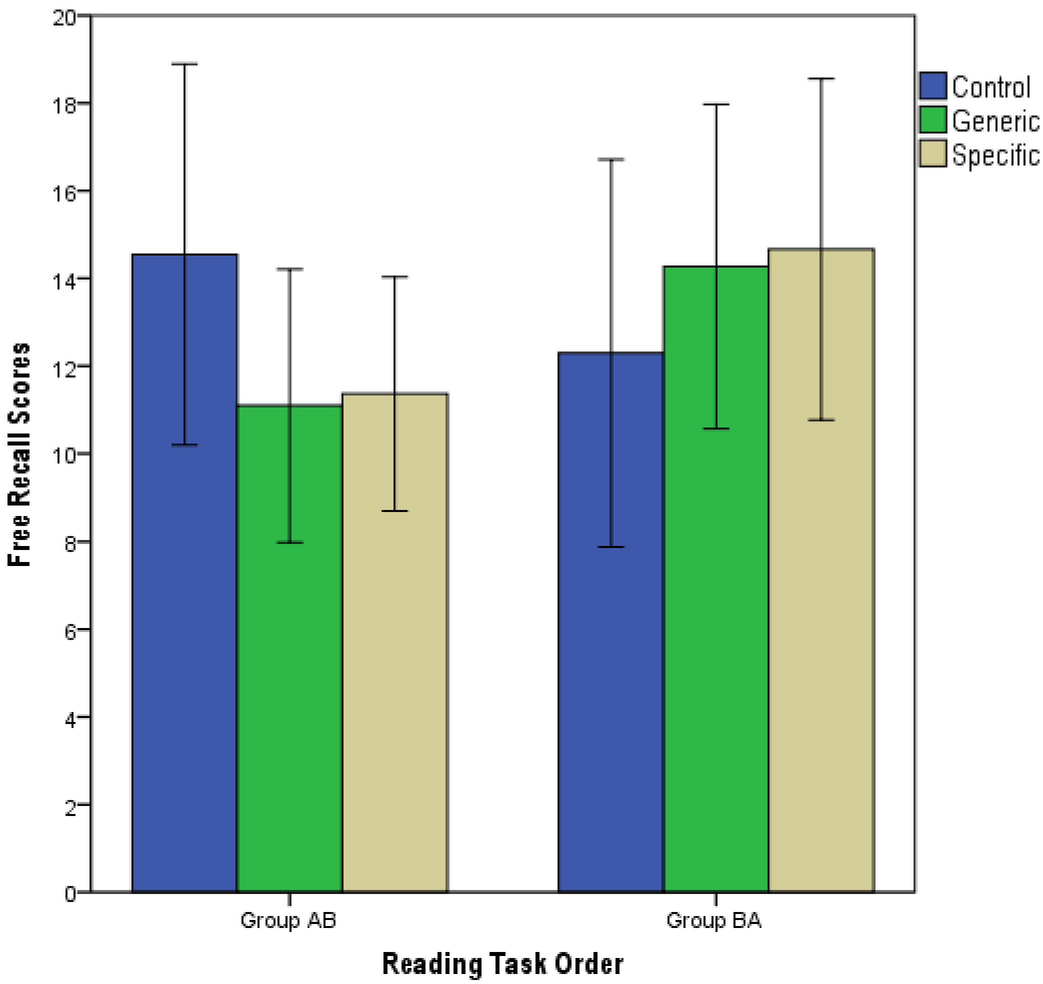
The following section will report an analysis using a three-way mixed ANOVA to investigate the effects of Reading Task order, prompt treatment, and Session 1 scores on each of the three dependent variables measured in Session 3.

A second analysis approach was used to verify the results. A two-way ANCOVA was conducted to examine the effects of reading task (A/B) and prompt treatment on Session 3 short-answer and multiple-choice question scores after controlling for Session 1 scores. As well, a two-way ANCOVA was conducted to examine the effects of reading task (A/B) and treatment on Session 3 free recall scores. The results for these three analyzes were the same as the results reported below.

3.9.2.1 *Free Recall*

Session 3 free recall mean scores and error bars for each prompt treatment group, by reading task, are presented in Figure 2.

Figure 2: Free Recall mean scores with 95% CI for each prompt treatment group by reading task order



The distribution of the free recall scores did not detectably depart from normal according to Shapiro-Wilk's test (ranging from p = .01 - .57), and they had acceptable levels of kurtosis and skewness (less than 1 and greater than -1) (Tabachnick & Fidell, 2007). Eight outliers were identified and removed, as assessed by inspection of a boxplot. There was not homogeneity of variances for these scores as assessed by Levene's test for equality of variances. After a data transformation (Log10), there was homogeneity of variances for both Session 1 (p = .102) and Session 3 free recall scores (p = .240), as assessed by Levene's test for equality of variances.

The three-way interaction between reading task, prompt treatment and Session1 scores was not statistically detectable, $F(2, 102) = .907$, $p = .907$ partial $\eta 2 = .002$. There was a statistically detectable two-way interaction between Session 1 and 3 free recall scores and reading task, $F(2, 102) = 8.595$ $p = .004$. All other two-way interactions were not statistically detectable ($p > .05$). Statistical detectability of a simple two-way interaction was accepted at a Bonferroni-adjusted alpha level of .025. There were no statistically detectable simple two-way interactions of reading task and prompt treatment on the Session 1 scores, $F(2, 106) = 1.65$, $p = .196$, or the Session 3 scores $F(2, 107) = 1.202$, $p = .305$.

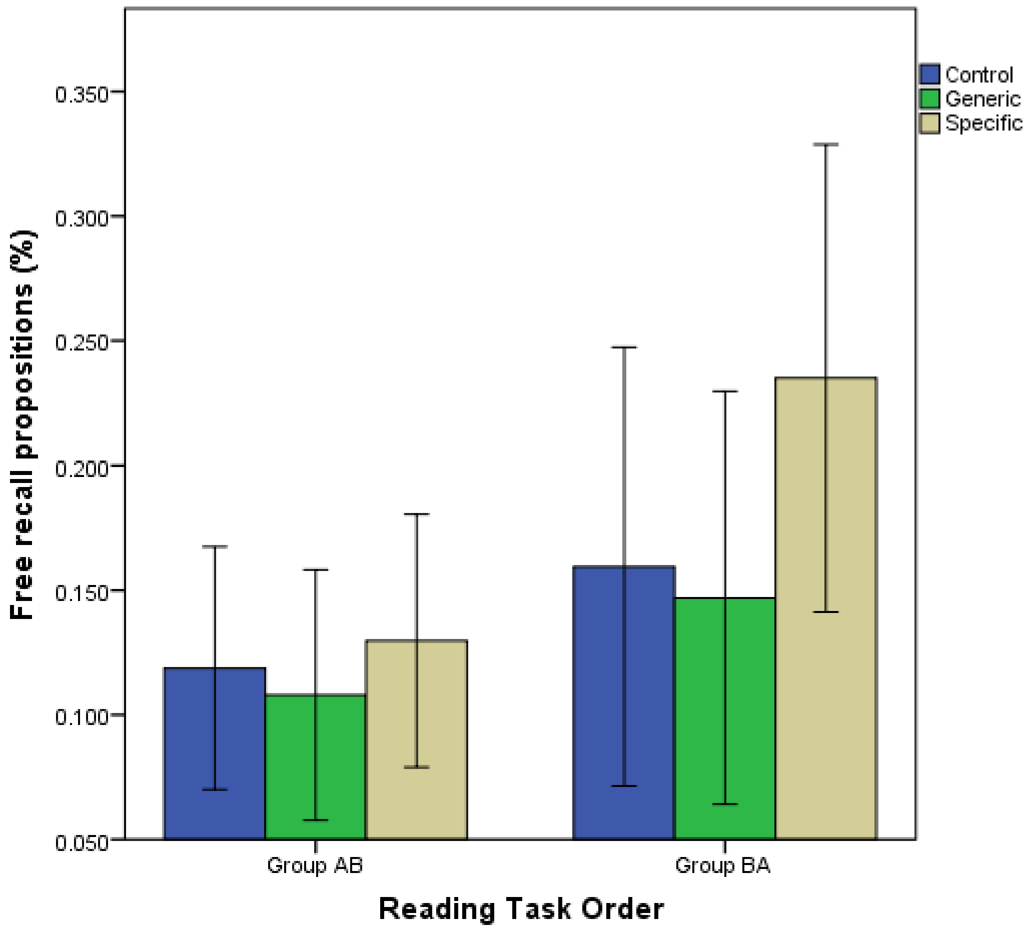
3.9.2.1.1 Recall of propositions related to the content of Specific SEPs

A two-way ANOVA was computed to examine the effects of prompt treatment and reading task on the recall of Session 3 propositions related to the specific SEPs. Table 17 displays the descriptive statistics for the recall scores. Figure 3 displays means with 95% confidence intervals.

Table 17: Descriptive Statistics of Recall scores on propositions directly related to specific SEPs both Groups AB and BA

		N	*M*	*SD*
Group AB	Control	22	10.4%	9.2%
	Generic	21	10.8%	11.1%
	Specific	18	11.3%	7.8%
Group BA	Control	21	15.9%	19.3%
	Generic	20	10.3%	13.0%
	Specific	19	23.5%	19.4%

Figure 3: Percentage of propositions related to the specific SEPs recalled by reading order and SEP type assignment



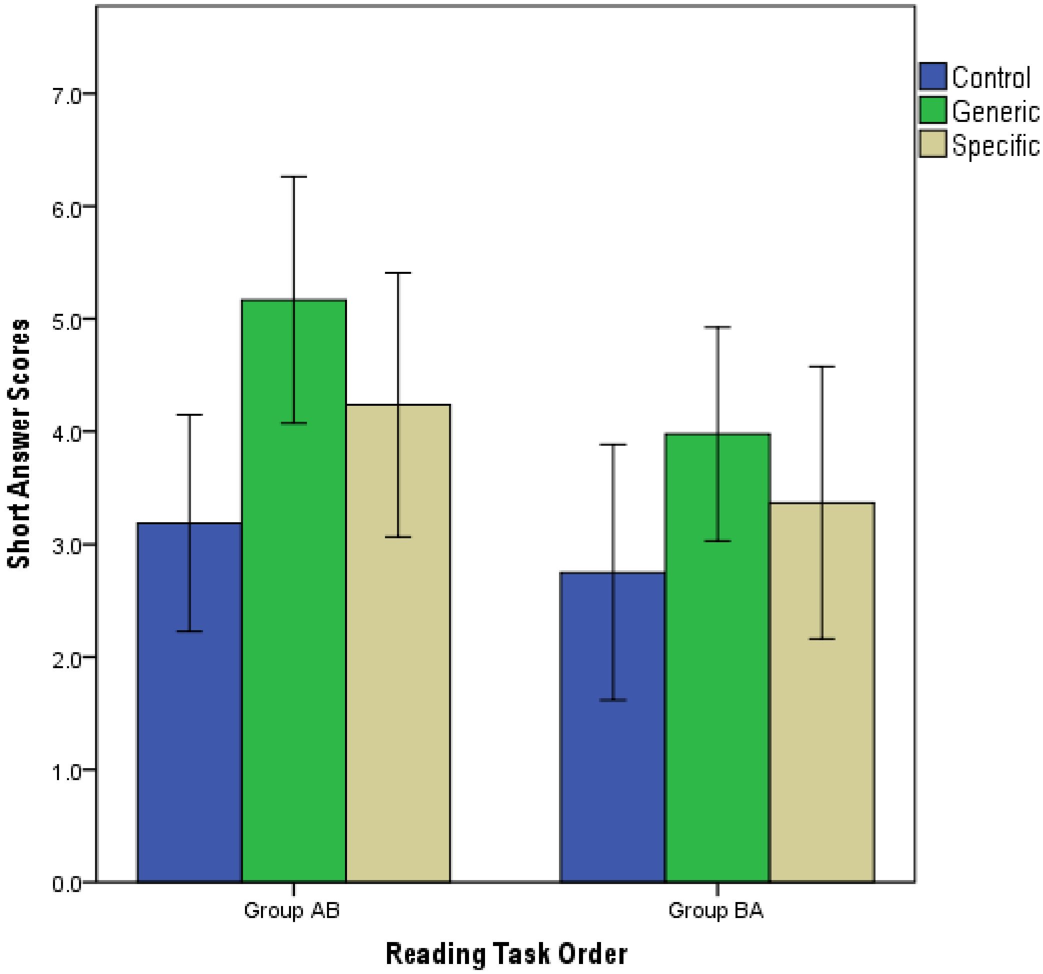
There were four outliers, as assessed as being greater than 1.5 box-lengths from the edge of the box in a boxplot. Since several of the cells (group assignment by reading task) displayed data that was not normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), the scores were transformed using Log10. The recall scores were mostly normally distributed considering acceptable levels of kurtosis and skewness (Tabachnick & Fidell, 2007). There was homogeneity of variances, as assessed by Levene's test for equality of variances, $p = .454$.

There was a statistically detectable interaction between group assignment and reading task for recall scores that were directly related to the specific SEPs, $F(2, 114) = 4.53$, $p = .013$, partial $\eta 2 = .074$. There was a statistically detectable main effect of Reading Task order, $F(1,114) = 19.284$, $p < .0005$, partial $\eta 2 = .145$. However, we are more interested in the main effect of group assignment, which was also statistically detectable, $F(2,114) = 3.336$, $p < .039$, partial $\eta 2 = .055$. Participants who received generic SEPs had a lower recall score on the propositions related to the content in specific SEPs, a statistically detectable difference, $p = .036$. No statistically detectable difference was found in these recall scores between the generic SEPs and no SEPs groups, nor the specific SEPs and no SEPs groups.

3.9.2.2 Short-Answer Question

Session 3 short-answer question mean scores and error bars for each prompt treatment group, by reading task, are presented in Figure 4.

Figure 4: **Session 3 short answer question mean scores with 95% CI for each of the prompt treatment groups, by reading task**



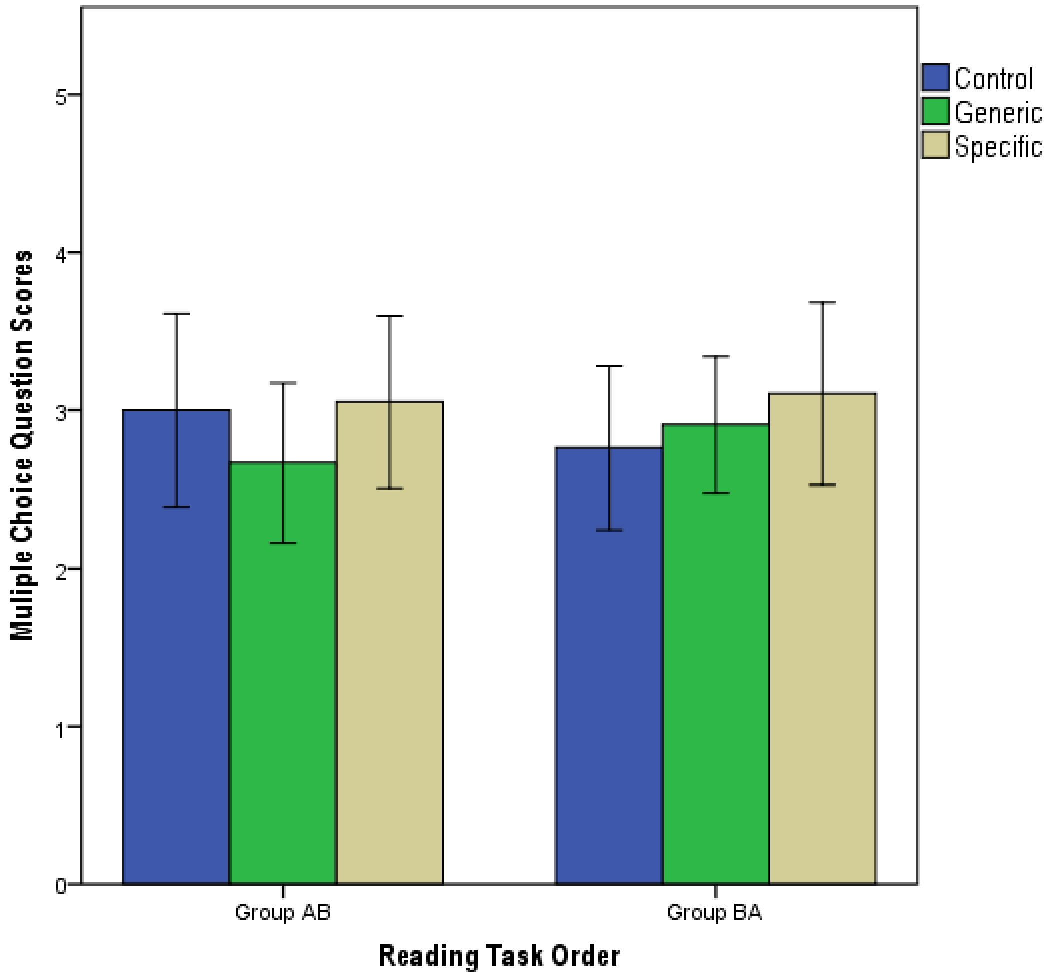
The distribution of the short answer scores did not detectably depart from normal according to Shapiro-Wilk's test (ranging from p = .01 - .82), and had acceptable levels of kurtosis and skewness (Tabachnick & Fidell, 2007). There were no outliers in the data, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. There was homogeneity of variance for both Session 1 short answer scores (p = .249) and Session 3 short answer scores (p = .960), as assessed by Levene's test.

The three-way interaction between reading task, prompt treatment and Session 1 scores on Session 3 short answer question scores was not statistically detectable, F(2, 117) = .091, p = .913 partial $\eta 2$ = .002. There was a statistically detectable two-way interaction between session and prompt treatment, F(2, 117) = 5.358, p = .006. All other two-way interactions were not statistically detectable ($p > .05$). In other words, the short answer questions scores were statistically detectably different from Session 1 to 3 between the prompt treatments. Statistical detectability of a simple two-way interaction was accepted at a Bonferroni-adjusted alpha level of .025. There was a statistically detectable simple main effect of prompt treatment during Session 3, $F(2, 117) = 4.784$, $p = .010$, but not during Session 1, $F(2, 117) = .208$, $p = .812$. All pairwise comparisons were performed for statistically detectable simple main effects. Bonferroni corrections were made with comparisons within each simple main effect considered a family of comparisons and only adjusted *p*-values are reported. Mean short answer score was higher in the generic SEPs than the no SEPs Group, a statistically detectable mean difference of 1.59 (95% *CI*, .340 to 2.830), $p = .007$.

3.9.2.4 Multiple-Choice Questions

Session 3 multiple-choice question mean scores and confidence intervals for each prompt treatment group, by reading task, are presented in Figure 5.

Figure 5: Session 3 multiple choice question mean scores with 95% CI for each of the prompt treatment groups, by reading task



The distribution of the multiple-choice answer scores did not detectably depart from normal according to Shapiro-Wilk’s test (ranging from p = .01 - .07), and considering acceptable levels of kurtosis and skewness (Tabachnick & Fidell, 2007). Three outliers in the data were identified and removed, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. There was not homogeneity of variances as assessed by Levene's test for equality of variances. The sample sizes are approximately equal, therefore the decision to proceed with a three-way mixed ANOVA was made.

The three-way interaction between reading task, prompt treatment and Session 1 on Session 3 short answer question scores was not statistically detectable, $F(2, 117) = .181$, $p = .834$ partial $\eta 2 = .003$. There were not statistically significant two-way interactions either.

3.9.3 Time-On-Task

Six outliers were identified and removed, as assessed by inspection of a boxplot (i.e. greater than 1.5 box-lengths from the edge) for Group AB. The time-on-task for each group (no SEPs, generic SEPs, specific SEPs) were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and considering acceptable levels of kurtosis and skewness (Tabachnick & Fidell, 2007). There was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .703$).

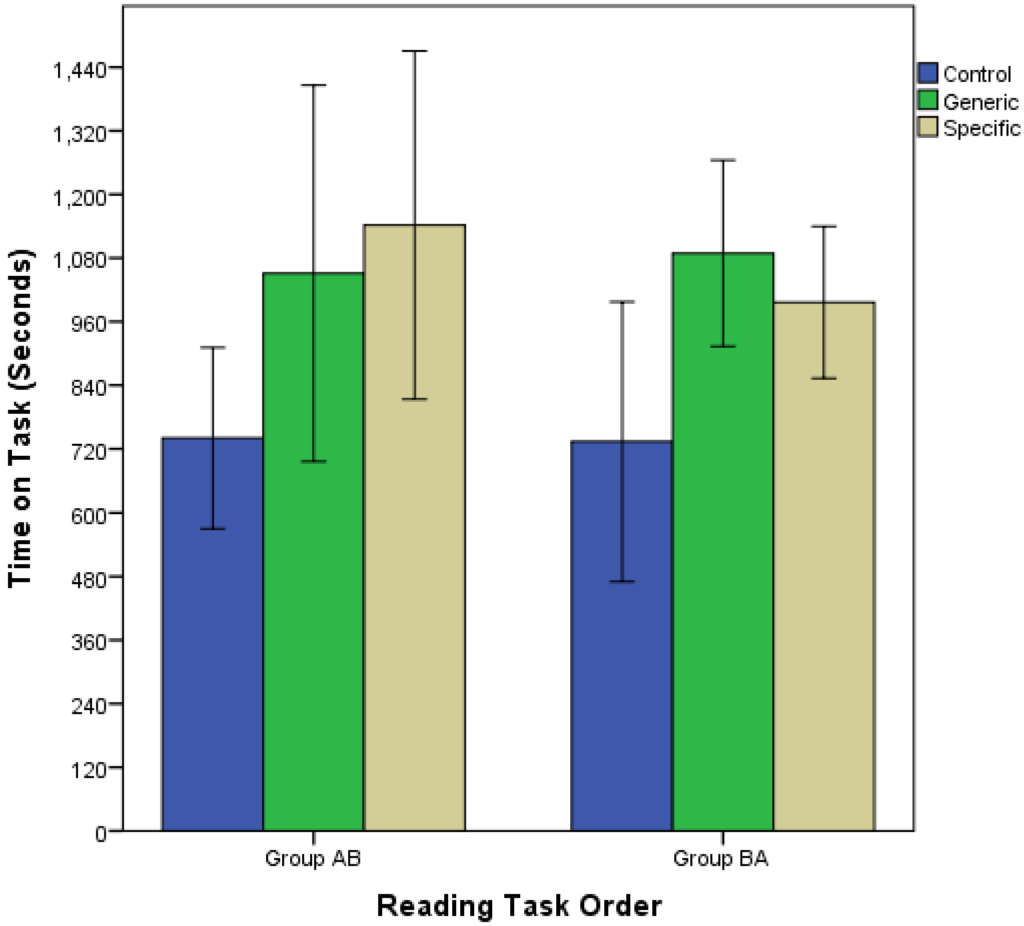
By inspection of the boxplot (i.e. greater than 1.5 box-lengths from the edge), 3 outliers were removed for Group BA. The time-on-task was not normally distributed, as assessed by Shapiro-Wilk's test ($p < .05$), so a data transformation was performed (SQRT). There was homogeneity of variances, as assessed by Levene's test ($p = .162$).

The mean and standard deviation are reported in the table below. As well, Figure 6 depicts the means and 95% confidence intervals by reading task order and prompt type.

Table 18: Descriptive Statistics of Time-on-Task by reading task order and prompt type.

		N	*M (m:s)*	*SD (m:s)*
Group AB	Control	19	12:21	5:53
	Generic	18	13:06	5:25
	Specific	21	14:58	7:20
Group BA	Control	18	11:50	7:06
	Generic	22	18:08	6:36
	Specific	19	16:36	3:59

Figure 6: **Mean time-on-task (seconds) with 95% CI Error Bars for each of the prompt treatment groups, by reading task**



There were no statistically detectable differences in time-on-task score among the different assignment groups, *F*(2, 55) = .920, *p* = .404 for Group AB. However, there were statistically significant differences in time-on-task among the different assignment groups for Group BA, *F*(2,56) = 6.57, *p* =.003.

There was an increase in time comparing the no SEPs group to the generic SEPs group, an increase of 6 minutes and 18 seconds (95% *CI*, 1:30 to 11:07), which was statistically significant (*p* = .007). Similarly, there was an increase in time from the no SEPs group to the specific SEPs group, an increase of 4 minutes and 46 seconds (95% *CI*, 0:12 to 9:47). However, there was not a statistically detectable difference in duration spent on task between the generic and specific SEPs groups.

3.10 Discussion

This study examined the effects of content-free (generic) and content-specific (specific) SEPs on achievement by participants studying expository text. The results support the hypotheses that generic SEPs are superior to specific SEPs and receiving no SEPs when achievement is measured using a short-answer question format.

Although two measures of reading comprehension were used in this study, the difference was only detected in the short-answer format. I speculate that this might be because responding to generic SEPs and short-answer questions use parallel mechanisms. Multiple-choice questions do not require participants to formulate answers in their own way. In contrast, participants can only respond to short-answer questions by using their own words and can communicate their understanding of the text by more fulsome means. Generic SEPs also may encourage users to fill-in their own knowledge gaps and use their own words to explain the text to themselves.

Another reason that one reading comprehension outcome yielded a statistically detectable difference among groups while another did not may be there were too few multiple-choice items. The decision to limit the number of questions to five was made as to limit study fatigue. However, the internal consistency of the multiple choice comprehension questions was weak, with Cronbach's alphas of 0.43 (Reading Task A) and .62 (Reading Task B). Perhaps the study could have benefited from adding several more parallel items to increase the reliability of these measures.

Strategies differ with respect to the types of learning outcomes they promote. For example, strategies such as keyword mnemonic or imagery of text have been shown to improve a participant's memory of facts, while others best serve comprehension (i.e. elaborative interrogation (Dunlosky et al., 2013). This study also investigated the effect of self-explanation on a free recall task in which reading an expository text was the learning activity. Contrary to the second hypothesis (page 54), no differences in scores were detected on this measure. This result bolsters the theoretical underpinnings of self-explanation as a mechanism which fills knowledge gaps in mental models (Chi & Bassok, 1989; Lin & Lehman, 1999; VanLehn et al., 1992), detects conflict in and revises mental model revision (Chi, De Leeuw, et al., 1994; DeLeeuw & Chi, 2003), but is not as effective during an information encoding task. Success on a free recall task is less linked to creating robust mental models than reading comprehension.

I also investigated whether receiving the content-specific SEPs impacted the recall of the propositions related to that content. The results revealed that the participants who received generic SEPs had a lower recall score on the content-specific propositions when compared to the group who received the specific SEPs. This result provides evidence that when a learner is provided with a specific SEP, they will attune themselves to content of the prompt. No difference in the recall of these propositions was statistically detected between specific SEPs group and the no SEPs group.

This result further supports the notion that receiving generic prompts drives learners to make meaning for themselves and fill in their own knowledge gaps. Since participants who received the content-specific SEPs tended to focus more on the embedded content, they were less likely to engage with the ideas presented by the entire text and how the entire text related to prior knowledge. Constraints on working memory may not have allowed the participants to engage in both cognitive tasks at the same time.

Current information processing theories draw upon ideas of working memory and its interplay with long-term memory. Working memory is the component of human memory that organizes and processes incoming information and connects it with prior knowledge (Baddeley & Hitch, 1974). It has a limited capacity and can only hold and manipulate a few units of information at any given time. It is the interactions between the new information in working memory and the prior knowledge loaded into working memory from long-term memory that result in learning (Novak & Cañas, 2006).

One of the factors in this research design is reading task order. There were two reading passages that were used in this study, and although best efforts were made to make them equal in terms of reading difficulty level, structure, and length, small differences were detected in measures of achievement. For example, the scores on the free recall task during Session 1 between Reading Task A and B were detectably different with participants remembering more about the “Nudge Policy” passage. This may have been due to a semantic construct of the passage that allowed participants to more easily recall the information.

The effect of prompt treatment on time-on-task was also examined. Time spent on the reading task page during Session 3 was operationalized as the time-on-task. A statistically detectable difference was found for reading task order Group BA participants, but not for Group AB. The no SEPs group not receiving SEPs spent less time on this page than the groups who received SEPs. This mixed result is somewhat consistent with the literature as some studies have found participants who receive SEPs spent more time-on-the task (Chi, De Leeuw, et al., 1994), while others have not reported a difference (Ainsworth & Burcham, 2007; Schworm & Renkl, 2006a).

It may also be important to note that the time-on-task in this type of an experiment might be difficult to measure. Since participants were not monitored and completed the study from various locations, it is difficult to ground inferences about what the participants were doing when the reading task page was active. For example, a participant may have left the page open and gone to the bathroom. In that case, the software would have logged the bathroom time as time-on-task. This type of measurement error might also help to explain the large standard deviations observed in the time-on-task scores. Chandler et al. (2014) reported that almost 1/5 participants engaged in activities simultaneously as MTurk tasks, while Necka, Cacioppo, Norman, & Cacioppo (2016) found 2/5 participants were multitasking.

However, in exchange for some uncertainty and lack of control over the environment, this web-based experiment gained validity in other respects. It was designed to closely mimic a student's natural study setting and practises and had participation from a broader demographic.

For one, the results demonstrated the benefits of self-explanation without over-prompting participants. For example, Chi et. al (1994) reported that participants who self-explained had greater gains on a biology exam than participants who did not receive those prompts. However, these participants were asked "to read each sentence out loud and then explain what it means to you." Focusing on each sentence is just not a practical and time-efficient study strategy for most students. The current experiment only had three SEPs embedded throughout the text and no mechanism to ensure that participants responded to each prompt. As such, participants could choose to respond to any or all of the prompts as they judged to be appropriate. This mimics real-life situations in which students are provided with tools and strategies, and it is up to them whether and when they use them.

The familiarity of the physical environment in which participants completed the sessions mimics a real-life study situation and therefore boosts the external validity of the experiment. Chandler et.al (2014) reported that 86% of participants completed MTurk tasks from home. One can imagine that in real-life study situation, participants might listen to music, be interrupted, or be checking their social media accounts. Online experiments allow participants the flexibility to create an environment that closely resembles their everyday study habits.

Not only are laboratory experiments typically restricted to a single location, but they are also typically restricted to daytime office hours. However, in this experiment, access to the sessions was available around the clock – whenever was convenient for participants. Studying in the evening or at night is a common habit for adult learners and research has supported that, for some people, studying during this time can lead to greater retention of information (Evans, Kelley, & Kelley, 2017). Providing participants with this freedom mimics real-life study situation in which students study during the hours that work best for them.

Another issue which arises when researchers restrict the physical location and time of day people can participate, as one would during an in-person laboratory experiment, is the risk of an overly narrow sample. Students who are working, have busy course schedules, or take courses online may be less likely to be available during specified laboratory hours. This online experiment contributed to the self-explanation literature by recruiting more broadly than previous studies.

One drawback of conducting an online experiment is that the anonymity of the participants could result in one participant submitting multiple responses into the system by creating multiple profiles. Although online experiment platforms take measures to prevent people from creating multiple accounts (Palan & Schitter, 2018), it is a possibility. For this experiment, participant IP addresses were tracked, however it was difficult to assess whether the same participant was logging back into the system, or two different people (e.g., roommates) were participating from the same location. Even if some participants were logging repeatedly into the system, the issue would mostly likely be small. One experiment with *MTurk* workers found that only less than 3% of participants behaved in a manner suggesting more than one account (Chandler et al., 2014). This experiment was also designed to be multi-session, which might deter some multi-account holders from participating.

Chapter 4. General Discussion and Conclusions

In this final chapter of the dissertation, I discuss the implications of the self-explanation meta-analysis (Chapter 2) and experiment (Chapter 3). The first section of the chapter considers the theoretical and practical contributions of the dissertation. The second section discusses the studies' limitations and possible future research avenues. Finally, in the last section, I offer concluding thoughts about how my work contributes to theoretical understanding of the self-explanation effect by supporting the imperfect mental model theory.

4.1 Theoretical and Practical Contributions

The results of the two studies reported in this dissertation confirmed and extended the theoretical underpinnings of the cognitive processes and utility of self-explanation.

4.1.1 Self-Explanation Meta-analysis

In the meta-analysis presented in Chapter 2, introducing an inducement to self-explain led to a moderate effect size. 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 0.45 to 0.65 favoring participants who were prompted or directed to self-explain ($p < .001$). With the exception of a few categories represented by a small number of studies, inducements to self-explain were associated with statistically detectable benefits.

Our findings have significant practical implications. The foremost is that having learners generate an explanation is usually 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.

Only two moderator variables showed statistically detectable effects that have implications for both theory and practice. Inducements to self-explain were associated with statistically detectable benefits in the two cases: (i) the comparison treatment of providing an instructional explanation; and (ii) the type of self-explanation elicited.

4.1.1.1 Self-Explanation Outperforms Instructional Explanation

Contrary to the coverage hypothesis, 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 opportunity to create an idiosyncratic match to prior knowledge of the new knowledge generated by self-explaining. We hypothesized that by retrieving relevant previously learned information from memory and elaborating it with relevant features in the new information, 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 produced by information that could be made available in 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 participants who received only SEPs outperformed participants who received SEPs 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.

4.1.1.2 How Self-Explanation is Elicited

The meta-analysis investigated three moderators of self-explanation inducement: inducement type, type of self-explanation elicited, and content specificity. One finding was that studies in which participants were prompted to give anticipatory self-explanations (g = 1.37) found substantially larger effects than those in which participants were asked to justify their own decisions (*g* = .42) or the decisions of another (*g* = .43). Participants gave anticipatory self-explanations after being presented with incomplete information and were subsequently given opportunity to obtain further information and revise their self-explanations. For example, de Bruin, Rickers and Schmidt (2007) asked participants to predict the next move of an expert chess program and explain why they believed the predicted move was the best. If their prediction turned out to incorrect, they were asked to revise their explanation. Participants in a comparison condition made the same prediction and received feedback but were not prompted to explain their prediction or revise their explanation. This anticipatory self-explanation plus correction paradigm may be highly effective because the original explanation is activated in working memory when the learner receives negative feedback and begins generating a revised explanation. The failed explanation and the feedback (i.e., the move actually made by the chess program) jointly cue effective revisions to the explanation.

Another finding about how self-explanation inducement boosts learning outcomes focused on content specificity. Effectively, we used three codes: “Specific”, “Generic”, and “Both”. A “Specific” code meant that the inducement contained content-specific information which may help learners engage in sufficient self-explanation as it signals suitable content to explain. On the other hand, “Generic” meant that the inducement did not contain content, allowing learners to apply their own criteria to decide what to self-explain which may support constructing explanations that are better adapted to prior knowledge. Lastly, if both types of inducements were used, then we coded the study as “Both”.

We found that the studies which used “Specific” inducements had a mean effect size of $g = .51$, $p < .05$ whereas the studies which used “Generic” inducements had a mean effect size of $g = .68$, $p < .05$. We only identified 6 papers which included both “Specific” and “Generic” SEPs and in four of these studies, all participants were given both types of prompts.

4.1.2 Self-Explanation Study

Although the meta-analysis examined how the cue strength associated with different forms of eliciting self-explanation might affect learning outcomes, many future avenues of investigation were identified because the prompts used to elicit self-explanations have been quite varied in the literature (Dunlosky et al., 2013). One such thread was whether learners who receive generic SEPs will outperform learners who receive specific SEPs and/or no self-explanation.

Chapter 3 described a study which examined the effects on reading comprehension of receiving: (i) specific SEPS; (ii) generic SEPS; and (iii) no SEPs. There were two competing hypotheses: (1) Participants who receive specific SEPs will perform better because these prompts cue participants to focus on certain elements of the text, which will lead them to remember more during the free recall task. (2) Participants who receive generic SEPs will perform better because these types of prompts allow participants to fill-in their own knowledge gaps.

The results support the hypotheses that generic SEPs are superior to both receiving specific SEPs and receiving no SEPs. Specifically, the group elicited to self-explain by generic SEPs outperformed the other two groups on the reading comprehension outcome in short-answer question format. Overall, these results are consistent with the conclusions from past studies that support the use of content-free self-explanations (Chou & Liang, 2009) as well the meta-analysis results described in Chapter 2.

4.1.2.1 Self-explanation and Reading Comprehension

The research reported in Chapter 3 contributes to the theoretical understanding of how self-explanations impact cognition while reading text. Self-explanation has most often been studied in the context of well-defined domains, such as Mathematics or Physics, and the learning task often involves problem solving, worked examples, or case studies (Bisra et al., 2018). Although all of these learning tasks require effort, planning, self-monitoring, selection and reflection (Winne & Hadwin, 1998), the types of strategies, generated inferences, and skills employed differ (Fuchs, Gilbert, Fuchs, Seethaler, & Brittany, 2018).

The central goal of reading is comprehension (Nation, 2005) and the negative consequences of comprehension difficulties are well documented in academic settings, career trajectories, and life satisfaction (Ricketts, 2011). Reading comprehension is a complex task involving numerous cognitive processes such as word recognition, accessing meaning, activating relevant prior knowledge, and generating inferences

(Diesen, 2014; Kintsch, 2004; Kintsch & van Dijk, 1978). Cognitively based theories of reading emphasize the interactive and constructivist nature of comprehension (Anderson, Reynolds, Schallert, & Goetz, 1977; Dole, Duffy, Roehler, & Pearson, 1991; van Dijk & Kintsch, 1993). Comprehension involves creating a mental representation, or a situation model that represent aspects of the text organized in a manner coherent with the reader's prior knowledge (Johnson-Laird, 1983; van Dijk & Kintsch, 1993).

Past studies have shown that readers who self-explain the text understand more from the text and it is assumed construct better mental models of the content (Chi & Bassok, 1989; Chi, De Leeuw, et al., 1994; Magliano et al., 1999). There are two theories which explain this result: (i) the incomplete text hypothesis; and (ii) the imperfect mental model hypothesis (see Chapter 1). The incomplete text hypothesis assumes that the text is incomplete in one manner or another (i.e. explanatorily incoherent, missing information) (Kintsch & van Dijk, 1978; Kintsch & Vipond, 1979), and that SEPs encourage learners to fill gaps in the text. The imperfect mental model view assumes that learners come to the text with different pre-existing mental models that have unique deficiencies and structures. As such, no matter how complete and coherent the text, it may still contain omissions from the reader's perspective (Chi, 2000, p. 198).

This study supports the imperfect mental model theory over the incomplete text hypothesis because the group that received content-free SEPs performed better on a reading comprehension outcome than groups that received content-specific SEPs or no SEPs. In addition to being prompted to generate inferences, the group that received specific SEPs could have been aided by providing missing links or by increasing cohesion of the text by the content embedded in the more specific prompt. If this group had outperformed the other two groups, then such result would have lent support for the incomplete text hypothesis.

However, the learners who received generic SEPs were able to generate inferences without being cued to focus on any specific content in the passage. These participants were prompted to generate inferences, unlike the group that did not receive SEPs, and they had greater freedom to fill gaps in their own mental models in comparison to the group that received specific SEPs. The finding that generic SEPs were more beneficial supports the imperfect mental model theory which describes the underlying cognitive mechanism of the self-explanation effect as one in which learners monitor gaps between their current mental model and new information presented in the text, and were then prompted to generate inferences to bridge the gaps (Chi & Bassok, 1989; Chi, Slotta, et al., 1994; Lin & Lehman, 1999; VanLehn et al., 1992).

This study also contributes to the imperfect mental model theory by adding a third consideration about integrating information while reading. If presented with a content-specific SEP, the reader must engage in additional work to incorporate the embedded content into their situation model. From the results of the study, one can infer that the participants who received specific SEPs did focus on the embedded content because they recalled the content-related propositions better than participants who received generic SEPs (p =.036). However, focusing on this content did not result in these participants outperforming the other two groups on the reading comprehension measures. Perhaps the added cognitive burden of making sense and incorporating the specific content embedded in the prompt interfered with a participant's ability to compare their mental model with the textbase, resulting in fewer inferences or incorrect inferences. Or, the specific SEPs caused a "tunneling" effect where participants only focused on the content related to the SEPs to the exclusion of other content in the passage. This left gaps in their comprehension.

4.1.2.2 Self-explanation and Real-life Study Situations

In general, learners do not spontaneously self-explain but do so when guided or prompted (Bielaczyc et al., 1995; Chi, De Leeuw, et al., 1994). As such, great attention has been payed to Intelligent Tutoring Systems or computer-supported self-explanation elicitation (Aleven, Koedinger, & Cross, 1999; Conati & Vanlehn, 2000). Although these interfaces have grown more complex over time, there is scant research on what will motivate a learner in a real-life situation to respond to a prompt. Learning strategies, be they practice testing, elaborative interrogation or self-explanation, will only be useful if students are motivated to use them appropriately (Dunlosky et al., 2013).

One criticism of prior work in the field of self-explanation has been that experimenters burdened participants by requiring them to respond to many SEPs (i.e., self-explain after reading each sentence of the text). This may demotivate learners in real-life situations as it would require a nontrivial amount of time and effort (Dunlosky et al., 2013). Students may find responding to so many SEPs tedious.

The experiment reported in Chapter 3 had only three SEPs embedded throughout the text and did not require participants to respond to each prompt. This decision was made to mimic real-life study situations in which, even if prompts are provided, learners must choose how much effort to put towards any prompt. One practical contribution this study is to demonstrate that even when not required to complete a SEP, learners will participate in self-explanation when they are trained to do so. In such cases, only a few prompts are necessary to promote learning gains.

Also, this is one of the few to examine self-explanation using only online participation. Online participation more closely mimics real-life study situations for learners in terms of environment (i.e., location, time of day) and study conditions (see Section 3.10). As such, it contributes to the body of literature on self-explanation by widening the ecological scope of the effect.

Bisra et. al (2018) found that the majority of studies on self-explanation recruited undergraduate students (61%) or school-aged participants (33%). This experiment adds to the literature as it is one of the few studies in which participants were older, on average almost 26 years old. Also, 56% had least one post-secondary degree. Although the results of this experiment cannot be generalized reliably to less educated samples, self-explanation is a study strategy that could be utilized by adults enrolled in higher education. As such, this sample does closely match one important target population which boosts generalizing to similar contexts (Shadish, Cook, & Campbell, 2001)

4.1.3 Self-explanation Usually Takes More Time. Does That Matter?

One issue which resurfaces in the self-explanation literature is whether the time demanded to generate self-explanations explain the learning outcome differences between groups which were prompted to self-explain and groups which were not. In order words, is the self-explanation effect due to cognitive differences between the groups or just a remnant of learners spending more time on the task.

Dunlosky et al. (2013) provided a brief overview of conflicting reports on this issue and concluded that further research was needed before a conclusion can be made. For example, they highlighted research by Chi et al. (1994) in which the self-explanation group spent almost twice the time as the rereading group (125 vs. 66

minutes), but then also point out that Ainsworth and Burcham (2007) reported the self-explanation group's performance was stronger even when controlling for time-on-task.

Our meta-analysis examined this issue empirically. 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* = .36), the difference was not statistically detectable. Our study also investigated whether time spent on task varied between the self-explanation groups and the control group. A statistically detectable difference was found for reading task order Group BA participants but not for Group AB. This mixed result is consistent with the literature as reported by Dunlosky et al. (2013).

We interpret these results as showing that inducing self-explanation is a time-efficient instructional method but, in some research, its efficiency may be overestimated because time-on-task was not controlled or reported. Most research on self-explanation has, regrettably, not fully reported time-on-task. We advocate for reporting the mean and standard deviation of learning task duration for all treatment groups in self-explanation research.

4.2 Limitations & Future Research

4.2.1 Self-Explanation Meta-analysis

The self-explanation meta-analysis has some limitations. While evidence was found to support certain factors that appear to moderate the overall effect of eliciting self-explanations, the need to refine our sense of these moderating effects through future investigations is significant. For example, more studies are needed on how the number and directedness of an SEP might affect learning outcomes.

Furthermore, most of the studies reviewed variation in the effects of SEPs depending on whether they were interspersed with the learning material (*g* = .52, 57 studies), presented at the beginning of the learning activity (*g* = 1.24, 4 studies), or presented after the participants had studied the learning material (*g* = .57, 8 studies). Although inconclusive, these results may suggest that the optimal timing of self-explanation inducement may be before the learning activity. Perhaps when a learner self-explains prior to engaging in a learning activity, she is primed to better access her prior knowledge or confront her misconceptions.

Within the field, there is a limited understanding of long-term impacts of self-explanation training. Most of the studies on self-explanation were conducted with the duration of learning activity taking less than 60 minutes and focus on only one learning activity. One of the strengths of self-explanation literature is that an effect was shown across subject areas and learning outcomes. Investigations focusing on the long-term impacts of self-explanation training, perhaps across subject domains and tracked over time, would greatly enhance the understanding of self-explanation as a cognitive process.

Finally, although reports of effect sizes in meta-analytical studies increase statistical power and provide a robust point estimate greater than any individual study (Cohn & Decker, 2003), they reflect the research quality of the primary studies. Feinstein (1995) described meta-analysis as “statistical alchemy” that attempts to turn something not so valuable, poor quality research, into gold. Ideally, to avoid aggregating results from flawed research methods, we would have excluded studies based on a research quality variable. However, we faced the same challenge as others: How does a meta-analyst evaluate research quality?

Our approach was to code four study quality variables and examine whether the results differed between poorly and well-designed studies. Categories include methodological quality (i.e., treatment fidelity) and participant-related factors (i.e., random assignment) (Sharpe, 1997). This approach aligns with the meta-analytic tradition of not excluding studies while accounting for research quality. However, there are more than four research quality-related variables (see Cook and Campbell, 1979)), and the reason we did not code for more was because few primary studies provide sufficiently detailed information about methodological variables.

For example, van Balkom et al., (1994) conducted a meta-analysis in which their coding form consisted of 49 items about treatment design, treatment integrity, internal validity, and construct validity. Few of their primary studies provided sufficient information to permit coding on all 49 items. As well, several attempts have been made to combine study features to create an overall quality index, however none of them have produced coding that is related to effect size (Weaver & Clum, 1995; Weiss & Weisz, 1995).

As such, one limitation of the meta-analysis is the possibility of including or inappropriately grouping low-quality studies with high quality studies. A summary effect size would have been calculated no matter the research quality of the primary studies. A meta-analysis answers research questions with “numbers” which, to some, may appear to be definitive. It is important to recognize that the conclusion of a meta-analytic review must be tempered by the quality of the empirical research comprising it (Little, 2013, p.18). The strength of meta-analyses lies in empirical generalizations of associations between classes of treatments and outcomes. Readers who focus on the summary effect and ignore heterogeneity are missing the major promise of research synthesis. Future primary research reports should include a greater description of research methodology and document step-by-step processes.

4.2.2 Self-Explanation Study

Although this work demonstrated that generic SEPs aid learners in reading comprehension better than specific SEPs, the context for this finding is limited. I had hypothesized that participants with lower reading comprehension abilities would have performed better with specific SEPs than generic SEPs. My reasoning was that learners with lower reading comprehension abilities might use the content in the SEPs to direct their attention in a beneficial manner.

However, because the sample was so varied, it was difficult to use any prior indicators (i.e. some participants reported a high school GPA while others reported a graduate degree GPA) as covariates for reading comprehension ability. Another method for conducting an analysis on reading ability and SEP type could have involved focusing on using reading comprehension scores from the first session as a moderator. Unfortunately, the research design necessitated using two reading task orders to control for differences in the two within-subject reading materials. A precise investigation into ability and SEP type would benefit the literature on self-explanation.

Although the participants in this study were mostly from WEIRD (Western, Educated, Industrialized, Rich and Democratic) countries, the geographic locations did span three large countries (Canada, USA, UK). This is a broader sample than much of the work previous published on self-explanation effects in which participants were selected from the same school (Hausmann & VanLehn, 2010b; Kramarski & Dudai, 2009; Yeh, Chen, Hung, & Hwang, 2010). I hypothesize that the reason the recruits were from mostly WEIRD nations is because one of the main restrictions placed upon the sample was ability to write and read in English. Future work could create several language versions of the experimental materials and actively recruit on platforms more popular in Asia, Africa, and other locations.

Research on self-explanation has achieved a stage where the general efficacy of self-explanation has been established. There is now a need for clearer mapping of 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. Another goal of future research should be to identify strategies whereby prompts are faded away so that self-explanation becomes fully self-regulated.

Current self-explanation literature tends to focus on problem solving, transfer, and inference as learning outcome measures with test formats focused on problems and short-answer questions (Bisra et al., 2018). There are few studies which use free recall as an outcome measure (Larsen, Butler, & Roediger, 2013), but the research on the effect of self-explanation on memory-related tasks is sparse. Perhaps one future area of research could concentrate on how to embed SEPs in text to bolster text retention.

Another issue impacting self-explanation is how the cue strength associated with different forms of self-explanation elicitation might affect learning outcomes. In our meta-analysis, multiple-choice prompts were associated with the smallest mean effect size (g = .24). Perhaps because of the small number of studies using multiple-choice prompts (k = 2) this result did not differ statistically from other prompt types. To address the question of optimal cue strength more thoroughly, we recommend 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 SEPs; as their knowledge increases, they may benefit more from less strongly cued prompts.

Although there are benefits and disadvantages of conducting online research, I encourage researchers to consider the changing landscape of learning technology when designing research projects. Between the ubiquity of smartphones in the classroom (Marinagi, Skourlas, & Belsis, 2013) and proliferation of online, blended, or massive open online courses (Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014; Swan, 2002; Willging & Johnson, 2009), students now expect to be able to access content and learn from anywhere (Roberts, Newman, & Schwartzstein, 2012). Although when cell phones first began to grow in popularity, they were seen as a disruption by instructors (Thomas, O'bannon, & Bolton, 2013), there is now a growing body of literature on the pedagogical benefits of cell phones (Daher, 2010; Thomas et al., 2013). In a similar vein, even if researchers view online experimentation as a disruption, its use will continue to grow. With this growth, there will be changing expectations on both the part of researchers and participants. Issues related to the ecological validity of laboratory experiments will emerge as students natural learning environments keep diverging from the commonly used computer-based laboratory. As well, learners may prefer to participate in online experiments as opposed to a physical laboratory.

By no means do I suggest that laboratory experiments, which offer a controlled environment, should be eschewed all together. Rather, I suggest that a variety of experimental modalities, each of which has distinctive shortcomings, be utilized in experimental designs in a complementary manner. Accordingly, perhaps a future research project could repeat the experiment described in Chapter 3 in a face-to-face setting. Not only would complementary experimentation shed more light on self-explanation, but it might also indicate how evidence of learning processes differs across research modalities.

4.3 Conclusion

Despite its limitations, this dissertation presents several important contributions to the literature on self-explanation. First, it includes a meta-analysis that investigates the cognitive effects and instructional effectiveness of self-explanation interventions. The findings of this meta-analysis have practical implications, namely that learner generated explanations are generally more effective than when learners are provided with an explanation. Another implication from the meta-analysis for teachers is that inducing self-explanations is beneficial across age groups, subject areas, and learning outcomes. Second, this dissertation reports on an experiment that investigated the learning effects of two types of SEPs. A practical implication of this study is to confirm that in some learning scenarios, including reading comprehension, content-free SEPs can offer benefit compared to content-specific SEPs and receiving no SEPs. The breadth of the learning scenarios where this hypothesis prevails is an important open question. This study also contributes to the theoretical understanding of the self-explanation effect by supporting the imperfect mental model theory. The group that received content-free SEPs had the greatest freedom to revise their mental models by filling in gaps or making changes also performed better than the other two groups.

In conclusion, this dissertation provides significant support for the continued study of how eliciting self-explanations can benefit learning. The results from this work can guide instructional designers, instructors, and students in how to foster and adopt self-explanation as a successful learning strategy.

References

Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*, *69*(2), 334–345. https://doi.org/10.1037/0021-9010.69.2.334

Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self-explanation. *Learning and Instruction*, *17*(3), 286–303. https://doi.org/10.1016/j.learninstruc.2007.02.004

Aleven, V., & Koedinger, K. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*. https://doi.org/10.1016/S0364-0213(02)00061-7

Aleven, V., Koedinger, K. R., & Cross, K. (1999). Tutoring Answer Explanation Fosters Learning with Understanding. In *AIED '99, 9th World Conference of Artificial Intelligence and Education,*. Le Mans, France.

Amir, O., Rand, D. G., & Gal, Y. K. (2012). Economic games on the internet: The effect of $1 stakes. *PLoS ONE*, *7*(2). https://doi.org/10.1371/journal.pone.0031461

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with Massive Online Courses. In C. Chung (Ed.), *23rd International Conference on the World Wide Web*. Seoul.

Anderson, R., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. (1977). Frameworks for Comprehending Discourse. *American Educational Research Journal*, *14*(4), 367–381. https://doi.org/10.3102/00028312014004367

Arnett, J. J. (2008). The Neglected 95%: Why American Psychology Needs to Become Less American. *American Psychologist*, *63*(7), 602–614. https://doi.org/10.1037/0003-066X.63.7.602

Aronson, E., Brewer, M., & Carlsmith, J. (1985). Experimentation in Social Psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology* (3rd ed.). New

York, NY: Random House.

Aronson, E., & Carlsmith, J. (1968). Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology* (Vol 2). MA: Addison-Wesley.

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* (Vol. 70). https://doi.org/10.3102/00346543070002181

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. https://doi.org/10.1037/0022-0663.95.4.774

Autumn, L., & Backhaus, M. (2009). Socioeconomic Status and Adjustment to College. *Higher Education*. Retrieved from https://www.apa.org/pi/ses/resources/indicator/2009/04/college

Azar, B. (2000). A web of research: They're fun, they're fast, and they save money, but do Web experiments yield quality results? *Monitor on Psychology*, *31*, 42–47.

Baddeley, A., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1

Bauman, C. W., Peter McGraw, A., Bartels, D. M., & Warren, C. (2014). Revisiting External Validity: Concerns about Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology but we worry that the judgment and decision-making processes people use in these unusual situations. *Social and Personality Psychology Compass*, *8*(9). https://doi.org/10.1111/spc3.12131

Benndorf, V., Moellers, C., & Normann, H.-T. (2017). Experienced vs. inexperienced participants in the lab: do they behave differently? *Journal of the Economic Science Association*, *3*(1), 12–25. https://doi.org/10.1007/s40881-017-0036-z

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, *37*(4), 345–363. https://doi.org/10.1007/s11251-008-9051-z

Berthold, K., & Renkl, A. (2009). Instructional Aids to Support a Conceptual Understanding of Multiple Representations. *Journal of Educational Psychology*, *101*(1), 70–87. https://doi.org/10.1037/a0013247

Bezuijen, X. M., van den Berg, P. T., van Dam, K., & Thierry, H. (2009). Pygmalion and Employee Learning: The Role of Leader Behaviors. *Journal of Management*, *35*(5), 1248–1267. https://doi.org/10.1177/0149206308329966

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in Self-Explanation and Self-Regulation Strategies: Investigating the Effects of Knowledge Acquisition Activities on Problem Solving. *Cognition and Instruction*. https://doi.org/10.1207/s1532690xci1302_3

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing Self-Explanation: a Meta-Analysis. *Educational Psychology Review*. https://doi.org/10.1007/s10648-018-9434-x

Bodvarsson, M. (2005). *Prompting students to justify their response while problem-solving: A nested mixed-methods study,*. University of Nebraska.

Bohannon, J. (2016, June 10). Mechanical Turk upends social sciences. *Science*. American Association for the Advancement of Science. https://doi.org/10.1126/science.352.6291.1263

Bonner, S. E., Hastie, R., Sprinkle, G. B., & Young, S. M. (2000). A Review of the Effects of Financial Incentives on Performance in Laboratory Tasks: Implications for Management Accounting. *Journal of Management Accounting Research*, *12*(1), 19–64. https://doi.org/10.2308/jmar.2000.12.1.19

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. https://doi.org/10.1111/j.2044-8279.2011.02041.x

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. https://doi.org/10.1016/j.learninstruc.2012.11.002

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. John Wiley & Sons. Retrieved from https://www.wiley.com/en-us/Introduction+to+Meta+Analysis-p-9780470057247

Brophy, J., & Good, T. (1974). *Teacher-student relationships: Causes and consequences.* New York: Holt, Rinehart & Wilson. Retrieved from https://psycnet.apa.org/record/1974-33079-000

Buchanan, T. (2000). Potential of the Internet for Personality Research. In *Psychological Experiments on the Internet* (pp. 121–140). Elsevier. https://doi.org/10.1016/b978-012099980-4/50006-x

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, *6*, 3–5. https://doi.org/10.1177/1745691610393980

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, *65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Canada, S. (2019). 2016 Census Data tables. Retrieved November 22, 2019, from https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dt-td/Rp-eng.cfm?LANG=E&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GID=0&GK=0&GRP=1&PID=109525&PRID=0&PTYPE=109445&S=0&SHOWALL=0&SUB=0&Temporal=2016&THEME=115&VID=0&VNAMEE=&VNAMEF=

Carey, S. (1999). Knowledge acquisition: Enrichment or conceptual change. In E. Margolis & S. Laurence (Eds.), *Concepts: core readings* (pp. 459–487). Cambridge, Massachusetts: MIT Press.

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using

Spacing to Enhance Diverse Forms of Learning: Review of Recent Research and Implications for Instruction. *Educational Psychology Review*, *24*, 369–378.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156–2160. https://doi.org/10.1016/j.chb.2013.05.009

Catrambone, R. (1996). Generalizing solution principles learned from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(4), 1020–1031.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112–130. https://doi.org/10.3758/s13428-013-0365-7

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, *8*(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2

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, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H. (2005). Commonsense Conceptions of Emergent Processes: Why Some Misconceptions Are Robust. *Journal of the Learning Sciences*, *14*(2), 161–199. https://doi.org/10.1207/s15327809jls1402_1

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. 61–81). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., & Bassok, M. (1989). Learning from Examples via Self-explanations. In B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 251–285). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

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* (Vol. 13). Retrieved from https://pdfs.semanticscholar.org/9f4b/eefc9b2a0eb8b92b4b79a878d39163c0d428.pdf

Chi, M. T. H., De Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439–477. https://doi.org/10.1016/0364-0213(94)90016-7

Chi, M. T. H., & Roscoe, R. (2002). The processes and challenges of conceptual change. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 3–27). Dordrecht, Netherlands: Kluwer Academic Publishers.

Chi, M. T. H., Slotta, J. D., & De Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, *4*(1), 27–43. https://doi.org/10.1016/0959-4752(94)90017-5

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

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. https://doi.org/10.1007/s12564-011-9187-4

Chou, C. Y., & Liang, H. T. (2009). Content-free computer supports for self-explaining: Modifiable typing interface and prompting. *Educational Technology and Society*, *12*(1), 121–133.

Christensen, L. (2001). *Experimental methodology* (8th ed.). Boston, MA: Allyn & Bacon.

Chu, J., Rittle-Johnson, B., & Fyfe, E. R. (2017). Diagrams Benefit Symbolic Problem Solving. *British Journal of Educational Psychology*, *87*(2).

Chung, S., Chung, M., & Severance, C. (1999). Design of support tools and knowledge building in a virtual university course: Effect of reflection and self-explanation prompts. In *WebNet 99 World Conference on the WWW and Internet Proceedings*. Honolulu, Hawaii.

Chung, S., Severance, C., & Moon-Jung, C. (2003). Design of Support Tools for Knowledge Building in a Virtual University Course. *Interactive Learning Environments*, *11*(1), 41–57.

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching*, *30*(10), 1241–1257. https://doi.org/10.1002/tea.3660301007

Cohen, J. R., Asarnow, R. F., Sabb, F. W., Bilder, R. M., Bookheimer, S. Y., Knowlton, B. J., & Poldrack, R. A. (2010). A unique adolescent response to reward prediction errors. *Nature Neuroscience*, *13*(6), 669–671. https://doi.org/10.1038/nn.2558

Cohn, L. D., & Decker, B. J. (2003). How Meta-Analysis Increases Statistical Power. *Psychological Methods*, *8*(3), 243–253. https://doi.org/10.1037/1082-989X.8.3.243

Conati, C., & Vanlehn, K. (2000a). Toward Computer-Based Support of Meta-Cognitive Skills: a Computational Framework to Coach Self-Explanation. *International Journal of Artificial Intelligence in Education*, (11), 398–415. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.1243

Conati, C., & Vanlehn, K. (2000b). Toward Computer-Based Support of Meta-Cognitive Skills: A computational Framework to Coach Self-explanation. *International Journal of Artificial Intelligence in Education*, (11). Retrieved from https://telearn.archives-ouvertes.fr/hal-00197335

Cook, T., & Campbell, D. (1979). *Quasi-experimental experimentation: Designs and analysis for field settings*. Boston: Houghton Mifflin.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

Daher, W. (2010). Building mathematical knowledge in an authentic mobile phone environment. *Australasian Journal of Educational Technology*, *26*(1), 85–104. https://doi.org/10.14742/ajet.1104

Davidov, E., & Depner, F. (2011). Testing for measurement equivalence of human values across online and paper-and-pencil surveys. *Quality and Quantity*, *45*(2), 375–390. https://doi.org/10.1007/s11135-009-9297-9

de Bruin, A. B. H., Rikers, R. M. J. P., & 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. https://doi.org/10.1016/j.cedpsych.2006.01.001

DeLeeuw, N., & Chi, M. T. H. (2003). Self-explanation: Enriching a situation model or repairing a domain model? In G. Sinatra & P. R. Pintrich (Eds.), *Intentional Conceptual Change* (pp. 55–78). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Diesen, D. (2014). *Situation model: A framework to study reading comprehension. Unpublished Master's Thesis*. Macquarie University.

Dole, J. A., Duffy, G. G., Roehler, L. R., & Pearson, P. D. (1991). *Moving From the Old to the New: Research on Reading Comprehension Instruction* (Vol. 61).

Drori, J. (2011). *The beautiful tricks of flowers*. Retrieved from https://www.ted.com/talks/jonathan_drori_the_beautiful_tricks_of_flowers?language=en

Dunlosky, J. (2013). Strengthening the Student Toolbox study strategies to Boost learning. *American Educator*, *37*(3), 12–21.

Dunlosky, J., Rawson, K., Marsh, E., Nathan, M., & Willingham, D. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, Supplement*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

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. https://doi.org/10.1016/j.compedu.2005.11.010

Eaton, J., Anderson, C., & Smith, E. (1983). When Students Don't Know They Don't Know. *Science and Children*, *20*(7), 6–9. Retrieved from https://eric.ed.gov/?id=EJ280026

Eaton, J., Anderson, C. W., & Smith, E. L. (1984). Students' Misconceptions Interfere with Science Learning: Case Studies of Fifth-Grade Students. *The Elementary School Journal*, *84*, 365–379. https://doi.org/10.2307/1001364

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583–610. https://doi.org/10.1007/s10683-011-9283-7

Eriksson, K., & Simpson, B. (2010). Emotional reactions to losing explain gender differences in entering a risky lottery. *Judgment and Decision Making*, *5*(3), 159–163.

Evans, M. D. R., Kelley, P., & Kelley, J. (2017). Identifying the Best Times for Cognitive Functioning Using New Methods: Matching University Times to Undergraduate Chronotypes. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00188

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994. https://doi.org/10.1257/aer.90.4.980

Feinstein, A. R. (1995). Meta-analysis: Statistical alchemy for the 21st century. *Journal of Clinical Epidemiology*, *48*(1), 71–79. https://doi.org/10.1016/0895-4356(94)00110-C

Ferguson-Hessler, M. G. M., & de Jong, T. (1990). Studying Physics Texts: Differences in Study Processes Between Good and Poor Performers. *Cognition and Instruction*. https://doi.org/10.1207/s1532690xci0701_2

Finley, A., & Penningroth, S. (2015). Online versus in-lab: pros and cons of an online prospective memory experiment. In A. Columbus (Ed.), *Advances in Psychology Research* (pp. 135–162). Hauppauge, NY: Nova Science Publishers, Inc.

Fisher, K. M. (1985). A misconception in biology: Amino acids and translation. *Journal of Research in Science Teaching*, *22*(1), 53–62. https://doi.org/10.1002/tea.3660220105

Fonseca, B., & Chi, M. (2011). Instruction based on self-explanation. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of Research on Learning and Instruction* (pp. 296–321). New York. Retrieved from https://books.google.ca/books?hl=en&lr=&id=PDNZBwAAQBAJ&oi=fnd&pg=PA296&dq=fonseca+and+chi+2011&ots=pt-YLvjSC9&sig=nLsJCyjXSydhSb1MtvMYc8ffJRk&redir_esc=y#v=onepage&q=fonseca and chi 2011&f=false

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., Cees, &, & Van Der Vleuten, P. M. (2012). Rethinking the globalisation of problem-based learning: how culture challenges self-directed learning. *Medical Education*, *46*, 738–747. https://doi.org/10.1111/j.1365-2923.2012.04290.x

Fuchs, L. S., Gilbert, J. K., Fuchs, D., Seethaler, P. M., & Brittany, B. L. (2018). Text Comprehension and Oral Language as Predictors of Word-Problem Solving: Insights into Word-Problem Solving as a Form of Text Comprehension. *Scientific Studies of Reading*, *22*(2), 152–166. https://doi.org/10.1080/10888438.2017.1398259

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. https://doi.org/10.1016/j.learninstruc.2011.06.002

Garnham, A., & Oakhill, J. (1987). Interpreting Elliptical Verb Phrases. *The Quarterly*

Journal of Experimental Psychology Section A, *39*(4), 611–627. https://doi.org/10.1080/14640748708401805

Gentner, D., & Stevens, A. L. (1983). *Mental models*. Erlbaum.

Gestsdottir, S., & Lerner, R. M. (2008). Positive Development in Adolescence: The Development and Role of Intentional Self-Regulation. *Human Development*, *51*, 202–224. https://doi.org/10.1159/000135757

Goldberg, M. E., & Gorn, G. J. (1987). Happy and Sad TV Programs: How They Affect Reactions to Commercials. *Journal of Consumer Research*, *14*(3), 387. https://doi.org/10.1086/209122

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*, 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. *American Psychologist*, *59*(2), 93–104. https://doi.org/10.1037/0003-066X.59.2.93

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395. https://doi.org/10.1037/0033-295X.101.3.371

Graf, E. (2000). *Designing a computer tutorial to correct a common student misconception in mathematics.* University of Washington.

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 and Cognition*, *36*(1), 93–103. https://doi.org/10.3758/MC.36.1.93

Hanel, P. H. P., & Vione, K. C. (2016). Do Student Samples Provide an Accurate Estimate of the General Public? https://doi.org/10.1371/journal.pone.0168354

Harrison, G., & List, J. (2004). Field Experiments. *Journal of Economic Literature*, *42*(4), 1099–1055.

Hartman, H. (2001). Developing students' metacognitive knowledge and skills. In *Metacognition in learning and instruction* (pp. 33–68). Springer Netherlands.

Hartman, H., Everson, H., Toblas, S., & Gourgey, A. (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 over 800+ meta-analyses relating to achievement*. Abingdon: Routledge. Retrieved from http://people.oregonstate.edu/~flayb/MY COURSES/H676 Meta-Analysis Fall2016/Examples of SRs & MAs of interventions/Hattie2000 Synthesis of over 800 meta-analyses relating to achievement - first 29 pages.pdf

Hausmann, R. G. M., & VanLehn, K. (2007). Explaining self-explaining: A contrast between content and generation. *Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, 417–424.

Hausmann, R. G. M., & VanLehn, K. (2010a). The effect of self-explaining on robust learning. *International Journal of Artificial Intelligence in Education*, *20*(4), 303–332. https://doi.org/10.3233/JAI-2010-010

Hausmann, R. G. M., & VanLehn, K. (2010b). The effect of self-explaining on robust learning. *International Journal of Artificial Intelligence in Education*, *20*(4), 303–332. https://doi.org/10.3233/JAI-2010-010

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press. Retrieved from https://idostatistics.com/hedges-olkin-1985-statistical-methods-for-meta-analysis/

Henrich, J., Heine, S. J., & Norenzayan, A. (2010, July 1). Most people are not WEIRD. *Nature*. https://doi.org/10.1038/466029a

Henry, P. J. (2008). College sophomores in the laboratory redux: Influences of a narrow data base on social psychology's view of the nature of prejudice. *Psychological Inquiry*, *19*(2), 49–71. https://doi.org/10.1080/10478400802049936

Hmelo-Silver, C. E. (2004). Problem-Based Learning: What and How Do Students Learn? *Educational Psychology Review*, *16*(3), 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, *14*(3), 399–425. https://doi.org/10.1007/s10683-011-9273-9

Johnson-Laird, P. N. (Philip N. (1983). *Mental models : towards a cognitive science of language, inference, and consciousness*. Harvard University Press. Retrieved from https://books.google.ca/books/about/Mental_Models.html?id=FS3zSKAfLGMC&redir_esc=y

Johnson-Laird, P. N. (Philip N. (1995). Inference and Mental Models. In S. Newstead & J. Evans (Eds.), *Perspectives on thinking and reasoning: Essays in honour of Peter Wason* (pp. 115–146). Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson, C. I., & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. *Computers in Human Behavior*, *26*(6), 1246–1252. https://doi.org/10.1016/j.chb.2010.03.025

Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R. Ruddell & N. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1270–1328). Newark: International Reading Association.

Kintsch, Walter, & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Kintsch, Walter, & Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L. Nilsson (Ed.), *Perspectives on Memory Research* (pp. 329–365). Retrieved from https://philpapers.org/rec/KINRCA-3

Kopp, C. B. (1982). Antecedents of self-regulation: A developmental perspective. *Developmental Psychology*, *18*(2), 199–214. https://doi.org/10.1037/0012-1649.18.2.199

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. https://doi.org/10.2190/ec.40.4.a

Kramarski, B., Weiss, I., & Sharon, S. (2013). Generic Versus Context-Specific Prompts for Supporting Self-Regulation in Mathematical Problem Solving Among Students With Low or High Prior Knowledge. *Journal of Cognitive Education and Psychology*, *12*(2), 197–214. https://doi.org/10.1891/1945-8959.12.2.197

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist*, *53*(9), 1017–1031. Retrieved from https://psycnet.apa.org/buy/1998-10886-001

Kwon, K., & Jonassen, D. H. (2011). The Influence of Reflective Self-Explanations on Problem-Solving Performance. *Journal of Educational Computing Research*, *44*(3), 247–263. https://doi.org/10.2190/ec.44.3.a

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 Www.Ncolr.Org/Jiol*, *10*(2). Retrieved from www.ncolr.org/jiol

Larsen, D. P., Butler, A. C., & Roediger, H. L. (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical Education*, *47*(7), 674–682. https://doi.org/10.1111/medu.12141

Lebel, C., & Beaulieu, C. (2011). Longitudinal development of human brain wiring continues from childhood into adulthood. *Journal of Neuroscience*, *31*(30), 10937–10947. https://doi.org/10.1523/JNEUROSCI.5302-10.2011

Levitt, S. D., & List, J. A. (2007). *What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World? The Journal of Economic Perspectives* (Vol. 21).

Lin, L., & Atkinson, R. K. (2013). Enhancing Learning from Different Visualizations by Self-Explanation Prompts. *Journal of Educational Computing Research*, *49*(1), 83–110. https://doi.org/10.2190/EC.49.1.d

Lin, X., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching*, *36*(7), 837–858. https://doi.org/10.1002/(SICI)1098-2736(199909)36:7<837::AID-TEA6>3.0.CO;2-U

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. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.396.9886

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage. Retrieved from https://idostatistics.com/lipsey-wilson-2001-practical-meta-analysis-2001/

List, J. (2006). The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions. *Journal of Political Economy*, *114*.

Little, T. D. (2013). *The Oxford Handbook of Quantitative Methods, Volume 1 : Foundations.* Oxford University Press, USA.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent Tutoring Systems and Learning Outcomes: A Meta-Analysis. https://doi.org/10.1037/a0037123.supp

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic Processing during Comprehension. *Journal of Educational Psychology*. https://doi.org/10.1037/0022-0663.91.4.615

Marambe, K. N., Vermunt, J. D., & Boshuizen, H. P. A. (2012). A cross-cultural comparison of student learning patterns in higher education. *Higher Education*, *64*(3), 299–316. https://doi.org/10.1007/s10734-011-9494-z

Marinagi, C., Skourlas, C., & Belsis, P. (2013). The 2nd International Conference on Integrated Information Employing ubiquitous computing devices and technologies in the higher education classroom of the future. *Procedia-Social and Behavioral Sciences*, *73*, 487–494. https://doi.org/10.1016/j.sbspro.2013.02.081

Martin, D. (1996). *Doing psychology experiments* (4th ed.). Pacific Grove, CA: Brooks/Cole.

Mayer, R. (1997). Multimedia Learning: Are We Asking the Right Questions? *Educational Psychologist*, *32*(1), 1–19. Retrieved from http://www.uky.edu/~gmswan3/544/mayer_1997.pdf

Mayer, R., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, *93*(1), 187–198. https://doi.org/10.1037/0022-0663.93.1.187

McCambridge, J., de Bruin, M., & Witton, J. (2012). The Effects of Demand Characteristics on Research Participant Behaviours in Non-Laboratory Settings: A Systematic Review. *PLoS ONE*, *7*(6), e39116. https://doi.org/10.1371/journal.pone.0039116

McCarney, R., Warner, J., Iliffe, S., Van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne Effect: A randomised, controlled trial. *BMC Medical Research Methodology*, *7*. https://doi.org/10.1186/1471-2288-7-30

Mceldoon, K. L., Durkin, K. L., & Rittle-Johnson, B. (2012). Is self-explanation worth the time? A comparison to additional practice. *British Journal of Educational Psychology*. https://doi.org/10.1111/j.2044-8279.2012.02083.x

McNamara, D. S. (2004). SERT: Self-Explanation Reading Training. *Discourse Processes*, *38*(1), 1–30. https://doi.org/10.1207/s15326950dp3801_1

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*. https://doi.org/10.1207/s1532690xci1401_1

McNemar. (1942). Opioion-attitude methodology. *Psychological Bulletin*, *43*, 289–374.

Meyers-Levy, J., Zhu, R., & Jiang, L. (2010). Context Effects from Bodily Sensations: Examining Bodily Sensations Induced by Flooring and the Moderating Role of Product Viewing Distance. *Journal of Consumer Research*, *37*(1), 1–14. https://doi.org/10.1086/649028

Nathan, M. (1994). Learning through Self-Explanation of Mathematics Examples: Effects of Cognitive Load. In *Annual Meeting of the American Educational Research Association*. New Orleans. Retrieved from https://eric.ed.gov/?id=ED372095

Nation, K. (2005). Children's reading comprehension difficulties. In *The science of reading: A handbook* (pp. 248–265). Oxford: Blackwell Publishing.

National Center for Education Statistics. (2019). *Profile of Very Low- and Low-Income Undergraduates in 2015–16*.

Necka, E. A., Cacioppo, S., Norman, G. J., & Cacioppo, J. T. (2016). Measuring the Prevalence of Problematic Respondent Behaviors among MTurk, Campus, and Community Participants. https://doi.org/10.1371/journal.pone.0157732

Nesbit, J. C., & Adesope, O. O. (2011). Learning from animated concept maps with concurrent audio narration. *Journal of Experimental Education*. https://doi.org/10.1080/00220970903292918

Nokes, T. J., Hausmann, R. G. M., VanLehn, K., & Gershman, S. (2011). Testing the instructional fit hypothesis: The case of self-explanation prompts. *Instructional Science*, *39*(5), 645–666. https://doi.org/10.1007/s11251-010-9151-4

Novak, J. D., & Cañas, A. J. (2006). The Origins of the Concept Mapping Tool and the Continuing Evolution of the Tool. *Information Visualization*, *5*(3), 175–184.

https://doi.org/10.1057/palgrave.ivs.9500126

Odilinye, L., Popowich, F., Zhang, E., Nesbit, J., & Winne, P. H. (2015). Aligning automatically generated questions to instructor goals and learner behaviour. In *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)* (pp. 216–223). IEEE. https://doi.org/10.1109/ICOSC.2015.7050809

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Orne, M. (1962). On the social psychological experiment: with particular reference to demand characteristics and their implications. *American Psychologist*, *17*(10), 776–783.

Orwin, R. G. (1983). A Fail-Safe N for Effect Size in Meta-Analysis. *Journal of Educational Statistics*, *8*(2), 157–159. https://doi.org/10.3102/10769986008002157

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon mechanical turk. *Judgment and Decision Making*, *5*(5), 411–419.

Peterson, R. A. (2001). On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-analysis. *Journal of Consumer Research*, *28*(3), 450–461. https://doi.org/10.1086/323732

Peterson, R. A., & Merunka, D. R. (2013). Convenience samples of college students and research reproducibility. https://doi.org/10.1016/j.jbusres.2013.08.010

Piaget, J. (1953). *The origin of intelligence in the child*. Retrieved from https://www.worldcat.org/title/origin-of-intelligence-in-the-child/oclc/5088448

Pintrich, P. R. (2000). The Role of Goal Orientation in Self-Regulated Learning. In

Handbook of Self-Regulation (pp. 451–502). Elsevier. https://doi.org/10.1016/B978-012109890-2/50043-3

Pirolli, P., & Recker, M. (1994). Learning Strategies and Transfer in the Domain of Programming. *Cognition and Instruction*. https://doi.org/10.1207/s1532690xci1203_2

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. https://doi.org/10.1002/sce.3730660207

Raffaelli, M., Crockett, L. J., & Shen, Y. L. (2005). Developmental stability and change in self-regulation from childhood to adolescence. *Journal of Genetic Psychology*, *166*(1), 54–76. https://doi.org/10.3200/GNTP.166.1.54-76

Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, *299*, 172–179. https://doi.org/10.1016/j.jtbi.2011.03.004

Reips, U. (2000). The web experiment method: Advantages, disadvantages, and Solutions. In M. Birnbaum (Ed.), *Psychological Experiments on the Internet* (pp. 89–117). San Diego, CA: Academic Press. Retrieved from https://web-b-ebscohost-com.proxy.lib.sfu.ca/ehost/ebookviewer/ebook/bmxlYmtfXzkxMjA1X19BTg2?sid=3aa676fc-b058-4ce7-87f8-213eefb92fda@pdc-v-sessmgr03&vid=0&format=EB&rid=1

Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*(5), 529–556. https://doi.org/10.1016/S0959-4752(01)00030-5

Renkl, Alexander. (1997). Learning from Worked-Out Examples: A Study on Individual Differences. *Cognitive Science*, *21*(1), 1–29. https://doi.org/10.1207/s15516709cog2101_1

Renkl, Alexander. (2005). The Worked-Out Examples Principle in Multimedia Learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 229–246). Cambridge: Cambridge University Press.

https://doi.org/10.1017/CBO9780511816819.016

Renkl, Alexander, Stark, R., Gruber, H., & Mandl, H. (1998). Learning from Worked-Out Examples: The Effects of Example Variability and Elicited Self-Explanations. *Contemporary Educational Psychology*. https://doi.org/10.1006/ceps.1997.0959

Rewey, K. L., Dansereau, D. F., Skaggs, L. P., Hall, R. H., & Pitre, U. (1989). Effects of scripted cooperation and knowledge maps on the processing of technical material. *Journal of Educational Psychology*, *81*(4), 604–609. https://doi.org/10.1037/0022-0663.81.4.604

Ricketts, J. (2011, November). Research Review: Reading comprehension in developmental disorders of language and communication. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. https://doi.org/10.1111/j.1469-7610.2011.02438.x

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, *77*(1), 1–15. https://doi.org/10.1111/j.1467-8624.2006.00852.x

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM - Mathematics Education*, *49*(4), 599–611. https://doi.org/10.1007/s11858-017-0834-z

Roberts, D. H., Newman, L. R., & Schwartzstein, R. M. (2012). Medical Teacher Twelve tips for facilitating Millennials' learning Twelve tips for facilitating Millennials' learning. https://doi.org/10.3109/0142159X.2011.613498

Rosenthal, R. (1976). *Experimenter effects in behavioral research, Enlarged ed. Experimenter effects in behavioral research, Enlarged ed.* Oxford, England: Irvington.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, *118*(2), 183–192. https://doi.org/10.1037/0033-2909.118.2.183

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, *3*(1), 16–20. https://doi.org/10.1007/BF02322211

Roy, M., & Chi, M. T. H. (2005). The Self-explanation Principle. In *Cambridge Handbook of Multimedia Learning*. https://doi.org/10.1016/0364-0213(89)90002-5

Sánchez-Meca, J., & Marín-Martínez, F. (2010). Meta-analysis in Psychological Research. *International Journal of Psychological Research*, *3*(1), 150–162. Retrieved from http://www.redalyc.org/articulo.oa?id=299023509012

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education*. Blackwell Publishing. https://doi.org/10.1111/j.1751-228X.2008.00041.x

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/10.1177/0735633115597625

Schworm, S., & Renkl, A. (2006a). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education*, *46*(4), 426–445. https://doi.org/10.1016/j.compedu.2004.08.011

Schworm, S., & Renkl, A. (2006b). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers and Education*, *46*(4), 426–445. https://doi.org/10.1016/j.compedu.2004.08.011

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. https://doi.org/10.1037/0022-0663.99.2.285

Sears, D. O. (1986). College Sophomores in the Laboratory. Influences of a Narrow Data Base on Social Psychology's View of Human Nature. *Journal of Personality and Social Psychology*, *51*(3), 515–530. https://doi.org/10.1037/0022-3514.51.3.515

Shadish, W., Cook, T., & Campbell, D. (2001). *Quasi-Experimental Designs for*

Generalized Causal Inference. Boston: Houghton Mifflin. Retrieved from https://www.researchgate.net/publication/51993098_Quasi-Experimental_Designs_for_Generalized_Causal_Inference

Shank, D. B. (2016). Using Crowdsourcing Websites for Sociological Research: The Case of Amazon Mechanical Turk. *American Sociologist*, *47*(1), 47–55. https://doi.org/10.1007/s12108-015-9266-9

Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: why validity issues in meta-analysis will not go away. *Clinical Psychology Review*, *17*(8), 881–901. https://doi.org/10.1016/s0272-7358(97)00056-1

Siegler, R. (2002). Microgenetic studies of self-explanation. In N. Granott & J. Parziale (Eds.), *Microdevelopment* (pp. 31–58). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511489709.002

Siegler, R. S. (1995). How Does Change Occur: A Microgenetic Study of Number Conservation. *Cognitive Psychology*, *28*(3), 225–273. https://doi.org/10.1006/cogp.1995.1006

Singer, E. (2002). *Survey Methodology Program The Use of Incentives to Reduce Nonresponse in Household Surveys*.

Singer, Eleanor, Van Hoewyk, J., Gebler, N., Raghunathan, T., & Mcgonagle, K. (1999). *The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys*. *Journal of Of®cial Statistics* (Vol. 15).

Smart, R. (1966). Subject selection bias in psychological research. *Canadian Psychologist*, *7a*, 115–121.

Smyth, J., & Pearson, J. (2011). Internet survey methods: A review of strengths, weaknesses, and innovations. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and behavioral research and the Internet: Advances in applied methods and, research strategies* (pp. 11–44). Retrieved from https://books.google.ca/books?hl=en&lr=&id=iXl0DwAAQBAJ&oi=fnd&pg=PT17&dq=Smyth,+J.,++Pearson,+J.+.+Internet+survey+methods:+A+review+of+strengths,

+weaknesses,+and+innovations.+In:+M.+Das,+P.+Ester,++L.+Kaczmirek+,+Social +and+behavioral+research+and+the+I

Stepans, J. (1991). Development of teachers' knowledge of conceptual change through content, seminars, practicum, methods, and student teaching. In A. Gardner, K. Cochran, & K. Tobin (Eds.), *Critical issues in reforming elementary teacher preparation in mathematics and science* (pp. 155–160). Greeley: University of Northern Colorado.

Sterman, J., & Ford, D. V. (1998). Expert Knowledge Elicitation to Improve Mental and Formal Models.'. *System Dynamics Review*.

Stewart, N., Chandler, J., & Paolacci, G. (2017, October 1). Crowdsourcing Samples in Cognitive Science. *Trends in Cognitive Sciences*. Elsevier Ltd. https://doi.org/10.1016/j.tics.2017.06.007

Suri, S., & Watts, D. J. (2011). Cooperation and Contagion in Web-Based, Networked Public Goods Experiments. *PLoS ONE*, *6*(3), e16836. https://doi.org/10.1371/journal.pone.0016836

Swan, K. (2002). Education, Communication & Information Building Learning Communities in Online Courses: the importance of interaction Building Learning Communities in Online Courses: the importance of interaction. *Communication & Information*, *2*(1). https://doi.org/10.1080/1463631022000005016

Sweller, J., & Chandler, P. (1994). Why Some Material is Difficult to Learn. *Cognition and Instruction*. https://doi.org/10.1207/s1532690xci1203_1

Sweller, J., & Cooper, G. A. (1985). The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra. *Cognition and Instruction*. https://doi.org/10.1207/s1532690xci0201_3

Sweller, J., Van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*. https://doi.org/10.1023/A:1022193728205

Tabachnick, B., & Fidell, L. S. (2007). *Using Multivarite Statistics*. *Boston: Allyn & Bacon* (Vol. 3).

Taylor, S. E., Pham, L. B., Rivkin, I. D., & Armor, D. A. (1998). Harnessing the Imagination: Mental Simulation, Self-Regulation, and Coping. *American Psychologist*. https://doi.org/10.1037/0003-066X.53.4.429

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. https://doi.org/10.1348/026151007X231057

The Condition of Education - Letter From the Commissioner. (2019). Retrieved from https://nces.ed.gov/programs/coe/commissioner.asp

Thomas, K. M., O'bannon, B. W., & Bolton, N. (2013). Cell Phones in the Classroom: Teachers' Perspectives of Inclusion, Benefits, and Barriers, Computers in the Schools, *30*(4), 295–308. https://doi.org/10.1080/07380569.2013.844637

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. https://doi.org/10.1007/s11422-008-9090-4

van Balkom, A., van Oppen, P., Bermeulen, A., van Dyck, R., Nauta, M., & Vorst, H. (1994). A meta-analysis on the treatment of obsessive compulsive disorder: A comparison of antidepressants, behavior, and cognitive therapy. *Clinical Psychology Review*, *14*(5), 359–381. https://doi.org/10.1016/0272-7358(94)90033-7

van Dijk, T. A., & Kintsch, W. (1993). *Strategies of discourse comprehension*. New York: Academic Press.

VanLehn, K. (1996). COGNITIVE SKILL ACQUISITION. *Annual Review of Psychology*, *47*(1), 513–539. https://doi.org/10.1146/annurev.psych.47.1.513

VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A Model of the Self-Explanation Effect. *Journal of the Learning Sciences*.

https://doi.org/10.1207/s15327809jls0201_1

Vosniadou, S. (2007). The cognitive-situative divide and the problem of conceptual change. *Educational Psychologist*. https://doi.org/10.1080/00461520709336918

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*. https://doi.org/10.1016/0010-0285(92)90018-W

Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science*. https://doi.org/10.1016/0364-0213(94)90022-1

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction*. https://doi.org/10.1016/S0959-4752(00)00038-4

Wasserman, S., Hedges, L. V., & Olkin, I. (1988). Statistical Methods for Meta-Analysis. *Journal of Educational Statistics*, *13*(1), 75. https://doi.org/10.2307/1164953

Weaver, T. L., & Clum, G. A. (1995). Psychological distress associated with interpersonal violence: A meta-analysis. *Clinical Psychology Review*, *15*(2), 115–140. https://doi.org/10.1016/0272-7358(95)00004-9

Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., … Blakemore, S. J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition*, *22*(1), 264–271. https://doi.org/10.1016/j.concog.2013.01.004

Weiss, B., & Weisz, J. R. (1995). Relative effectiveness of behavioral versus nonbehavioral child psychotherapy. *Journal of Consulting and Clinical Psychology*, *63*(2), 317–320. https://doi.org/10.1037/0022-006X.63.2.317

Wichmann, A. (2010). Multi-level support with respect to inquiry : Explanations and regulation during an inquiry cycle. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=pdx&AN=0226985&site=ehost-live

Wickstrom, G., & Bendix, T. (2000). The ″Hawthorne effect″ - What did the original Hawthorne studies actually show? *Scandinavian Journal of Work, Environment and Health*. https://doi.org/10.5271/sjweh.555

Willging, P. A., & Johnson, S. D. (2009). Factors that Influence Students' Decision to Dropout of Online Courses. *Journal of Asynchronous Learning Networks*, *13*, 115.

Winkler, R. L., & Murphy, A. H. (1973). Experiments in the laboratory and the real world. *Organizational Behavior and Human Performance*, *10*(2), 252–270. https://doi.org/10.1016/0030-5073(73)90017-2

Winne, P. (2000). Self-Regulated Learning Viewed from Models of Information Processing. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement : theoretical perspectives* (pp. 153–189). Mahwah, NJ : Lawrence Erlbaum Associates.

Winne, P., & Hadwin, A. (1998). Studying as self-regulated learning. In *Metacognition in educational theory and practice*.

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. https://doi.org/10.1007/s10648-010-9136-5

Wong, R., Lawson, M., & Keeves, J. (2002). The effects of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction*, *12*, 233–262.

Wylie, R., Koedinger, K., & Mitamura, T. (2009). Is self-explanation always better? The effects of adding self-explanation prompts to an English grammar tutor. *Proceedings of the 31st …*.

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 and Education*, *54*(4), 1089–1100. https://doi.org/10.1016/j.compedu.2009.10.013

Zhu, X., & Simon, H. A. (1987). Learning Mathematics From Examples and by Doing.

Cognition and Instruction, *4*(3), 137–166. https://doi.org/10.1207/s1532690xci0403_1

Zimmerman, B. (2000). Self-Efficacy: An Essential Motive to Learn. *Contemporary Educational Psychology*. https://doi.org/10.1006/ceps.1999.1016

Appendix A. Advertisement for Research Participants

Research Participants Needed

Earn $20 through Paypal

What do I have to do?

- You will be asked to provide some information, read a passage, and answer questions about the passage.

Study details

- This study has 3 sessions and each session can last between 20-30 minutes.
- Each session is completely online.
- Each session must be completed in one sitting. Therefore, you cannot log out and then log back into the system. Do not log into the system unless you have time set aside.
- If you do not complete a session in one sitting, you will not receive your payment for that session or pair of sessions

Payment

- Once you have completed Session #1, you will receive $10 through your email.
- You will receive the remaining $10 after you complete both Session #2 and Session #3.

If you'd like to participate, please email [redacted] and I will send you an invite for Session #1.

Appendix B. Ethics Approval

SFU OFFICE OF RESEARCH ETHICS

Street Address
Simon Fraser University
Discovery 2
Room 230, 8900 Nelson Way
Burnaby, BC Canada V5A 4W9

Mailing Address
8888 University Drive
Discovery 2
Burnaby, BC Canada
V5A 1S6

http://www.sfu.ca/ore.html

Delegated Minimal Risk Approval

Study Number: 2016s0072
Study Title: The effects of domain specific and generic self-explanation prompts on reading comprehension.

Approval Date: 2016 March 8
Principal Investigator: Bisra, Kiran
SFU Position: Graduate Student

Expiry Date: 2017 March 8
Supervisor: Nesbit, John
Faculty/Department: Education

SFU Collaborator: n/a
External Collaborator: n/a
Research Personnel: n/a

Funding Source: n/a
Grant Title: none

Documents Approved in this Application:

- Study Details, uploaded 2016 February 29
- Consent Form, uploaded 2016 February 29
- Recruitment Poster, February 25
- Demographic Survey, February 24
- Need for Cognition Scale, February 24
- Reading Passages, February 24
- Self Explanation Training Document, February 24

I am pleased to inform you that the above referenced study has been approved by the Associate Director, Office of Research Ethics, on behalf of the Research Ethics Board in accordance with University Policy R20.01 (http://www.sfu.ca/policies/research/r20.01.htm). The Board reviews and may amend decisions or subsequent amendments made independently by the Associate Director, Director, Chair or Deputy Chair at its regular monthly meeting.

The approval for this protocol expires on the **Expiry Date. An annual renewal form must be completed every year prior to the Expiry Date. Failure to submit an annual renewal form will lead to your study being suspended and potentially terminated.**

Your application has been categorized as "Minimal Risk". "Minimal Risk" occurs when potential participants can reasonably be expected to regard the probability and magnitude of possible harms to be no greater than those encountered by the participant in those aspects of his or her everyday life that relate to the research. Please note that it is the responsibility of the researcher, or the responsibility of the Faculty Supervisor if the researcher is a graduate student, to maintain written or other forms of documented consent for a period of 2 years after the research has been completed, or as required according to relevant regulations and funder/sponsor requirements.

SFU

OFFICE OF RESEARCH ETHICS

The Office of Research Ethics **must** be notified of any changes to the approved protocol, or study documents, including unanticipated problems and the completion of the study. To notify the ORE, complete the appropriate form, and append it to an email to [redacted] In all email correspondence relating to this application, please reference the application number shown on this letter in the subject line, in square brackets; this will ensure that all correspondence is saved to the electronic study file.

All correspondence with regards to this application will be sent to your SFU email address.

This letter is your official research ethics approval documentation for this project. Please keep this document for reference purposes and acknowledge receipt by email to [redacted], including the study number in square brackets as the first item in the Subject Line.

Best wishes for success in this research.

Sincerely,

Dina Shafey

Dina Shafey, PhD, MBA
Associate Director
Office of Research Ethics

Appendix C. Consent Form

Consent Form

The effects of domain specific and generic self-explanation prompts on reading comprehension.

Who is conducting the study?

This investigation is undertaken by Kiran Bisra [redacted] and John Nesbit [redacted] in the Faculty of Education.

Who is funding this study?

This research is funded by the Social Sciences and Humanities Research Council of Canada.

Why are we doing this study?

The purpose of this study is to understand the relationship between personal factors collected via the questionnaires, reading comprehension, and study techniques. It is hoped this research will lead to improvements in the way many university courses are taught.

Your participation is voluntary

You have the right to refuse to participate in this study without any negative consequences to your grades. If you decide to participate, you may withdraw from the study at any time without any negative consequences to your grades.

What happens if you say "Yes, I want to be in the study"?

If you say yes, you are agreeing to: Give the researcher access to data collected during your online session. If you choose not to participate, your data will not be used for the research.

What are the risks?

The risks to you are minimal. Although all reasonable measures will be taken to secure the data and preserve your anonymity, no computer system is 100% secure. There is a small risk someone outside the research team might be able to link the data with your identity.

What are the benefits of this research?

As this is an observational study, there may be no direct benefits to you from participating. However, it is hoped this research will lead to improvements in study techniques that will benefit to future students. Because the results of this research will be published, it may lead to improvements in many other university course

Payment?

All participants will receive $15 completion of the survey (approximately 60 mins) and $10 for completing the first two modules.

How will your identity be protected?

You will not be identified by name in any research reports. Key data such as questionnaire scores will be transferred to a spreadsheet file for data analysis. All identifying information will be removed from the spreadsheet file before data analysis and publication. All data stored on servers will be password protected. If data are downloaded onto a computer for analysis, the files will be password protected and the computer will be kept secure in a locked cabinet in the researcher’s office. The collected data will be anonymous and the risk of identification of a particular individual is low or very low.

What If you decide to withdraw your consent?

You can withdraw from the research by logging out of the system.

Publication of results

The results of this research will likely be presented at a conference and published in a peer- reviewed journal. If you write your email address on the consent form, we will send you a link to the published report. Keep in mind that this may take a year or more.

Who to contact for further information

Contact the lead researcher [redacted] or her supervisor John Nesbit [redacted] if you have questions about the study.

Complaints or concerns?

If you have any concerns about your rights as a research participant and/or your experiences while participating in this study, you may contact Dr. Jeffery Toward, Director, Office of Research Ethics at [redacted].

How might the data be re-used?

We will post a spreadsheet on the Web that presents the key data we collected. It will not include identifying information such as your name, age or gender. This is done so that other researchers can check our work and replicate our analysis. It is possible that researchers might re-use the data to investigate research questions not identified here.

Consent Signature Page

Taking part in this study is entirely up to you. You have the right to refuse to participate in this study. If you decide to take part, you may choose to pull out of the study at any time. Refusal to participate or withdrawal after agreeing to participate will have no adverse effects on your grades or evaluation in the course or classroom.

Your signature below indicates that you have received a copy of this consent form for your own records.

Your signature indicates that you consent to participate in this study

____

Participant signature Date (dd/mm/yyyy)

____

Printed name of the participant signing above

Participant’s email address (to receive a link to the published report)

Appendix D. Study Details

PROJECT TITLE

The effects of domain specific and generic self-explanation prompts on reading comprehension.

COLLABORATORS

None

BACKGROUND

Self-explanation (SE) is the activity of explaining to oneself new information presented in text, animation, diagram or another medium (Chi, 2000; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Roy & Chi, 2005). It is a learner-generated explanation, as opposed to an explanation provided by an external source (i.e. an instructor or a textbook), and it may occur spontaneously (Chi et al., 1989; Pirolli & Recker, 1994) or in response to instruction (Bielaczyc, Pirolli, & Brown, 1995).

Engaging in SE allows the learner to integrate current learning with prior knowledge (Chi, 2000; Chi, DeLeeuw, Chiu, & LaVancher, 1994). Research has shown that when encouraged to self-explain, learners perform better on problem-solving tasks, repair flawed mental models, and generate inferences (Chi, 2000; Chi, DeLeeuw, et al., 1994).

Research findings have shown that readers who self-explain the text, either spontaneously or when prompted to do so, understand more from the text and therefore construct better mental models of the content (Chi & Bassok, 1989; Chi, DeLeeuw, et al., 1994; Magliano, Trabasso, & Graesser, 1999; Trabasso & Magliano, 1996). Self-explanation prompts (SEPs) can be verbal cues from a person (Chi, DeLeeuw, et al., 1994), generated by computer tutors (Aleven & Koedinger, 2002; McNamara, 2004) or embedded into the learning materials themselves (Renkl, Stark, Gruber, & Mandl, 1998).

Self-explanation prompts can be designed to induce particular cognitive activity in a learner. For example, prompts have asked the learner to justify the correct step (Berthold, Eysink, & Renkl, 2009; Rittle-Johnson, 2006), engage in reflection (Chi, DeLeeuw, et al., 1994), reference terms in a glossary (Aleven & Koedinger, 2002), etc. Furthermore, prompts can be specific to the domain content, or more general in nature.

Research has shown that less able students learn better with specific prompts and more able students learn better with generic prompts in well-structured domains (Aleven, Pinkwart, Ashely, Lynch, 2015). However, the cognitive effects of domain specific and generic self-explanation prompts while a learner is reading have not been well researched.

STUDY PURPOSE

This study is being conducted to gain further insight into whether specific or generic self-explanation prompts better aid in reading comprehension. Furthermore, we will investigate which type of leaner benefits most from domain specific and generic prompts during reading.

HYPOTHESIS/AIMS

Research question 1: Do participants who receive a domain specific self-explanation prompt perform better on outcome measures than participants who receive a domain generic self- explanation prompt?

Hypothesis 1: Participants who receive the domain generic self-explanation prompt will outperform participants who receive the domain specific self-explanation prompts. Domain generic self-explanation prompts allow participants to fill-in their own knowledge gaps.

Research question 2: Do participants with poor reading comprehension skills perform better on outcome measures when they receive a domain specific self-explanation prompt?

Hypothesis 2: Participants with poor reading comprehension skills will perform better on outcome measures when they receive a domain specific self-explanation prompt. Domain specific self- explanation prompts allow this group of participants to focus on specific sections of the text.

LOCATION WHERE RESEARCH WILL BE CONDUCTED

This study will be available for participants to complete online. Therefore, participants could potentially partake in the study at home, in class, or elsewhere. A location on campus with computers able to access the study instruments will also be made available. This location is a lab in the basement of the Education building and contains multiple computers.

PROSPECTIVE PARTICIPANT INFORMATION

<u>Inclusion Criteria</u>

Each participant must be 18 years old or older. This study specifically is focusing on post- secondary teaching and learning and hence the learning materials are written at an undergraduate level. Participants who can read and write in English

<u>Exclusion Criteria</u>

Minors will be excluded from this study. As well as anyone cannot read and write in English. All learning materials are written in English. No one else will be excluded back on attributes such as

culture, religion, race, disability, sexual orientation, ethnicity, or gender.

<u>Number of Participants</u>

There will be between 75 and 125 participants.

DETAILED RESEARCH PROCEDURES

This study will be conducted over a single 60 minute session in which participants will be required to provide information and interact with materials on a computer. The research design is an experiment with simple random assignment into three groups. Each participant will be provided with one of three randomly assigned URL links that will determine which treatment they receive. Randomization is important to this study because the statistical modelling that will be used to analyze the data require random sampling in order to more confidently attribute any differences between group outcomes to the experimental procedures or treatment.

The participants will work their way through a survey that is administered through FluidSurveys, and online survey tool. All data will be collected by this software and the servers for FluidSurveys are located in Canada.

Step 1: Each participant will read a consent form online and agree to participant.

Step 2: Each participant will receive a demographic questionnaire. This questionnaire will ask the participants their age, gender, country, EAL or ESL status, past education, and CGPA (where applicable).

Step 3: Each participant will read a short text passage (less than 600 words) and then complete a 4 question multiple choice reading comprehension test. This passage will be referred to as the Practice Passage.

Step 4: At this point, participants will be filtered into Groups: A, B and C. Groups A and B will read about self-explanation, how to self-explain and what is not self-explanation. Meanwhile, Group C will read a passage.

Step 5: Each group will receive a different set of instructions:

Group A: Will be asked to read the following passage and prompted with a self-explanation domain- specific prompt.

Group B: Will be asked to read the following passage and prompted with a self-explanation domain- generic prompt.

Group C: Will be asked to read the following passage and make notes about the passage.

Step 6: Each participant will read a second short text passage (less than 600 words) labelled Passage 1.

Step 7: Once Passage 1 has been read, each participant will be asked to recall as much of the passage as possible.

Step 8: After the recall task, each participant is asked to complete a 4 question multiple choice read comprehension test about Passage B. Steps 6-8 will be repeated for Passages 2 and 3.

Step 9: Lastly, each participant will complete the Need for Cognition 18-item scale. This scale reflects an individual difference that has been well researched in the literature and reflects an individual's level of desire to engage in complex, inquisitive, and analytical thoughts.

Recruitment

Participants will be recruited through ads posted on boards around the SFU campus, online ads, and in class announcements. The ads that will be placed around campus are attached in the materials. Ads will also be placed on online social platforms, such as Facebook. Lastly, in-class announcements will be made to SFU students. Student recruited from the classroom will be ensured their participation will have no impact on their course grade.

Reimbursements and Payments

Each participant will be compensated with $15 for their time.

Secondary Data/Tissue Analysis

Not Applicable

Deception/Partial Disclosure

Not Applicable

Photography, Video / Audio Recording

Not Applicable

Instrumentation

Not Applicable

Online Surveys

FluidSurveys is an online survey tool that will be used in this study. All data that is collected is held on servers located inside Canada.

Obtaining Consent

If participants are logging in from home or elsewhere, they will be provided with a digital consent form at the beginning of the survey. The software will record the participant's consent and timestamp. If participants come to the lab, then they will be provided with a written consent form. This form will be kept and stored in a cabinet under lock and key. The consent form will include a short description of the study, any foreseeable risks and potential benefits, contact information of the researcher, and assurance that the participant can choose to end their involvement at any time. Student recruited from the classroom will be ensured their participation will have no impact on their evaluation nor their relationship with the professor.

Competency and Capacity Consent for minors

No minors will be asked to participate in this study.

Consent for individuals who lack the capacity to consent for themselves

This study is not targeting participants who lack capacity.

POTENTIAL BENEFITS

For those participants who receive training on how to self-explain, they may learn a new study technique. This study technique has been investigated in the research literature and no negative effects have been observed. the results of the study may increase knowledge about teaching and learning and lead to the development of better learning strategies

POTENTIAL RISKS

Participants may not be able to cope with the duration of the study. Otherwise, no foreseeable risks to the participants or the investigator and research staff.

DESIGNATION OF THE STUDY AS MINIMAL OR NON-MINIMAL RISK

Minimal risk review

MAINTENANCE OF CONFIDENTIALITY

All data stored on servers will be password protected. If data are downloaded onto a computer for analysis, the files will be password protected and the computer will be kept secure in a locked cabinet in the researcher's office. The collected data will be confidential (no names) and the risk of identification of a particular individual is low or very low.

Access to the Data- Investigators and Staff

Principal Investigator – Kiran Bisra (Phd Student) Graduate Supervisor – John Nesbit (SFU professor)

DATA ANALYSIS PLAN

The goal of the study is to investigate whether differences in treatment lead to differences in reading comprehension. As well, other variables will be used as covariates; such as the scores from the Passage A reading comprehension test. Therefore, an analysis of co-variance will be used to analyze the data.

RETENTION AND DESTRUCTION OF DATA

The data will be kept on *FluidSurveys* servers for two years. After two years has elapsed, the account will be deleted. This server is located in Canada and access is protected through a userID and password. Downloaded data will be stored on a computer with password protection for five years, after which the data will be destroyed. The computer will be kept secure in a locked cabinet in the researcher's office. Once the graduate student has left the university, the data will be possession of the graduate supervisor.

FUTURE USE OF DATA

We may post a spreadsheet on the Web that presents the key data we collected. It will not include identifying information such as age or gender. This is done so that other researchers can check our work and replicate our analysis. It is possible that researchers might re-use the data to investigate research questions not identified here

DISSEMINATION OF RESULTS

Once the study is completed, a write-up will be made available online. Participants will be informed of the website before they begin the survey. Only participants who provide their email addresses on the consent form will have the results of the investigation sent to them.

Appendix E. Reading Passages with SEPs and Multiple Choice Questions

Reading Passage A - Overpopulation

Generic Self-explanation Prompt Instructions: **Generate self-explanations by sentence, paragraph and across paragraphs as you read this passage. Three boxes are provided throughout the passage. Type out what you've explained to yourself in the boxes.**

Specific Self-explanation Prompt Instructions: **There are three self-explanation prompts provided for you. Please respond to the prompts and type out what you've explained to yourself in the boxes.**

Control Instructions: **Textboxes have been provided for you if you wish you take notes. You are NOT required to you use them.**

Traditionally, healthcare has been delivered with a paternalistic approach. Behind this approach is the idea that medical decisions are best left in the hands of those providing healthcare: doctors. Patients present problems to doctors, who make decisions about treatment, prescriptions and further testing. However, the paternalistic approach may threaten patients' ability to make their own health care decisions.

Generally speaking, doctors may not know what matters most to patients regarding their ability to get and stay well. Therefore, a patient's right to make medical decisions, without a health care provider trying to influence the decision, is crucial. The autonomous decision-making approach emphases the right and responsibility of individuals to decide how they live and how they die.

Generic Self-explanation Prompt: **Type out your self-explanations below:**

Specific Self-explanation Prompt: **Type out your self-explanations about Malthus' theory below:**

Control Prompt: **Notes - Optional**

In 1966, United States President Lyndon Johnson shipped wheat to India to reduce famine on the condition that the country accelerate its family planning campaign. Johnson was among the US presidents who were concerned with the harmful effects of rapid population growth in developing countries.

But the 1970s surprised those who were concerned about the population growth. Instead of being a period of widespread famine, new crop strains (especially grains such as rice and wheat) caused a large increase in the global production of cereals, the main energy source in the global diet. People became more optimistic. By the end of the decade, the public health community felt so positive, they came up with the slogan, "Health for All by the Year 2000." Average life expectancy continued to rise almost everywhere.

Generic Self-explanation Prompt: **Type out your self-explanations below:**

Specific Self-explanation Prompt: **Type out your self-explanations about "Health for all by the year 2000" below:**

Control Prompt: **Notes - Optional**

World oil shocks in 1973 and 1979 also fueled this dramatic policy shift. World oil shocks contributed both to “stagflation” — a combination of rising unemployment with higher prices — and to increased economic power for the oil-producing countries of the Third World. Stagflation was accompanied by a decline in concern for Third World development among politicians and the general public. Unlike his predecessors, US president Ronald Reagan considered concerns about global population size to be "vastly exaggerated." In 1980, the US surprised the world by reducing its previous efforts to promote global family planning.

As foreign aid budgets fell, the Health for All targets began to slip from reach. The result of market deregulation and generally high birth rates for many Third World countries has led to poorer outcomes than predicted in the 1970s. The failure of development is most obvious in many sub-Saharan countries, where life expectancy has fallen substantially. Life expectancy has also fallen in Haiti, Russia, North Korea, and a handful of other nations.

Amid the many different explanations for the horrific 1994 Rwandan genocide, the possibility of a Malthusian check was scarcely mentioned. There is even less discussion about it in relation to the sub-Saharan epidemic of HIV/AIDS. However, the issues surrounding human carrying capacity and overpopulation deserve fresh consideration. Perhaps putting greater resources and efforts into reducing the population can reduce suffering and improve the health outcomes of those in Third World countries.

Generic Self-explanation Prompt: **Type out your self-explanations below:**

Specific Self-explanation Prompt: **Type out your self-explanations about stagflation below:**

Control Prompt: **Notes - Optional**

1. United States President Lyndon Johnson
   - A. was concerned with the harmful effects of rapid population growth in developing countries
   - B. refused to ship wheat to India during their famine
   - C. decreased foreign aid during his presidency
   - D. did not support family planning initiatives
2. According to the passage, which of the following is not a reason the USA's foreign aid budget fell starting in the 1980s.
   - A. improvements in motor-vehicle safety
   - B. the 1973 "oil price shock"
   - C. improved life expectancy in Third World countries
   - D. stagflation

3. Which of the following statements, if true, would most weaken the author's arguments concerning carrying capacity and overpopulation?
   - A. Vaccines contributed more to increased life expectancy in the 20th century than antibiotics.
   - B. Malthus's views were widely criticized at the time of their publication.
   - C. Historical population reduction is fairly uncommon.
   - D. Historical population reduction is most common when resources exceed population demands.

4. Which of the following would the author most likely consider a modern day "negative check"?

A. the rise of national vaccination programs

B. widespread famine among sub-Saharan nations.

C. escalating violence in the Middle East.

D. growing rates of birth control usage

5. The passage author most likely believes the Rwandan Genocide and AIDS crisis are…

A. examples of the dangerous consequences of excessive foreign aid.

B. less serious than population declines in Haiti, Russia, and North Korea.

C. possible consequences of overpopulation.

D. not related to one another.

1. A - was concerned with the harmful effects of rapid population growth in developing countries
2. A - improvements in motor-vehicle safety
3. D - Historical population reduction is most common when resources exceed population demands.
4. D - growing rates of birth control usage
5. C - possible consequences of overpopulation.

Reading Passage B – Nudge policy

Generic Self-explanation Prompt: **Generate self-explanations by sentence, paragraph and across paragraphs as you read this passage. Three boxes are provided throughout the passage. Type out what you've explained to yourself in the boxes.**

Specific Self-explanation Prompt: **There are three boxes provided for you to type out your self-explanations.**

Traditionally, healthcare has been delivered with a paternalistic approach. Behind this approach is the idea that medical decisions are best left in the hands of those providing healthcare: doctors. Patients present problems to doctors, who make decisions about treatment, prescriptions and further testing. However, the paternalistic approach may threaten patients' ability to make their own health care decisions.

Generally speaking, doctors may not know what matters most to patients regarding their ability to get and stay well. Therefore, a patient's right to make medical decisions, without a health care provider trying to influence the decision, is crucial. The autonomous decision-making approach emphases the right and responsibility of individuals to decide how they live and how they die. Under this approach, doctors are required to ensure patients understand what kind of treatment they are agreeing to, known as "informed consent", and to share with patients all relevant information about their health.

Generic Self-explanation Prompt: **Type out your self-explanations below:**

Specific Self-explanation Prompt: **Type out your self-explanations about the paternalistic approach below:**

Control Prompt: **Notes - Optional**

One main difficulty with autonomous decision making is that it requires patients to have comprehensive and objective information. This can cause problems because patients may struggle to interpret complex figures (such as percentages and probabilities) provided by doctors. For example, they may over or under-estimate the risks or potential benefits of treatment. Also, most patients do not have specific medical training or the experience to be able to understand the details of their illness. And lastly, patients may be influenced by strong emotions, such as fear and grief. Studies of cancer patients have found that their treatment decisions are not based on a reasonable evaluation of risks and benefits, but instead are based on hope − even when considering a treatment with a high likelihood of failure.

The differences between these two approaches can be made clear through the following example. A doctor is discussing dosage options for hypertension with a patient, when the patient suggests stopping treatment. Under a paternalistic approach, the doctor may exaggerate the possible unpleasant side-effects of this action and state, “you will almost certainly suffer a stroke”, a claim the doctor knows to be untrue. The doctor may justify misleading the patient because treatment is better for the patient in the long term. On the other hand, a doctor with a different philosophy may respect the patient’s wish to stop all treatment.

Generic Self-explanation Prompt: **Type out your self-explanations below:**

Specific Self-explanation Prompt: **Type out your self-explanations about autonomous decision making below:**

Control Prompt: **Notes - Optional**

An approach known as libertarian paternalism, however, aims to provide a framework where doctors influence choices. In other words, by changing the "choice architecture" for decision making, patients can be "nudged" into making informed choices. For example, a doctor might present several treatment options to her patient, but discuss the treatment the doctor feels is most appropriate in greater depth. Supporters of libertarian paternalism argue that illness may create a temporary loss of patient autonomy caused by physical or mental side effects. As a result, to some degree, paternalistic input is justified to restore an individual’s autonomy.

Critics of libertarian paternalism have suggested that it is perhaps more accurate to consider the nudge techniques a form of "manipulation”. Any attempt to influence the patient will result in a decision that is not truly being made by the patient. However, being a patient with a potentially life-threatening condition puts a considerable strain on one’s ability to process information. This raises the question of whether it is the duty of a doctor to recognize and correct reasoning failure in their patients. If so, would this be considered a paternalistic act?

Generic Self-explanation Prompt: **Type out your self-explanations below:**

Specific Self-explanation Prompt: **Type out your self-explanations about libertarian paternalism below:**

Control Prompt: **Notes - Optional**

1. In which scenario did the doctor not act paternalistically toward the patient?
    a. A patient is diagnosed as being terminally ill. Believing that it would be better for him to die still hopeful of a recovery, the physician does not inform the patient of this diagnosis.
    b. A patient complains of severe back pains. The physician discovers, after extensive examination, that there is no problem. He prescribes a placebo, and the symptoms disappear.
    c. A patient's 13-year old daughter is discovered to be pregnant by a doctor. Although she treats the daughter, at the request of the daughter, the doctor does not inform the patient about her daughter's condition.
    d. A physician discovers a serious condition which requires a delicate operation. Because of the patient's emotional tendencies, she misleads the patient about the diagnosis.

2. Based on passage information, critics of medical paternalism would most likely respond to news that billions of dollars per year could be saved if clinicians downplayed the effectiveness of a powerful but often unnecessary new treatment by:
    a. Supporting this practice because the money saved is significant enough to outweigh the loss of autonomy.
    b. Opposing this practice because the money saved is significant enough to outweigh the loss of autonomy.
    c. Supporting this practice because patients are not receiving all relevant information.
    d. Opposing this practice because patients are not receiving all relevant information.

3. Suppose a new chemotherapy drug is developed that has the potential to extend one out of every ten thousand cancer patients' lives by up to four months. The drug costs $125,000 per year and has horrific side effects that can lead to prolonged hospitalization. The passage author would most likely respond to this news by arguing that clinicians should:
    a. Focus on informing patients of the side effects and costs of the treatment.
    b. Prescribe the treatment in all cancer cases.
    c. Provide patients with all known information on the drug so they can decide their treatment options.
    d. Refuse to prescribe the drug to any patient under any circumstances.

4. Which of the following is true about the supports of the “nudge” approach?
    a. They believe informed consent is crucial to an ethical doctor-patient relationship.
    b. They do not support patient autonomy.
    c. They believe only an individual to whom that decision relates can truly appreciate his own values and preferences and therefore doctors should not attempt to manipulate him into a decision.
    d. They believe utilizing biases to trick patients into certain decisions is right.

5. Libertarian paternalism …
    a. Threatens patients’ ability to make their own health care decisions.
    b. Is an approach where doctors influence choices, while patients make decisions from a range of available options.
    c. Describes a patient’s wish to stop all treatment.
    d. Allows doctors to make decisions about treatment, prescriptions and further testing without patient consent.

Reading Passage C – Common Limited Property

Generic Self-explanation Prompt: Explain to yourself the meaning and relevance of each sentence or paragraph. Type out what you've explained to yourself.

Specific Self-explanation Prompt: Please explain to yourself the relevance of common property to this passage. Type out what you've explained to yourself.

In British Columbia, condominium developments are known as "strata corporations". A strata corporation is comprised of two types of property: (1) strata lots (owned entirely by individual owners) and (2) common property (of which ownership is shared by everyone). Common property include areas shared by all of the owners, such as the lobby, gym, pool, or grounds around the building.

Generally speaking, an owner is responsible for repairing and maintaining everything inside the bounds of his or her own strata lot. The strata corporation is responsible for repairing and maintaining the common property. For instance, in the above example, the strata corporation would hire a gardener to maintain the grounds and a janitor to clean the pool and gym area, and then divide the cost among all of the strata lot owners. On the other hand, if you needed to repair a broken faucet in your strata lot, you would have to personally arrange for a plumber to fix the problem. In this way, living in a strata corporation is different than living in a rental building with a landlord.

Limited common property is a special subset of common property. While limited common property is owned by everyone in the strata corporation, its use is restricted for one or more specific strata lots within the strata corporation. For instance, balconies are often limited common property. Practically speaking, although a balcony may be owed by all the owners as common property, only the owner of the adjacent strata lot and his or her guests is allowed to use it. This can create a lot of confusion and anger regarding who is responsible for repairing and maintaining the balcony.

Many owners may be frustrated and wonder why they have to pay to repair their neighbour's balcony. Especially since the owner receive no benefit from her neighbour's space. At first, this issue may seem like a minor annoyance; however, the cost of balcony and patio repairs can be in the millions of dollars. The costs are even higher if you live in a wet climate, such as Vancouver.

The justification for this legislation is that even though the limited common property is only used by one owner, the government has decided it is better to leave major repairs to the whole strata corporation. More specifically, issues that affect the integrity of the building's structure or its weather-proofing up are legally considered a joint-expense for all of the owners. That way, the strata corporation can ensure that the repairs are done properly and do not create structural or water leak problems for other owners.

Another consequence of a balcony being limited common property is that the strata council can make rules about what you can and cannot keep on your balcony. For instance, it may pass a rule that you cannot store recycling or your bicycle outside. Once again, confusion over the level of independence an individual owner has over his or her balcony is not infrequent.

In order to clarify what parts of your strata corporation are part of your strata lot, and which are common property, including limited common property, you can consult a document called a "strata plan". This strata plan is available for approximately $15 from the Land Title Office, and before purchasing into a condominium corporation it is important to get a copy as it can save a lot of frustration later on.

Which of the following is most likely to be something an individual owner is responsible for, rather than their strata corporation?

a. washing the cement floor in the parkade;
b. cleaning the pool in the rec room;
c. replacing a cracked toilet bowl in their unit;
d. trimming the hedges around the strata complex;

2. Which of the following is not a reason given in the passage for why the strata corporation is responsible for private balconies in some strata corporations:

   a. ensuring the building stays protected against the weather

   b. protecting the integrity of the building's structure

   c. making sure that the repairs are done properly

   d. to reduce conflict between neighbours in case something goes wrong

3. An owner who enjoys his or her autonomy would MOST likely want to buy in a building in which the balconies, patios, garages, etc. are which of the following:

   a. part of the strata lot

   b. part of the common property

   c. limited common property

   d. registered at the Land Title Office

4. Which of the following, if true, most weakens the BC government's logic for making certain balconies limited common property:

   a. the strata plan can be amended at any time by a unanimous vote of all the owners

   b. balconies can almost always be effectively repaired without affecting the building's exterior or structure;

   c. balconies tend to rot quickly in British Columbia because of the damp climate;

   d. trash stored on balconies can be unsightly and attract vermin;

5. A strata plan…

   a. outlines which parts of the strata corporation are part of an owner's strata lot, and which are common property.

   b. clarifies that balconies are limited common property

   c. is a document that the government has put together to outlines rules that govern stratas.

   d. can create a lot of confusion and anger regarding who is responsible for repairing and maintaining the balcony.

1. C - replacing a cracked toilet bowl in their unit;
2. D - to reduce conflict between neighbours in case something goes wrong
3. A - part of the strata lot
4. B - balconies can almost always be effectively repaired without affecting the building's exterior or structure;

5. A - outlines which parts of the strata corporation are part of an owner's strata lot, and which are common property.

	Words	Sentences	Paragraphs	Flesch-Kincaid Grade level
Passage A- overpopulation	535	27	7	12.5
Passage B- nudge policy	573	29	6	12.6
Passage C- common limited	569	25	7	12.8

Appendix F. Session 1 – Version A

Page #1

Welcome to Session 1!

Thank you for agreeing to participate in this study examining how people learn. You will be asked to provide some information, read a passage, and answer questions about the passage. This first session will approximately take you about 30 mins. Be assured that all of your responses will be kept confidential.

You must complete this study in one sitting. Therefore, you cannot log out and then log

back into the system. Do not hit the "Back" button on your browser.

Please click 'Next' when you are ready to begin.

Page 2

Taking part in this study is entirely up to you. You have the right to refuse to participate in this study. If you decide to take part, you may choose to pull out of the study at any time.

Download Consent Form

To proceed, you must click both boxes:(Consent)

By clicking this box, you acknowledge that you've read the consent form.

By clicking this box, you indicate that you consent to participate in this study.

Page #3

To receive payment for your participation, please enter the number that was emailed to you.

[Textbox]

Page #4

In the next section, you will read a short passage. Once you've finished reading the passage, you will take a short quiz about this passage.

Page #5

Traditionally, healthcare has been delivered with a paternalistic approach. Behind this approach is the idea that medical decisions are best left in the hands of those providing healthcare: doctors. Patients present problems to doctors, who make decisions about treatment, prescriptions and further testing. However, the paternalistic approach may threaten patients' ability to make their own health care decisions.

Generally speaking, doctors may not know what matters most to patients regarding their ability to get and stay well. Therefore, a patient's right to make medical decisions, without a health care provider trying to influence the decision, is crucial. The autonomous decision-making approach emphases the right and responsibility of individuals to decide how they live and how they die. Under this approach, doctors are required to ensure patients understand what kind of treatment they are agreeing to, known as "informed consent", and to share with patients all relevant information about their health.

One main difficulty with autonomous decision making is that it requires patients to have comprehensive and objective information. This can cause problems because patients may struggle to interpret complex figures (such as percentages and probabilities) provided by doctors. For example, they may over or under-estimate the risks or potential benefits of treatment. Also, most patients do not have medical training or the experience to be able to understand the details of their illness. And lastly, patients may be influenced by strong emotions, such as fear and grief. Studies of cancer patients have found that their treatment decisions are not based on a reasonable evaluation of risks and benefits, but instead are based on hope − even when considering a treatment with a high likelihood of failure.

The differences between these two approaches can be made clear through the following example. A doctor is discussing dosage options for hypertension with a patient, when the patient suggests stopping treatment. Under a paternalistic approach, the doctor may exaggerate the possible unpleasant side-effects of this action and state, "you will almost certainly suffer a stroke", a claim the doctor knows to be untrue. The doctor may justify misleading the patient because treatment is better for the patient in the long term. On the other hand, a doctor with a different philosophy may respect the patient's wish to stop all treatment.

An approach known as libertarian paternalism, however, aims to provide a framework where doctors influence choices. In other words, by changing the "choice architecture" for decision making, patients can be "nudged" into making informed choices. For example, a doctor might present several treatment options to her patient, but discuss the treatment the doctor feels is most appropriate in greater depth. Supporters of libertarian paternalism argue that illness may create a temporary loss of patient autonomy caused by physical or mental side effects. As a result, to some degree, paternalistic input is justified to restore an individual's autonomy.

Critics of libertarian paternalism have suggested that it is perhaps more accurate to consider the nudge techniques a form of "manipulation". Any attempt to influence the patient will result in a decision that is not truly being made by the patient. However, being a patient with a potentially life-threatening condition puts a considerable strain on one's ability to process information. This raises the question of whether it is the duty of a doctor to recognize and correct reasoning failure in their patients. If so, would this be considered a paternalistic act?

Page #6

Please type out as much as you can remember about the passage you just read (note form).

[Textbox]

Page #7

1. Why is informed consent important to attain from patients?

[Textbox]

2. What is a criticism of the “nudge” approach?

[Textbox]

3. What is the difference between a paternalistic approach and a libertarian paternalistic approach?

[Textbox]

Page #8

Multiple Choice Questions:

4. In which scenario did the doctor not act paternalistically toward the patient?

A patient is diagnosed as being terminally ill. Believing that it would be better for him to die still hopeful of a recovery, the physician does not inform the patient of this diagnosis.

A patient complains of severe back pains. The physician discovers, after extensive examination, that there is no problem. He prescribes a placebo, and the symptoms disappear.

A patient’s 13-year old daughter is discovered to be pregnant by a doctor. Although she treats the daughter, at the request of the daughter, the doctor does not inform the patient about her daughter’s condition.

A physician discovers a serious condition which requires a delicate operation. Because of the patient's emotional tendencies, she misleads the patient about the diagnosis.

5. Based on passage information, critics of medical paternalism would most likely respond to news that billions of dollars per year could be saved if clinicians downplayed the effectiveness of a powerful but often unnecessary new treatment by:

Supporting this practice because the money saved is significant enough to outweigh the loss of autonomy.

Opposing this practice because the money saved is significant enough to outweigh the loss of autonomy.

Supporting this practice because patients are not receiving all relevant information.

Opposing this practice because patients are not receiving all relevant information.

6. Suppose a new chemotherapy drug is developed that has the potential to extend one out of every ten thousand cancer patients’ lives by up to four months. The drug costs $125,000 per year and has horrific side effects that can lead to prolonged hospitalization. The passage author would most likely respond to this news by arguing that clinicians should:

Focus on informing patients of the side effects and costs of the treatment.

Prescribe the treatment in all cancer cases.

Provide patients with all known information on the drug so they can decide their treatment options.

Refuse to prescribe the drug to any patient under any circumstances.

7. Which of the following is true about the supporters of the “nudge” approach?

They believe informed consent is crucial to an ethical doctor-patient relationship.

They do not support patient autonomy.

They believe only an individual to whom that decision relates can truly appreciate his own values and preferences and therefore doctors should not attempt to manipulate him into a decision.

They believe utilizing biases to trick patients into certain decisions is right.

8. Libertarian paternalism ...

- Threatens patients’ ability to make their own health care decisions.
- Is an approach where doctors influence choices, while patients make decisions from a range of available options.
- Describes a patient’s wish to stop all treatment.
- Allows doctors to make decisions about treatment, prescriptions and further testing without patient consent.

Page #9 You are almost finished this session!

The next section is a very short questionnaire about you.

Branching Information

Page #10

- If not What is the highest degree or level of school you ... is one of [Bachelor's Degree, Master's Degree, Doctorate Degree] then Hide grad_fac
- If not Are you currently enrolled in an ... = Yes then Hide Which degree or level of school are you currently ...
- If not Are you currently enrolled in an ... = Yes then Hide Which Faculty are you currently enrolled in?
- If not Are you currently enrolled in an ... = Yes then Hide What is your cumulative GPA?
- If not What is the highest degree or level of school you ... is one of [Bachelor's Degree, Master's Degree, Doctorate Degree] then Hide What was your graduating cumulative GPA?
- If not Which country are you currently living in? = Canada then Hide Which province are you currently living in?

What is your age?

Drop Down 18-50+

What is your gender?

- Female
- Male

Which country are you currently living in?

- Canada
- USA
- Afghanistan
- Albania
- Algeria
- Andorra
- Angola
- Antarctica
- Antigua and Barbuda
- Argentina

... 172 additional choices hidden

Which province are you currently living in?

- Alberta
- British Columbia
- Manitoba
- New Brunswick
- Newfoundland and Labrador
- Northwest Territories
- Nova Scotia
- Nunavut
- Ontario
- Prince Edward Island
- Quebec
- Saskatchewan
- Yukon

Do you consider yourself as someone for whom English is an additional language (EAL) or English is a secondary language (ESL)?

- Yes
- No

What is the highest degree or level of school you have completed?

- High School Diploma
- Bachelor's Degree
- Master's Degree
- Doctorate Degree

Which Faculty did you graduate from?

- Applied Science
- Arts and Social Sciences
- Business
- Communication, Art and Technology
- Education
- Environment
- Health Sciences/Medicine
- Law
- Science
- Other___________

What was your graduating cumulative GPA?

Letter grade A or 4.33 - 3.34

Letter grade B or 3.33 - 2.34

Letter grade C or 2.33 - 1.68

Below letter grade D or 1.6

Page #11

Branching Information

- If not Are you currently enrolled in an ... = Yes, then Hide Which degree or level of school are you currently ...
- If not Are you currently enrolled in an ... = Yes, then Hide Which Faculty are you currently enrolled in?
- If not Are you currently enrolled in an ... = Yes, then Hide What is your cumulative GPA?
- If not Are you currently enrolled in an ... = Yes, then Hide credits

Are you currently enrolled in an academic program?

Yes

No

Which degree or level of school are you currently enrolled in?

Certificate Associate Degree Bachelor's Degree Master's Degree Doctorate Degree

How many credits have you completed? (credits)

Which Faculty are you currently enrolled in?

o Applied Science
o Arts and Social Sciences
o Business
o Communication, Art and Technology
o Education
o Environment
o Health Sciences/Medicine
o Law
o Science
o Other _

What is your cumulative GPA?

Letter grade A or 4.33 - 3.34 Letter grade B or 3.33 - 2.34 Letter grade C or 2.33 - 1.68 Below letter grade D 1.68

Do not have a CGPA yet

Page #12

Your payment will be sent through Paypal. Please type in the email account you'd like your payment to be sent.

[Textbox]

Page #13 You're done Session 1!

Thanks for your participation.

Appendix G. Session 1 – Version B

Page #1

Welcome to Session 1!

Thank you for agreeing to participate in this study examining how people learn. You will be asked to provide some information, read a passage, and answer questions about the passage. This first session will approximately take you about 30 mins. Be assured that all of your responses will be kept confidential.

You must complete this study in one sitting. Therefore, you cannot log out and then log back into the system. Do not hit the "Back" button on your browser.

Please click 'Next' when you are ready to begin.

Page #2

Taking part in this study is entirely up to you. You have the right to refuse to participate in this study. If you decide to take part, you may choose to pull out of the study at any time.

Download Consent Form

To proceed, you must click both boxes:(Consent)

By clicking this box, you acknowledge that you've read the consent form.

By clicking this box, you indicate that you consent to participate in this study.

Page #3

To receive payment for your participation, please enter the number that was emailed to you.

[Textbox]

Page #4

In the next section, you will read a short passage. Once you've finished reading the passage, you will take a short quiz about this passage.

Page #5

The question of human overpopulation and its relationship to human carrying capacity − the planet's limited ability to support its people − has been controversial for over two centuries. In 1798 the Reverend Thomas Malthus came up with a theory that growth in the number of humans on earth would exceed the growth of resources humans need to live. According to this theory, too much population growth would lead to a reduction in the number of humans by either "positive checks," such as disease, famine, and war, or "negative checks," by which Malthus meant restrictions on marriage.

Malthus's worst fears did not occur during the century following his death in 1834. Food production largely kept pace with the slowly growing global population. However, soon after 1934, the global population began to rise quickly as antibiotics, vaccines, and technology increased the expected lifespan of humans. By the 1960s, many people were again becoming concerned about the global population exceeding global food supply.

In 1966, United States President Lyndon Johnson shipped wheat to India to reduce famine on the condition that the country accelerate its family planning campaign. Johnson was among the US presidents who were concerned with the harmful effects of rapid population growth in developing countries.

But the 1970s surprised those who were concerned about the population growth. Instead of being a period of widespread famine, new crop strains (especially grains such as rice and wheat) caused a large increase in the global production of cereals, the main energy source in the global diet. People became more optimistic. By the end of the decade, the public health community felt so positive, they came up with the slogan, "Health for All by the Year 2000." Average life expectancy continued to rise almost everywhere.

World oil shocks in 1973 and 1979 also fueled this dramatic policy shift. World oil shocks contributed both to “stagflation” — a combination of rising unemployment with higher prices — and to increased economic power for the oil-producing countries of the Third World. Stagflation was accompanied by a decline in concern for Third World development among politicians and the general public. Unlike his predecessors, US president Ronald Reagan considered concerns about global population size to be "vastly exaggerated." In 1980, the US surprised the world by reducing its previous efforts to promote global family planning.

As foreign aid budgets fell, the Health for All targets began to slip from reach. The result of market deregulation and generally high birth rates for many Third World countries has led to poorer outcomes than predicted in the 1970s. The failure of development is most obvious in many sub-Saharan countries, where life expectancy has fallen substantially. Life expectancy has also fallen in Haiti, Russia, North Korea, and a handful of other nations.

Amid the many different explanations for the horrific 1994 Rwandan genocide, the possibility of a Malthusian check was scarcely mentioned. There is even less discussion about it in relation to the sub-Saharan epidemic of HIV/AIDS. However, the issues surrounding human carrying capacity and overpopulation deserve fresh consideration. Perhaps putting greater resources and efforts into reducing the population can reduce suffering and improve the health outcomes of those in Third World countries.

Page #6

Please type out as much as you can remember about the passage you just read (note form).

[Textbox]

Page #7

1. Why did so many people become concerned about the global population exceeding global food supply in the 1960s?

[Textbox]

2. Why did USA president Ronald Reagan reduce aid to promote family planning in developing countries?

[Textbox]

3. According to the author, what does the Rwandan genocide and HIV/AIDS have in common?

[Textbox]

Page #8

Multiple Choice Questions:

4. United States President Lyndon Johnson ...

- ○ was concerned with the harmful effects of rapid population growth in developing countries
- ○ refused to ship wheat to India during their famine
- ○ decreased foreign aid during his presidency
- ○ did not support family planning initiatives

5. According to the passage, which of the following is not a reason the USA's foreign aid budget fell starting in the 1980s.

- ○ improvements in motor-vehicle safety
- ○ the 1973 "oil price shock"

- ○ improved life expectancy in Third World countries
- ○ stagflation

6. Which of the following statements, if true, would most weaken the author's arguments concerning carrying capacity and overpopulation?

- ○ Vaccines contributed more to increased life expectancy in the 20th century than antibiotics.
- ○ Malthus' views were widely criticized at the time of their publication.
- ○ Historical population reduction is fairly uncommon.
- ○ Historical population reduction is most common when resources exceed population demands.

7. Which of the following would the author most likely consider a modern day "negative check"?

- ○ the rise of national vaccination programs
- ○ widespread famine among sub-Saharan nations.
- ○ escalating violence in the Middle East.
- ○ growing rates of birth control usage

8. The passage author most likely believes the Rwandan Genocide and AIDS crisis are...

- ○ examples of the dangerous consequences of excessive foreign aid.
- ○ less serious than population declines in Haiti, Russia, and North Korea.
- ○ possible consequences of overpopulation.
- ○ not related to one another.

Please see Appendix D for the remaining pages 9-13

Appendix H. Session 2 – No Self-explanation Treatment

Page #1

Welcome to Session 2!

This session will take you approximately 20 mins. You will be asked to watch a video and answer questions about the video. Be assured that all of your responses will be kept confidential. You must complete this study in one sitting. Therefore, you cannot log out and then log back into the system. Do not hit the "Back" button on your browser. Please click 'Next' when you are ready to begin.

Page #2

To receive payment for your participation, please enter the number that was emailed to you.

[Textbox]

Page #3

Please watch this video once

Embedded Video

Page 4

1. How does pollen disseminate?

[Textbox]

2. What are some ways in which flowering plants attract insects?

[Textbox]

3. The flowering plant philodendron selloum raises itself to 115-degrees Fahrenheit, 43-degrees centigrade for two days a year? Why does it do this?

[Textbox]

Page 5

You are almost finished this session!

The next section is a very short questionnaire about you.

	Extremely uncharacteristic of me	Somewhat uncharacteristic of me	Uncertain	Somewhat characteristic of me	Extremely characteristic of me
I would prefer complex to simple problems.	❍	❍	❍	❍	❍
I like to have the responsibility of handling a situation that requires a lot of thinking.	❍	❍	❍	❍	❍
Thinking is not my idea of fun.	❍	❍	❍	❍	❍
I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.	❍	❍	❍	❍	❍
I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.	❍	❍	❍	❍	❍
I find satisfaction in deliberating hard and for long hours.	❍	❍	❍	❍	❍

I only think as hard as I have to.	❍	❍	❍	❍	❍
I prefer to think about small, daily projects to long-term ones.	❍	❍	❍	❍	❍
I like tasks that require little thought once I've learned them.	❍	❍	❍	❍	❍
The idea of relying on thought to make my way to the top appeals to me.	❍	❍	❍	❍	❍
I really enjoy a task that involves coming up with new solutions to problems.	❍	❍	❍	❍	❍
Learning new ways to think doesn't excite me very much.	❍	❍	❍	❍	❍
I prefer my life to be filled with puzzles that I must solve.	❍	❍	❍	❍	❍
The notion of thinking abstractly is appealing to me.	❍	❍	❍	❍	❍
I would prefer a task that is intellectual, difficult, and	❍	❍	❍	❍	❍

important to one that is somewhat important but does not require much thought.					
I feel relief rather than satisfaction after completing a task that required a lot of mental effort.	❍	❍	❍	❍	❍
It’s enough for me that something gets the job done; I don’t care how or why it works.	❍	❍	❍	❍	❍
I usually end up deliberating about issues even when they do not affect me personally.	❍	❍	❍	❍	❍

Page 7

You’re done Session 2! Thank you for participating.

Appendix I. Session 2 – Self-explanation Training

Page #1

Welcome to Session 2!

In this session, you will learn about a new studying technique called "self-explanation". This session will take you approximately 20 mins. Be assured that all of your responses will be kept confidential. You must complete this study in one session. Therefore, you cannot log out and then log back into the system. Do not hit the "Back" button on your browser. Please click 'Next' when you are ready to begin.

Page #2

To receive payment for your participation, please enter the number that was emailed to you.

[Textbox]

Page #3

"**Self-explanation**" is a study strategy you can use to improve your understanding of what you read. Basically, you explain what you are reading to yourself. Self-explanations include: Identifying the main ideas in the text. Expanding on the main ideas of the text and explaining them to yourself. Attempting to explain each sentence in your own words. These may be ideas from previous text or ideas from your own knowledge of the topic area.

Page #4

As you read, you should generate self-explanations that best meet your personal learning needs. Your personal learning needs depend on how much knowledge and understanding you already may have about the topic of the text. Self-explanations can be:

For each sentence   For each paragraph. And across paragraphs

The next section will provide examples of self-explanations generated by a student.

Page #5

Sample Passage:

Surface water currents are largely driven by wind and the Earth's rotation. They commonly exhibit a circular pattern that which is clockwise in the northern hemisphere and counterclockwise in the southern hemisphere. This difference is due to the coriolis effect. The coriolis effect drives the flow of warm surface currents from the tropics to the polar regions via large current flows known as gyres.

Examples of self-explanation by sentence: The first sentence is about what causes surface currents.   What do I already know about the Earth's rotation? I know the Earth completes one rotates in a day.

Page #6

Sample Passage:

Surface currents are largely driven by wind and the Earth's rotation. They commonly exhibit a circular pattern that which is clockwise in the northern hemisphere and counterclockwise in the southern hemisphere. This difference is due to the coriolis effect. The coriolis effect drives the flow of warm surface currents from the tropics to the polar regions via large current flows known as gyres.

Examples of self-explanation by paragraph: The purpose of this paragraph is to explain the coriolis effect. Changes in the temperature of water currents cause gyres. If there are surface currents, are there also deep water currents too?

Page #7

Sample Passage:

Surface currents are largely driven by wind and the Earth's rotation. They commonly exhibit a circular pattern that which is clockwise in the northern hemisphere and counterclockwise in the southern hemisphere. This difference is due to the coriolis effect. The coriolis effect drives the flow of warm surface currents from the tropics to the polar regions via large current flows known as gyres.

Events, such as underwater earthquakes, can also trigger serious currents. Huge storms move water masses and underwater earthquakes may trigger devastating tsunamis. Both move masses of water inland when they reach shallow water and coastlines. Earthquakes may also trigger rapid downslope movement of water-saturated sediments, creating turbidity currents.

Examples of self-explanation across paragraphs:
These two paragraphs are discussing what causes ocean currents
The difference between the two paragraphs is that the first one is about currents that can occur at any time, and the second paragraph is about specific events that trigger current changes.
What do turbidity currents have to do with surface currents?

Page #8

There are many other study strategies. For example, monitoring or **summarizing** are two other examples of strategies. But they are very different from self-explanation.

Page #9

Monitoring - To observe or keep track of one's thought processes.

Sample Passage:

Surface currents are largely driven by wind and the Earth's rotation. They commonly exhibit a circular pattern that which is clockwise in the northern hemisphere and counterclockwise in the southern hemisphere. This difference is due to the coriolis effect.

The coriolis effect drives the flow of warm surface currents from the tropics to the polar regions via large current flows known as gyres.

Examples of monitoring statements: Ok, I understand this passage Hm, maybe I need re re-read the last sentence These are not self-explanation statements because they do not explain the passage.

Page #10

Summarizing - To express the meaning of a sentence using different words.

Sample Passage:

Surface currents are largely driven by wind and the Earth's rotation. They commonly exhibit a circular pattern that which is clockwise in the northern hemisphere and counterclockwise in the southern hemisphere. This difference is due to the coriolis effect. The coriolis effect drives the flow of warm surface currents from the tropics to the polar regions via large current flows known as gyres.

Summary example: Gyres are large current flows that flow from the tropics to the polar regions. This is not a self-explanation. The student just re-arranged the words in the sentence. No additional information is added or linked together.

Page #11

In conclusion:

There are high quality self-explanations and low quality self-explanations. High quality self-explanations:

- identify the main ideas of the passage
- make links to other information
- do not contains monitoring statements or summaries

- include questions that you may have about what you just read
- provide answers for previous questions you may have had about the passage

Page #12

Try to write a self-explanation for this paragraph:

In sociology, the social disorganization theory is one of the most important theories. The theory directly links crime rates to neighborhood ecological characteristics; a core principle of social disorganization theory is that place matters. In other words, a person's residential location is a substantial factor shaping the likelihood that that person will become involved in illegal activities. The theory suggests that, among determinants of a person's later illegal activity, residential location is as significant as or more significant than the person's individual characteristics (e.g., age, gender, or race). For example, the theory suggests that youths from disadvantaged neighborhoods participate in a subculture which approves of delinquency, and that these youths thus acquire criminality in this social and cultural setting.

[Textbox]

Page #13

Consider your Self-explanation and ask yourself if you did the following:

- I did not use monitoring
- I did not summarize
- I made linkages with my previous knowledge
- I explained the main ideas in the passage to myself
- I made linkages between ideas in the passage

- o I asked myself questions about ideas I did not fully understand

In my opinion, the quality of my self-explanation was: (out of 3)

[Textbox]

Page #14

Now, practice what you just learned:

Passage:

Trees in urban landscape settings are often subject to disturbances, whether human or natural, both above and below ground. They may require care to improve their chances of survival following damage from either biotic or abiotic causes. Arborists can provide appropriate solutions, such as pruning trees for health and good structure, for aesthetic reasons, and to permit people to walk under them (a technique often referred to as "crown raising"), or to keep them away from wires, fences and buildings (a technique referred to as "crown reduction"). Timing and methods of treatment depend on the species of tree and the purpose of the work. To determine the best practices, a thorough knowledge of local species and environments is essential.

Drag the following statements on the left to the categories on the right:

	Self-explanation	Monitoring	Summary
I have understood this paragraph	○	○	○
Timing and methods of treatment depend on tree species	○	○	○
Crown raising and crown reduction are two ways that arborists control trees	○	○	○
I wonder how knowing about local species helps arborists?	○	○	○
I finished reading this paragraph.	○	○	○
Trees need care to survive following damage.	○	○	○

Page #15

Below is the correct categorization for the statements

Page #16

You are almost finished this session!

The next section is a very short questionnaire about yourself.

Please see pages 6-7 from Appendix F for the remainder

Appendix J. Session 3 – Self-explanation Generic Prompts

Page #1

Welcome to Session 3!

This is your last session. You will be asked to provide some information, read a passage, and answer questions about that passage. This session should take approximately 20 mins. Be assured that all of your responses will be kept confidential.

You must complete this study in one session. Therefore, you cannot log out and then log back into the system.

Do not hit the "Back" button on your browser.

Please click 'Next' when you are ready to begin.

Page #2

To receive payment for your participation, please enter the number that was emailed to you.

[Textbox]

Page #3

Next, you will read a short passage. Once you've finished reading the passage, you will take a short quiz about the passage. There will be several self-explanation prompts throughout the passage. Please read them carefully and respond to these prompts as you read the passage.

Page #4

Generate self-explanations by sentence, paragraph and across paragraphs as you read this passage. Three boxes are provided throughout the passage. Type out what you've explained to yourself in the boxes.

Traditionally, healthcare has been delivered with a paternalistic approach. Behind this approach is the idea that medical decisions are best left in the hands of those providing healthcare: doctors. Patients present problems to doctors, who make decisions about treatment, prescriptions and further testing. However, the paternalistic approach may threaten patients' ability to make their own health care decisions.

Generally speaking, doctors may not know what matters most to patients regarding their ability to get and stay well. Therefore, a patient's right to make medical decisions, without a health care provider trying to influence the decision, is crucial. The autonomous decision-making approach emphases the right and responsibility of individuals to decide how they live and how they die. Under this approach, doctors are required to ensure patients understand what kind of treatment they are agreeing to, known as "informed consent", and to share with patients all relevant information about their health.

Type out your self-explanation below:

[Textbox]

One main difficulty with autonomous decision making is that it requires patients to have comprehensive and objective information. This can cause problems because patients may struggle to interpret complex figures (such as percentages and probabilities) provided by doctors. For example, they may over or under-estimate the risks or potential benefits of treatment. Also, most patients do not have medical training or the experience to be able to understand the details of their illness. And lastly, patients may be influenced by strong emotions, such as fear and grief. Studies of cancer patients have found that their treatment decisions are not based on a reasonable evaluation of risks and benefits, but instead are based on hope − even when considering a treatment with a high likelihood of failure.

The differences between these two approaches can be made clear through the following example. A doctor is discussing dosage options for hypertension with a patient, when the patient suggests stopping treatment. Under a paternalistic approach, the doctor may exaggerate the possible unpleasant side-effects of this action and state, “you will almost certainly suffer a stroke”, a claim the doctor knows to be untrue. The doctor may justify misleading the patient because treatment is better for the patient in the long term. On the other hand, a doctor with a different philosophy may respect the patient’s wish to stop all treatment.

Type out your self-explanation below:

[Textbox]

An approach known as libertarian paternalism, however, aims to provide a framework where doctors influence choices. In other words, by changing the "choice architecture" for decision making, patients can be "nudged" into making informed choices. For example, a doctor might present several treatment options to her patient, but discuss the treatment the doctor feels is most appropriate in greater depth. Supporters of libertarian paternalism argue that illness may create a temporary loss of patient autonomy caused by physical or mental side effects. As a result, to some degree, paternalistic input is justified to restore an individual’s autonomy.

Critics of libertarian paternalism have suggested that it is perhaps more accurate to consider the nudge techniques a form of "manipulation”. Any attempt to influence the patient will result in a decision that is not truly being made by the patient. However, being a patient with a potentially life-threatening condition puts a considerable strain on one’s ability to process information. This raises the question of whether it is the duty of a doctor to recognize and correct reasoning failure in their patients. If so, would this be considered a paternalistic act?

Type out your self-explanations below:

[Textbox]

Please see pages 6-8 from Appendix A for the remainder.

Page #8

You're done!

Thank you for your participation.

Appendix K. Sample of Free Recall Scoring

Idea #	Idea Unit (Recall)	11101	12201	12102	12202	12103
1	The question of human overpopulation and its relationship to human carrying capacity has been controversial for over two centuries.	1				
2	Human carrying capacity is the planet's limited ability to support its people					
3	In 1798 the Reverend Thomas Malthus came up with a theory	1				
4	Malthus came up with a theory that growth in the number of humans on earth would exceed the growth of resources humans need to live.		2	2		2
5	Malthus predicts too much population growth would lead to a reduction in the number of humans					
6	A reduction in the number of humans by "positive checks"			2		2
7	A reduction in the number of humans by "negative checks"			2		2
8	Positive checks such as disease					2
9	Positive checks such as famine					2
10	Positive checks such as war					2
11	Negative checks, by which Malthus meant restrictions on marriage.					2
12	Malthus's worst fears did not occur during the century following his death					2
13	Malthus died in 1834.					
14	Food production largely kept pace with the slowly growing global population.					
15	Soon after 1934, the global population began to rise quickly					
16	Global population began to rise as vaccines increased the expected lifespan of humans.	2		2		

Appendix L. Sample of Short Answer Question Scoring

ID	Version	1. Why is informed consent important to attain from patients?	Patient understands treatment	Right over body/autonomy	Patient knows what's best for their body	Total
12202	A	So that it is clear, that the patient is well known of the circumstances and it is his decision. It also acts as legal evidence.	1			1
12102	A	Ensures that autonomy has been respected		1		1
12103	A	So they know what they are getting into and the risks of doing or not doing a treatment	1			1
11101	A	Because it is the patient's life and not everyone has the same opinions of how they want to live or die.			1	2
12104	A	patients have a right to be actively involved in their healthcare decisions.		1		1
12105	A	It is important because it provides information in a way that the patient can understand so they can make to best decision for their wants and needs.	1		1	3
12106	A	Because the treatment option they choose have likelihood of altering the quality of their life, so they should have control over decisions made about their body.		1		1
12207	A	Because patients must be informed of what will happen to them, and they must agree to what will happen if they are conscious and of sound mind.		1		1