IMPACT OF SELF-EXPLANATION AND ANALOGICAL COMPARISON SUPPORT ON LEARNING PROCESSES, MOTIVATION, METACOGNITION, AND TRANSFER

by

J. Elizabeth Richey

B.A., B.S., University of Pittsburgh, 2006

M.S., University of Pittsburgh, 2011

Submitted to the Graduate Faculty of the

Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

J. Elizabeth Richey

It was defended on

May 13, 2015

and approved by

Kevin Crowley, Professor, Learning Sciences and Policy, University of Pittsburgh

Kenneth R. Koedinger, Professor, Human-Computer Interaction Institute, Carnegie Mellon University

Christian D. Schunn, Professor, Department of Psychology, University of Pittsburgh

Dissertation Advisor: Timothy J. Nokes-Malach, Associate Professor, Department of Psychology, University of Pittsburgh

Impact of Self-Explanation and Analogical Comparison Support on Learning Processes, Motivation, Metacognition, and Transfer

J. Elizabeth Richey, PhD

University of Pittsburgh, 2015

IMPACT OF SELF-EXPLANATION AND ANALOGICAL COMPARISON SUPPORT ON LEARNING PROCESSES, MOTIVATION, METACOGNITION, AND TRANSFER

J. Elizabeth Richey, PhD

University of Pittsburgh, 2015

Research examining analogical comparison and self-explanation has produced a robust set of findings about learning and transfer supported by each instructional technique. However, it is unclear how the types of knowledge generated through each technique differ, which has important implications for cognitive theory as well as instructional practice. I conducted a pair of experiments to directly compare the effects of instructional prompts supporting self-explanation, analogical comparison, and the study of instructional explanations across a number of fine-grained learning process, motivation, metacognition, and transfer measures. Experiment 1 explored these questions using sequence extrapolation problems, and results showed no differences between self-explanation and analogical comparison support conditions on any measure. Experiment 2 explored the same questions in a science domain. I evaluated condition effects on transfer outcomes; self-reported self-explanation, analogical comparison, and metacognitive processes; and achievement goals. I also examined relations between transfer and self-reported processes and goals. Receiving materials with analogical comparison support and reporting greater levels of analogical comparison were both associated with worse transfer performance, while reporting greater levels of self-explanation was associated with better performance. Learners' self-reports of self-explanation and analogical comparison were not related to condition assignment, suggesting

that the questionnaires did not measure the same processes promoted by the intervention, or that individual differences in processing are robust even when learners are instructed to engage in self-explanation or analogical comparison.

TABLE OF CONTENTS

PREFACE XIV

1.0 INTRODUCTION 1

1.1 SELF-EXPLANATION 5

1.1.1 Mechanisms that support learning through self-explanation 6

1.1.2 How has self-explanation been studied? 7

1.1.3 Types of knowledge that self-explanation supports 11

1.2 ANALOGICAL COMPARISON 13

1.2.1 Mechanisms that support learning through analogical comparison 13

1.2.2 How has analogical comparison been studied? 14

1.2.3 Types of knowledge that analogical comparison supports 16

1.3 WORKED EXAMPLES 17

1.4 INSTRUCTIONAL EXPLANATIONS 18

1.5 METACOGNITION 19

1.6 MOTIVATION 21

1.7 FRAMEWORK FOR COMPARING INSTRUCTIONAL TECHNIQUES 24

1.7.1 Prior work comparing self-explanation and analogical comparison 25

1.7.2 Current studies 26

2.0 EXPERIMENT 1: SEQUENCE EXTRAPOLATION THROUGH ANALOGICAL COMPARISON, SELF-EXPLANATION, INSTRUCTIONAL EXPLANATION, AND PROBLEM-SOLVING PRACTICE 30
3.0 METHOD 32
3.1 PARTICIPANTS 32
3.2 DESIGN 32
3.3 MATERIALS 32
3.3.1 Instructional text 32
3.3.2 Training and example problems 33
3.3.2.1 Self-explanation 35
3.3.2.2 Analogical comparison 36
3.3.2.3 Instructional explanation 38
3.3.2.4 Practice 39
3.3.3 Test materials 39
3.3.4 Questionnaires 42
3.3.4.1 Task-framed Achievement Goals Questionnaire 42
3.3.4.2 Learning processes questionnaire 43
3.4 PROCEDURE 44
4.0 RESULTS 45
4.1.1 Learning materials completion and accuracy 45
4.1.2 Test accuracy 47
4.1.3 Solution times 49
4.1.4 Detection times 50

4.1.5 Extrapolation times ... 52
4.1.6 Completion patterns ... 53
4.1.7 Error correction ... 54
4.1.8 Achievement goals ... 55
5.0 DISCUSSION ... 59
6.0 EXPERIMENT 2: SELF-EXPLANATION AND ANALOGICAL COMPARISON OF SCIENCE MATERIALS ... 63
7.0 METHOD ... 67
7.1 PARTICIPANTS ... 67
7.2 DESIGN ... 67
7.3 MATERIALS ... 68
7.3.1 Learning materials ... 68
7.3.2 Test materials ... 70
7.3.3 Questionnaires ... 72
7.3.3.1 Open-ended learning processes question ... 72
7.3.3.2 Learning processes questionnaire ... 73
7.3.3.3 Metacognitive questionnaire ... 73
7.3.3.4 Achievement goals questionnaires ... 74
7.4 PROCEDURE ... 75
8.0 RESULTS ... 76
8.1.1 Assessment of learning quality ... 76
8.1.2 Condition effects on learning ... 82
8.1.3 Open-ended learning processes question ... 85

8.1.4 Learning processes questionnaire ... **88**

8.1.5 Task-framed achievement goal surveys ... **91**

8.1.6 Science class-framed surveys ... **93**

9.0 DISCUSSION ... **95**

10.0 CONCLUSION ... **103**

APPENDIX A ... **107**

APPENDIX B ... **117**

BIBLIOGRAPHY ... **120**

LIST OF TABLES

Table 1. *Levels of abstraction targeted by learning and test material problems with a simple demonstration example. For all examples, base for transfer is ABAB.* 33

Table 2. *Examples and test items used in Experiment 1.* 35

Table 3. *Targeted knowledge and predictions for each measure in Experiment 1.* 41

Table 4. *Learning processes questionnaire items and the level at which they were framed.* 43

Table 5. *Rubric parameters and examples of prompt responses directed at the worked example shown in Figure 12 for scores 1-4.* 77

Table 6. *Rubric parameters and examples of prompt responses directed at the worked example shown in Figure 13 for scores 1-4.* 80

Table 7. *Correlation (r) between prompt response scores, average word count, and test performance by condition.* 82

Table 8. *Mean and standard deviation of pretest and learning packet accuracy by condition.* ... 83

Table 9. *Correlations between cognitive, metacognitive, and goal word use and posttest performance.* 87

Table 10. *Summary of multiple regression analyses for self-reported learning process use predicting subsets of the posttest.* 90

Table 11. *Correlations between cognitive, metacognitive, and goal word use and self-reported responses on the strategies questionnaire.* 90

Table 12. *Summary of multiple regression analyses for task-framed achievement goals predicting subsets of the posttest.* 92

Table 13. *Summary of multiple regression analyses for science class-framed achievement goals predicting subsets of the posttest.* 94

Table 14. *LIWC dictionary words for cognitive processes.* 117

Table 15. *LIWC dictionary words for metacognitive processes.* 118

Table 16. *LIWC dictionary words for achievement goals.* 118

LIST OF FIGURES

Figure 1. Sample sequence extrapolation pattern with rules labeled. 30

Figure 2. Worked example steps formatted for the self-explanation condition. 36

Figure 3. Worked examples steps formatted for the analogical comparison condition. Other conditions received the same steps separately, one example sequence at a time. 37

Figure 4. Sample of instructional explanation paired with worked example in the instructional explanation condition. 38

Figure 5. Accuracy by problem type across conditions. 49

Figure 6. Solution times in seconds by problem type across conditions. 50

Figure 7. Detection times in seconds by problem type across conditions. 51

Figure 8. Detection times in seconds by problem type across conditions. 52

Figure 9. Mean number of changes to final extrapolation pattern by condition 55

Figure 10. Condition effects on achievement goal endorsement. 57

Figure 11. Worked example with instructional explanations. 71

Figure 12. Worked example that served as the target of prompt responses included in Table 5. 77

Figure 13. Worked example that served as the target of prompt responses included in Table 6. 79

Figure 14. Posttest accuracy on subsets of items classified by transfer distance. 84

Figure 15. Use of cognitive and metacognitive processes and goal language. 87

Figure 16. Endorsement of cognitive and metacognitive processes by condition. 89

Figure 17. Endorsement of task-framed achievement goals by condition. 91

Figure 18. Endorsement of science class-framed achievement goals by condition. 93

PREFACE

I would like to thank those who have mentored me throughout graduate school, most especially my advisor, Dr. Timothy Nokes-Malach, and committee members, Drs. Christian Schunn, Kevin Crowley, and Kenneth Koedinger, for their time, wisdom, and support. I'm grateful for the Cognitive Science Learning Laboratory, the Higher Order Cognitive Collective, and the Pittsburgh Science of Learning Center, and all the people who have made those communities an incredible place for exploring ideas. I'm indebted to Drs. Michael Sayette, Thomas Kirchner, and Kasey Creswell for first fostering and nourishing my interest in psychology research during my time as an undergraduate at Pitt. I'm thankful for the support, inspiring discussions, and invaluable assistance provided by my labmates and the dedicated undergraduates who have contributed to this work, including Dr. Joel Chan, Cristina Zepeda, Kelly Boden, Sarah Honsaker, and Aleza Wallace. Finally, I'm grateful to all the educators who have taken time to share their experiences and insights with me and especially to Paul Ronevich and Brian Rose for their years of close collaboration across many projects.

On a personal note, I'd like to thank my husband, Dan, for his love, encouragement, and patience; my daughter, Sarah, for the joy and inspiration she has brought to our lives; and our soon-to-arrive daughter, Lilah, who has already provided tremendous support by not making an early debut. I also thank my parents, Bob and Janie Strohm, for the unconditional love and guidance that have shaped my life, and the friends who have made life richer along the way.

1.0 INTRODUCTION

One goal of cognitive science is to examine the cognitive processes that support learning and transfer, or the application of knowledge to a new situation or problem. Analogical comparison and self-explanation are hypothesized to be two constructive, sense-making instructional techniques for acquiring knowledge that transfers (Chi, 2009; Koedinger, Booth, & Klahr, 2013; Renkl, 2014; Richey & Nokes-Malach, 2015), and both have shown consistent benefits for learning in the laboratory as well as the classroom (for reviews, see Alfieri, Nokes-Malach, & Schunn, 2013; Fonseca & Chi, 2011). Both analogical comparison and self-explanation support learning of relational features and transfer to new contexts (Ainsworth & Burcham, 2007; Alfieri et al., 2013; Atkinson, Renkl, & Merrill, 2003; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Crowley & Siegler, 1999; Dellarosa, 1985; Gentner, Loewenstein, & Thompson, 2003; Gick & Holyoak, 1983; Kurtz, Miao, & Gentner, 2001; Rittle-Johnson, 2006). While they appear to rely on some of the same mechanisms (e.g., inference generation), prior work suggests they also operate through different mechanisms (e.g., mental model revision in self-explanation versus relational abstraction in analogical comparison) and the exact nature of the knowledge acquired through each is not clear. For example, how does each instructional technique promote procedural versus declarative knowledge acquisition, and how abstract is the knowledge acquired? The answers have important implications for when students should explain or compare in class, particularly if there are instructional scenarios to which one approach is better suited than the other.

Comparing the two techniques could provide answers to these questions. While each instructional technique has been studied extensively on its own, little work has directly compared the two (Renkl, 2014; cf. Edwards, Williams, Gentner, & Lombrozo, 2014; Gadgil, Nokes-Malach, & Chi, 2012; Nokes-Malach et al., 2013). Further, the wide variety of tasks, amount of scaffolding, and measurement employed in prior work make it difficult to compare across experiments examining each technique separately. Consequently, there is little evidence available to suggest which instructional technique is most appropriate for particular instructional goals (e.g., near or far transfer). The relative dearth of prior comparative work combined with the strong foundations established in the separate literatures examining analogical comparison and self-explanation pose an opportunity for testing cognitive theories about both processes and taking the first steps toward creating an instructional theory that incorporates both. By integrating literatures on analogical comparison and self-explanation, I seek to identify differences in the types of knowledge each supports. I also investigate measures designed to capture individual differences in the degree to which learners engage in these learning processes, both spontaneously and when prompted. Direct comparison creates opportunities to identify differences, if any, in the knowledge representations acquired through each process, and this approach could provide evidence to guide instructors in determining which processes to support for certain situations.

It is also possible that instructions to engage in either analogical comparison or self-explanation promote use of the other process as well (Edwards et al., 2014; Edwards, Williams, & Lombrozo, 2013; Edwards, 2014; Neuman & Schwarz, 1998). For example, comparison often involves explicit explanation of features and their relations within examples, and self-explanation invites comparisons between prior knowledge and new information or different pieces of information. For these reasons, it is interesting to explore not only the types of knowledge supported by instructing learners to engage in either self-explanation or analogical comparison, but also the degree to which students report engaging in both of these processes after receiving either type of instruction. Thus, the current work aims to directly compare learning outcomes when participants receive either self-explanation or analogical comparison support, while also exploring new questionnaire measures targeting individual differences in self-explanation and analogical comparison.

I conducted two experiments to assess what is learned and transferred when participants are encouraged to engage in either self-explanation or analogical comparison of the same worked examples. Experiments 1 and 2 compared knowledge acquired through self-explanation and analogical comparison on a series of near-, intermediate-, and far-transfer tasks. Experiment 1 employed a sequence extrapolation task, which was hypothesized to contain sufficiently rich principles to capture cognitive processes and types of knowledge representations similar to those employed in academic learning while also affording greater experimental control of knowledge types and test items. Experiment 2 aimed to explore the same questions in an ecologically valid academic domain. Both experiments examine tasks designed to be equally well suited to self-explanation and analogical comparison, with similar levels of scaffolding and support offered for each instructional technique to facilitate clearer comparison of learning processes and outcomes.

In the following sections, I review prior research on self-explanation and analogical comparison, highlighting findings regarding the types of knowledge and performance each one supports and the hypothesized mechanisms thought to explain those findings. I discuss the key features compared across instructional techniques in the present work, and then provide conceptual background for the other variables of interest, including metacognition and motivation. My examination of these techniques is grounded in the literature on example-based learning (Renkl, 2014), with instructional prompts for each condition targeting worked examples and a worked example condition with instructional explanations used as a third condition across studies.

This work adopts Barnett and Ceci's (2002) model of transfer distance, which accounts for two dimensions of distance, content and context, to determine where a task lies on the transfer spectrum in relation to a knowledge source. Transfer can occur across a number of different factors within these dimensions (e.g., knowledge domain, temporal context) and its distance is a reflection of how different the transfer task is from the initial learning task. Different types of transfer are associated with different types of knowledge representations and their characteristics. For example, performing well on a near-transfer task suggests the learner has acquired procedures and concrete features of the initial task, while performing well on a far-transfer task suggests the learner has acquired abstract principles that can be flexibly applied to new features. This makes transfer an especially useful measure for understanding and differentiating the nature of the knowledge representations acquired across conditions.

In the following experiments, I classify test items based on Barnett and Ceci's (2002) and Nokes-Malach et al.'s (2013) models of transfer distance with categories of near, intermediate, and far transfer, as well as Bransford and Schwartz's (1999) preparation for future learning. Items classified as near transfer involve either identical problem-solving procedures or the same surface features or relations (e.g., identical phrases of text) that can cue memory retrieval of prior learning materials. Intermediate-transfer items involve the same representations (e.g., structures, principles, problem-solving procedures) but may require some abstraction or inference based on changes in surface features. Far-transfer items require the learner to draw inferences and reason about new problem features or relations by applying concepts or principles from prior learning materials. Finally, preparation for future learning (PFL) is a transfer measure that looks at how well a learner is able to acquire and apply knowledge from a new learning resource after initial instruction (Bransford & Schwartz, 1999; Schwartz et al., 2005; Schwartz & Martin, 2004). In other words, while the other transfer measures look at the application of knowledge to a new question, PFL transfer looks at the application of knowledge to a new learning opportunity. Knowledge must be increasingly abstract and flexible to facilitate transfer as distance increases, meaning that more concrete knowledge should support near transfer while more abstract knowledge should support far transfer and PFL.

1.1 SELF-EXPLANATION

Self-explanation is a powerful, well-studied learning process that involves generating explanations of provided content (e.g., worked examples, instructional text; Atkinson et al., 2003; Catrambone, 1998; Fonseca & Chi, 2011; Rittle-Johnson, 2006). It is a particular type within the broader category of explanations (e.g., instructional explanations, disciplinary explanations; Keil, 2006; Leinhardt, 2010). While all explanations share certain features, such as generating answers to questions (Edwards, 2014), each type has specific features with implications for how they arise, the mechanisms through which they promote learning, and how they can be used in instruction. In this section, I review the features of self-explanations, which are explanations that learners generate for themselves.

1.1.1 Mechanisms that support learning through self-explanation

Because they are generated with minimal concerns about external evaluation or coherence, self-explanations are often incomplete and tailored to the learner's prior knowledge. While this can make self-explanations less accurate according to disciplinary standards, it also means that they are especially useful for building and revising a learner's existing knowledge about a domain. Several theories have proposed that self-explanation operates through gap-filling and the correction of errors in mental representations of complex concepts (Chi, 2000; Nokes, Hausmann, VanLehn, & Gershman, 2011; VanLehn & Jones, 1993). The benefits of self-explanation seem to be critically linked to the prior or personal knowledge of the learner and the process of activating that knowledge to construct explanations (Fonseca & Chi, 2011). For example, Hausmann & VanLehn (2010) compared learning outcomes for college students instructed either to self-explain or paraphrase the content of the examples they were shown. To help students better follow the manipulation instructions, both conditions received brief instruction about the nature of self-explanation or paraphrasing and studied student examples modeling the behavior. Students instructed to self-explain demonstrated greater learning during the intervention, as measured by their proportion of errors and requests for help out of all their problem-solving entries. They also required less assistance on related homework problems solved after the training intervention, suggesting the instructions to self-explain supported deeper learning that facilitated performance after the intervention had ended. These results indicate that self-explanation operates not through exposure to additional content, but through the act of generating content.

A primary theoretical argument about the nature and utility of self-explanation suggests that learning comes from the process of working through one's own understanding in relation to new information; based on this explanation, it is less important that the explanations be accurate or complete and more important that they involve one's prior knowledge or understanding. Thus, Chi (2000) argued that inaccurate self-explanations should not necessarily reduce learning. Prior work has shown that elaborations are associated with better comprehension even when they are not accurate (McNamara, 2004), although other results suggested that inaccurate elaborations were associated with less procedural learning from examples (Berthold & Renkl, 2009).

1.1.2 How has self-explanation been studied?

Self-explanation has primarily been studied through examination of spontaneous behaviors and measurement of learning following instructional interventions designed to either train or prompt self-explanation. Learners who engage in more self-explanation on their own demonstrate better understanding of principles compared to those who engage in less self-explanation, suggesting that prompting explanation might be a productive avenue for improving learning (Chi et al., 1989). Even when explanations are prompted, however, there is still high variability in self-explanation quality (Chi, De Leeuw, Chiu, & Lavancher, 1994), suggesting that prompting for explanations may not be very effective on its own in promoting the types of deep, relational explanations that support robust learning. For example, in a pilot study using a cognitive tutor that prompted high school geometry students to explain example problem-solving steps, Aleven and Koedinger (2000) reported low success on a variety of measures including attempts at explanation, accuracy of explanations, and completeness. The reasons for low-quality responses could include low student motivation, uncertainty about how to respond to the prompts, limited prior knowledge or instructional content available as a source for explanations, and the inherent difficulty of the task. Generating novel explanations is more difficult and errorprone than restating information provided through instruction, and so learners may be more inclined to paraphrase or generate shallow explanations.

A number of interventions have tried to address these issues by employing more robust self-explanation prompting. Interventions to promote self-explanation can be broadly classified as instructional training in self-explanation (Bielaczyc, Pirolli, & Brown, 1995; McNamara, 2004) or specific prompts to encourage self-explanation (Atkinson et al., 2003; Chi et al., 1994; Crowley & Siegler, 1999; Hausmann, van de Sande, & VanLehn, 2008; Rittle-Johnson, 2006). The former generally entails instruction about different types of self-explanations and their utility, modeling of effective self-explaining, and practice generating self-explanations. Training can vary in length from a short tutorial at the start of the lesson (e.g., Hausmann & VanLehn, 2010) to an extensive series of sessions spanning multiple days.

Prompting may include some basic practice, modeling, or instruction about the nature of self-explanations, but it primarily relies on instructions embedded within learning materials to promote self-explanation. Prompts can be specifically tailored to the learning context (e.g., "Can you say something about the functions of different parts" in a learning activity about the circulatory system; Gadgil et al., 2012) or much more general (e.g., "Explain aloud the reasoning or justification for each step of the solution" in a learning activity about physics problem solving; Nokes-Malach et al., 2013). Some are open-ended and encourage elaboration, as in the previous examples, while other prompts require students to provide the name of a principle or to select a principle or another short response from the a list of options (e.g., Aleven & Koedinger, 2002; Atkinson et al., 2003). Others combine different types of prompts; for example, Berthold and Renkl (2009) employed very detailed prompts targeting specific elements of worked examples early on, and then removed scaffolding by switching to open-ended prompts as learners continued to study examples. While the previous examples targeted the information provided in the instructional materials, some prompts are focused more directly on monitoring the learner's understanding (e.g., "What parts of this page are new to me? Is there anything I still don't understand?"; Wong, Lawson, & Keeves, 2002), and others prompt activities such as explaining how model students arrived at correct or incorrect answers (e.g., "the experimenter asked the participant to explain verbally both how the other child had obtained the answer and why each answer was correct or incorrect," Rittle-Johnson, 2006).

Sometimes learners are prompted to explain verbally (Ainsworth & Burcham, 2007; Chi et al., 1994; Hausmann, Nokes, VanLehn, & van de Sande, 2009; Neuman & Schwarz, 1998; Nokes-Malach et al., 2013; Renkl, Stark, Gruber, & Mandl, 1998; Rittle-Johnson, 2006; Wong et al., 2002), while other times they are prompted to write out their explanations (Berthold & Renkl, 2009; Schworm & Renkl, 2006, 2007). Some evidence suggests that the two processes might trigger different results. Hausmann and Chi (2002) found that students typed fewer spontaneous, unprompted self-explanations and more paraphrasing compared to results in prior research examining spontaneous verbal self-explanations. They proposed that the process of typing may have led participants to filter their thoughts and focus on writing complete, accurate notes, or that the act of typing required more cognitive resources than talking aloud and thus left fewer resources available for the constructive processes of self-explanation. However, another study comparing spontaneous self-explaining and paraphrasing across modalities found that participants did less paraphrasing when typing compared to talking aloud (Muñoz, Magliano, Sheridan, & McNamara, 2006). Given that relatively little work has examined the effects of verbal and written modalities on learning from self-explanation, and those results have been contradictory, it is unclear whether prompting learners to verbalize or write their explanations affects learning.

Interventions also vary in terms of whether learners receive corrective feedback or repeated questioning following incomplete or inaccurate self-explanations (Ainsworth & Loizou, 2003; Aleven, Popescu, & Koedinger, 2002; Conati & VanLehn, 2000; Neuman & Schwarz, 1998), or no follow-up prompting beyond the initial prompt to self-explain (Schworm & Renkl, 2007). For example, Aleven et al.'s (2002) cognitive tutor assessed the completeness and accuracy of learners' self-explanations. In response to incomplete or inaccurate explanations, the tutor would provide an additional prompt to foster a more complete or accurate explanation, such as "Please add to your explanation or type something different" or "That may not be the right number. Can you state the geometry rule that justifies your answer?" In cases where the experimenter provides the feedback or questioning, it is typically used when learners fail to meet a certain set of criteria with their explanations, such as being incorrect, leaving out important points, or remaining silent for too long (Ainsworth & Burcham, 2007; Chi et al., 1994; Neuman & Schwarz, 1998; Renkl et al., 1998). Both approaches produce similar effects of encouraging more elaborate explanations.

Measurement of self-explanation is generally based on verbal or written protocol analysis (e.g., asking participants to talk aloud while studying material and then coding each utterance based on whether it is a particular type of self-explanation statement) and outcome measures including performance on tests and revision of inaccurate mental models or misconceptions. A number of studies have attempted to classify the types of responses students produce, either when simply instructed to talk aloud (e.g., Hausmann, Nokes, VanLehn, & van de Sande, 2009) or when prompted to self-explain (e.g., Berthold & Renkl, 2009). Although the specific categories and features classified have varied across research, most focus on differentiating several dimensions including the degree to which the learner mentions prior knowledge or examples, the degree to which the learner explains compared to paraphrasing, and the degree to which he or she engages in metacognitive behaviors, typically in the form of monitoring understanding. For example, Hausmann et al. (2009) created a coarse-grained coding scheme to analyze individuals’ and dyads’ dialogues while studying examples based on how frequently they explained, monitored their understanding, mentioned prior knowledge or examples, and paraphrased the content of the example they were studying. Berthold & Renkl (2009) coded learners’ written responses to self-explanation prompts according to whether they were based on identifying and describing underlying principles; providing a rationale for why a principle applied; or applying a misunderstanding about principles. Ainsworth and Burcham (2007) developed a coding scheme that was based on Renkl's (1997) work and classified explanations as referencing underlying principles, referencing a goal or purpose for the information, elaborating on the information, assessing the coherence between a new piece of information and previous information, and monitoring understanding. Identifying the different types of information learners generate when constructing explanations has provided some evidence about which types of explanations support learning and consequently provides clues regarding the mechanisms that make self-explanation a robust learning process.

1.1.3 Types of knowledge that self-explanation supports

The types of knowledge supported by self-explanation depend on both the content being explained and the types of explanations the learner generates. Berthold and Renkl (2009) examined correlations between their classification of principle-based, rationale-based, and incorrect self-explanations and conceptual and procedural knowledge. The authors found that principle-based explanations were correlated positively with only conceptual knowledge, incorrect self-explanations were correlated negatively with only procedural knowledge, and rationale-based self-explanations were correlated positively with both. Using a different coding scheme, Ainsworth and Burcham (2007) found that explanations focused on principles, monitoring positive understanding, and paraphrasing were positively correlated with posttest performance, while explanations focused on goals, elaborations on the text, noticing coherence in the text, and monitoring a lack of understanding were not correlated with performance. Only false explanations were negatively correlated with performance. This evidence suggests complex relations between the content of explanations and learning, but given the many different coding schemes and learning outcomes measured across studies, it is difficult to derive clear distinctions about which types of self-explanation support different kinds of knowledge.

Self-explanation prompts have been applied to a wide variety of instructional materials, including text (Chi et al., 1994), diagrams (Ainsworth & Loizou, 2003), and example problems (Rittle-Johnson, 2006). While much work has focused on deepening conceptual understanding (Ainsworth & Loizou, 2003; Atkinson et al., 2003; Chi et al., 1989; Gadgil et al., 2012), other work has emphasized learning procedures, generalizing them, or connecting them to concepts (Aleven, Koedinger, & Popescu, 2003; Crowley & Siegler, 1999; Rittle-Johnson, 2006). Some evidence suggests that the effects of self-explanation are most powerful on difficult (Chi et al., 1994) or far-transfer problems (Wong, Lawson, & Keeves, 2002). There is also evidence of the learning process itself transferring in the form of self-explanation training or prompting changing students' behaviors in later problem-solving scenarios (Wong, Lawson, & Keeves, 2002).

1.2 ANALOGICAL COMPARISON

Analogical comparison is an instructional technique in which learners receive multiple exemplars and engage in mapping features and relations between them, which leads to better encoding of abstract relations that can be applied to novel cases (Gentner et al., 2003; Gick & Holyoak, 1983). Analogical comparison is a particular type of analogical reasoning, which entails applying relations from one case or multiple cases to a novel one; however, retrieving an appropriate case upon which to base the analogy is a major stumbling block in the process (Blanchette & Dunbar, 2000; Holyoak & Koh, 1987; Ross & Kilbane, 1997; Sternberg, 1977). In the case of analogical comparison, which I review here, the two cases or examples are provided. This eliminates potential retrieval failures, and the learning process focuses on identifying features and relations within and across the two cases that can then be used to understand novel cases. In the following sections, I review the mechanisms through which analogical comparison is hypothesized to support learning and the approaches used to examine and test learning through analogical comparison. Based on this review, I identify the types of knowledge analogical comparison may support, considering both theoretical claims and experimental evidence. Through this review, I aim to identify the ways in which analogical comparison might support learning and the critical features necessary to obtain positive learning outcomes.

1.2.1 Mechanisms that support learning through analogical comparison

Analogical case comparisons that are designed to highlight shared structural features and relations can help a learner encode those structural features and relations without the distracting surface features that are also present in individual cases (Gentner et al., 2003). A pair of “contrasting cases” should be carefully selected to highlight similarities that are relevant to principles shared across cases while reducing emphasis on surface features by minimizing superficial similarities (Gick & Paterson, 1992; Schwartz & Bransford, 1998). A schema that captures abstract, structural details without superficial features can facilitate transfer to novel problems with different surface details but the same structural features, making it more likely that a learner who has engaged in case comparison will be able to apply the same principles to a new problem in a different context (Gick & Holyoak, 1983).

1.2.2 How has analogical comparison been studied?

Much work on analogical comparison has examined the degree to which learners engage in it spontaneously and the degree to which it can be successfully prompted or trained. Although the learning benefits of analogical comparison have been extensively demonstrated, research has shown that there is a great deal of individual variability in the extent to which learners engage in analogical comparison spontaneously, and many learners do not make fruitful comparisons even when instructed to do so (Gick & Holyoak, 1983).

Comparisons may be prompted, meaning the learners receive some kind of clue or hint that they should identify similarities or differences but no direction regarding which features they should focus on. Scaffolding is a more robust approach to prompting, in which learners are not only instructed to compare but given clues or explicit instructions about which features and relations to align across cases (Gentner et al., 2003). Many prompts take the form of a hint or direct instruction to compare (Gick & Holyoak, 1983), identify similarities and/or differences between cases (Edwards et al., 2013; Loewenstein, Thompson, & Gentner, 1999), select a match out of a set of options (Gentner & Namy, 1999), or rate similarity (Markman & Gentner, 1993).

Less direct prompts and activities can also be effective in encouraging the learner to engage in comparison (Gentner & Medina, 1998), including instructions to elaborate on relations, providing perceptual cues to align features or relations (Kolodner, 1997), or using names or labels to highlight categories, features, or relations (Graham, Namy, Gentner, & Meagher, 2010; Namy & Gentner, 2002). Multiple strategies from prior work were incorporated into the materials in the present set of experiments to promote analogical comparison, including prompting, modeling, and perceptual alignment.

While analogical comparison can be a powerful tool for deep learning, a meta-analysis of analogical case comparison conducted by Alfieri et al. (2013) showed that several design factors were significant moderators of learning outcomes across studies. Specifically, instructing learners to focus on similarities was more effective than focusing on both similarities and differences, which was more effective than focusing on differences alone. Direct instruction on the key principle across cases *after* the learner engaged in comparison was more effective than identifying it before or not at all. Studies tended to report larger effect sizes when examining perceptual content compared to procedural content, and when the test was administered immediately after the comparison rather than on a subsequent day. In contrast, results did not vary between prompted and guided experiments, generated or provided features, or classroom or laboratory studies. The ages and experience levels of the learners also did not moderate results across studies, suggesting that participants' levels of prior knowledge may not determine their learning outcomes from analogical comparison. The authors found some evidence of a significant difference in results depending on whether the domain was math, science, or another domain, whether the dependent measure targeted near or far content, and whether the cases were rich or minimal, but results were inconsistent. Based on prior literature, it seems less important that learners compare a particular type of information and more important that the information being compared shares critical features and relations, reduces emphasis on surface similarities, and provides sufficient scaffolding to help the learner focus on important features and relations.

1.2.3 Types of knowledge that analogical comparison supports

Measurement of analogical comparison is generally based on transfer outcomes (e.g., if the learner successfully transfers an idea from provided examples to a new problem, one may infer that analogical comparison has been successful), responses to scaffolded prompts (e.g., a participant might be asked to fill in specific features and relations as they align across examples), and performance on assessments of learning (Nokes-Malach et al., 2013; Rittle-Johnson & Star, 2007) or misconceptions (Brown & Clement, 1989; Gadgil et al., 2012). Prior work has shown that comparison can support learning of procedures (Rittle-Johnson & Star, 2007), structural features and relations (Dellarosa, 1985; Kurtz et al., 2001), and abstract principles (Gentner et al., 2003; Novick & Holyoak, 1991). Comparing features and relations supports the encoding of abstract information, which may make analogical comparison especially well-suited for supporting transfer of knowledge to novel contexts or problems (Alfieri, Nokes-Malach, & Schunn, 2013; Gentner et al., 2003; Gick & Holyoak, 1983). Conversely, some evidence has suggested that it may not be as beneficial as other types of instruction, including self-explanation and worked-example study, for facilitating knowledge of specific problem-solving procedures (Nokes-Malach et al., 2013). Comparison can also support the correction of misconceptions or revision of mental models if cases for comparison are selected to highlight learners' misconceptions, either by asking learners to compare a scientifically accurate model to a flawed model that they hold (Gadgil et al., 2012) or by helping the learner to realize that two superficially different cases are instantiations of the same principles (Brown & Clement, 1989).

1.3 WORKED EXAMPLES

Across both studies in the current work, the target of learners' analogical comparisons and self-explanations were worked examples. Worked examples are problems that include problem-solving steps and final solutions, and may contain instructional explanations of the steps or solutions. Prior work has shown that providing learners with worked examples instead of simply giving them problem-solving practice can help them acquire a greater amount of procedural and conceptual knowledge and learn more efficiently (Burns & Vollmeyer, 2002; Fong & Nisbett, 1991; Renkl, 1997, 2005, 2014; Sweller, 1988). Worked examples have been shown to support learning across a variety of math and science domains, and theoretical accounts argue that they operate through several mechanisms. First, they provide constraints on the solution space by highlighting the correct solution path so students do not waste time on incorrect or unfruitful searches or encode incorrect solution strategies (Paas, 1992). Second, they are hypothesized to reduce irrelevant cognitive load, or the amount of cognitive resources a learner employs during a task, by highlighting the important elements of the problem and solution for the learner to focus on, encode, and reason about (Chandler & Sweller, 1991; Paas & Van Merriënboer, 1994; Ward & Sweller, 1990). Third, by virtue of reducing cognitive load and increasing available working memory resources, they are hypothesized to encourage constructive cognitive processes such as self-explanation, in which learners explain to themselves the underlying conceptual logic and justifications behind each step (Catrambone, 1998; Chi et al., 1989; Renkl, 1997, 2014).

Worked examples have often been used as the target of prompts and scaffolding to induce analogical comparison (e.g., Gerjets, Scheiter, & Schuh, 2008) and self-explanations (e.g., Atkinson et al., 2003; Renkl, 2002). The lower cognitive load associated with worked examples means learners are better able to devote sufficient cognitive resources necessary for identifying, aligning, and drawing inferences across structural and relational features or generating deep explanations and revising existing knowledge.

1.4 INSTRUCTIONAL EXPLANATIONS

Worked examples have also been combined with a different type of explanation called instructional explanations, which provide explanations of the concepts, principles, or procedures applied in the worked example (Leinhardt, 2010; Renkl, 2002; Wittwer & Renkl, 2008). While instructional explanations lack some of the advantages of self-explanations, including the constructive element of self-explaining and their incorporation of learners' prior knowledge, they offer the potential for greater disciplinary accuracy and completeness than learners may be able to generate on their own, which may better address learners' comprehension issues or help them recognize errors in their understanding (Renkl, 2002). Instructional explanations therefore might be particularly beneficial in cases in which students lack sufficient prior knowledge to generate fruitful explanations or comparisons, or in situations in which inaccurate or incomplete self-explanations reduce learning (e.g., Berthold & Renkl, 2009). Their utility likely depends on factors concerning the content of the instructional explanations, such as whether they focus on deep principles or restatement of procedural steps, and on how they are implemented, such as whether learners are given opportunities to apply instructional explanations to new problems after receiving them (Wittwer & Renkl, 2008, 2010).

Evidence regarding the efficacy of instructional explanations has been mixed, with some experiments showing an advantage when they are used to support self-explanation (Renkl, 2002) or delivered in a computer-based tutoring environment (Atkinson, 2002) and other experiments showing reduced or equivalent learning compared to simply studying worked examples (Richey & Nokes-Malach, 2013; van Gog, Paas, & van Merriënboer, 2008). A meta-analysis found that instructional explanations slightly improved conceptual learning from worked examples but not procedural learning, and only in certain domains, suggesting that exposure to principle-based explanations can have a positive effect on learning even when they are not self-generated (Wittwer & Renkl, 2010). The same meta-analysis failed to find evidence that instructional explanations were more effective than self-explanation at supporting learning from worked examples, although designs that encourage some form of active processing and application of instructional explanations are expected to improve learning (Wittwer & Renkl, 2008). Consistent with prior work (Schworm & Renkl, 2006), instructional explanations were embedded in each worked example in the following experiments to serve as a comparison condition that suppresses spontaneous self-explanation and controls for the amount of information processed across conditions while manipulating the processes (i.e., reading explanations, generating self-explanations, generating comparisons).

1.5 METACOGNITION

Metacognition, or the ability to think about one's own cognitions, is a critical skill for self-regulated learning (Boekaerts & Corno, 2005; Efklides, 2011; Winne, 1995; Zimmerman, 2000, 2011). Several key components of metacognition include monitoring understanding, controlling one's strategies, and evaluating solutions. For an example of monitoring understanding, students listening to a teacher's instruction, reading a text, or solving practice problems benefit from being able to recognize *when* they don't understand something and *what* they don't understand. With that recognition, they can ask questions, re-read, or seek out other resources to improve their understanding. Otherwise, they might continue with an incomplete or inaccurate understanding, which would hurt their comprehension of later lessons building on those ideas as well as their performance on homework and exams. An awareness of one's learning processes and how to change them can help a student redirect her efforts if she does not understand a lesson or is not moving toward a problem solution; such control can reduce the likelihood of a student trying the same, ineffective strategy until eventually giving up. Finally, the ability to evaluate one's solutions helps a learner assess her comprehension at the end of a lesson or a portion of a lesson, as well as the accuracy of her answers on a problem-solving task or assignment. Students with poor evaluation abilities frequently stop thinking about a problem as soon as they arrive at *any* answer, which prevents them from recognizing flaws or inaccuracies in their answers. This leads to missing out on opportunities for error correction and can propagate misconceptions, if the student thinks she is getting correct answers or has a complete understanding when she does not.

Metacognition has been studied extensively in classroom and laboratory learning environments, and it has been associated with better learning and transfer and more positive motivational outcomes (Schraw, Dunkle, Bendixen, & Roedel, 1995; Veenman, Elshout, & Meijer, 1997; Veenman & Verheij, 2001; Wolters & Pintrich, 1998). Less work has tested interventions to improve learners' metacognitive skills, although some research has suggested promising effects of metacognitive training on use of metacognitive behaviors as well as motivation and near- and far-transfer learning outcomes (Brand, Reimer, & Opwis, 2003; Lin & Lehman, 1999; Palinscar & Brown, 1984; Zepeda, Richey, Ronevich, & Nokes-Malach, in press). As an alternative to direct instruction of metacognition, I examine whether prompting different instructional techniques can have a similar effect of increasing learners' use of metacognition, and whether metacognitive behaviors contribute to the effects of different instructional conditions on learning outcomes. Given the mechanisms of prior knowledge activation and mental model revision associated with self-explanation, it may be that an intervention modeling and prompting self-explanation leads students to increase their use of metacognitive processes. I investigate this question in Experiment 2 by assessing students' self-reports of metacognition in relation to their instructional condition and learning outcomes.

1.6 MOTIVATION

Motivation is associated with a variety of learner behaviors, including the type of study strategies learners employ, their learning and transfer of the content being studied, their persistence in the face of difficulty, their enjoyment and interest in a course, and their decision to pursue a topic either outside the classroom context or in future courses (Elliot, McGregor, & Gable, 1999; Elliot & McGregor, 2001; Hulleman, Godes, Hendricks, & Harackiewicz, 2010; Hulleman & Harackiewicz, 2009; Pintrich & de Groot, 1990; Pintrich, Marx, & Boyle, 1993; Pintrich, 2003). In addition to reducing learning and performance in a course, poor motivation can have negative repercussions on students' long-term career trajectories. Prior research has shown that students' motivation to do science and math declines during middle school (Collins & Osborne, 2001;

Haworth, Dale, & Plomin, 2008; Jarvis & Pell, 2001; Jenkins & Nelson, 2005; Lyons, 2006), potentially contributing to the decrease in students who report wanting to pursue science and math careers in high school or who actually do so in college and beyond compared to those who express interest in middle school. For these reasons, it is important to understand how different instructional techniques support changes in motivation.

One approach for assessing learner motivation is Elliot and colleagues' achievement goals framework (Elliot & McGregor, 2001; Elliot & Murayama, 2008). Achievement goals are classified along two dimensions. First, the definition of the goal is determined by whether the learner makes competency judgments based on intrapersonal or absolute criteria (mastery goal) or interpersonal criteria (performance goal). Second, the valence of the goal is determined by whether the learner strives to attain positive outcomes (approach goal) or avoid negative ones (avoidance goal). These two dimensions are crossed to produce four distinct, though not mutually exclusive, goals: mastery approach (e.g., "My aim is to completely master the material presented in this class"), mastery avoidance (e.g., "I am striving to avoid an incomplete understanding of the course material"), performance approach (e.g., "My goal is to perform better than other students"), and performance avoidance ("My aim is to avoid doing worse than other students").

Prior work has associated each of these distinct goals with different motivations, learning behaviors, and outcomes. In general, mastery-approach goals are associated with effective learning strategies and positive learning outcomes. Performance-approach goals are associated with a mixture of effective and ineffective learning strategies and generally positive learning outcomes, while mastery-avoidance goals are associated with a mixture of effective and ineffective strategies as well as positive and negative learning outcomes. Finally, performance-avoidance goals are associated with ineffective strategy use and negative learning outcomes (Ames & Archer, 1988; Elliot et al., 1999; Elliot & McGregor, 2001; Howell & Watson, 2007; Hulleman, Schrager, Bodmann, & Harackiewicz, 2010; Linnenbrink-Garcia, Tyson, & Patall, 2008; Middleton & Midgley, 1997; Roussel, Elliot, & Feltman, 2011). In the current work, I focus on achievement goals as a measure of motivation because their relation to multiple levels of learning and performance outcomes has been well established in classroom and laboratory research, and because they can be framed at both the domain and task level, allowing assessment of how instruction changes a learner's goals for a particular task. Different instructional techniques may promote the adoption of different achievement goals. For example, being repeatedly prompted to explain may lead students to more strongly endorse mastery-approach goals that are consistent with deep processing behaviors, while prompts to study completed examples might lead students to more strongly endorse performance-approach goals consistent with an emphasis on correct answers.

In both studies, I measured achievement goals to assess whether different instructional conditions supported students' adoption of different task-framed achievement goals. Examining this question could contribute to theoretical understanding of the nature and mechanisms of each instructional technique. It also has practical implications for instructional design. For example, it may suggest that instructional approaches are an avenue for shaping and redirecting students' motivations toward more fruitful achievement goals based on the types of learning behaviors and outcomes the teacher wishes to promote.

1.7 FRAMEWORK FOR COMPARING INSTRUCTIONAL TECHNIQUES

The knowledge representations created by a particular instructional technique have implications for understanding when that technique should be supported by the instructor (or selected by the student) based on the instructional goals, prior knowledge of the learner, and other affordances of the task and environment (Koedinger et al., 2013; Koedinger, Corbett, & Perfetti, 2012; Nokes et al., 2011). For example, Nokes et al. (2011) showed that the efficacy of self-explanation prompts depends on instructional fit, the idea that there must be alignment among task goals, instructional prompts, and the cognitive processes and knowledge representations they support. Along similar lines, the knowledge-learning-instruction framework (Koedinger et al., 2012) proposes a systematic approach to understanding how processes such as self-explanation or analogical comparison facilitate specific learning events, which then interact with and produce changes in knowledge. This approach of emphasizing alignment between instruction, learning, and outcomes advances the cognitive sciences by shifting focus away from pitting one instructional principle against another to find which is "better" and instead emphasizes understanding the particular learning events and knowledge features each instruction facilitates. However, while this framework outlines the general structure of interactions among instruction, learning, and knowledge, there is a great deal of work to done in better understanding the learning events and knowledge representations associated with specific instructional principles, particularly in comparison to others thought to support similar types of knowledge.

As discussed in the sections above, there is a great deal of overlap between mechanisms thought to support learning through self-explanation and analogical comparison, as well as the knowledge that both produce. For example, inference generation is a primary mechanism of learning from both analogical comparison and self-explanation. However, there are also critical differences in the hypothesized mechanisms underlying each technique that suggest they might support different knowledge representations. For example, analogical comparison seems especially well suited for prompting abstract knowledge that transfers, while self-explanation seems especially well suited for error correction and revising misconceptions. Analogical comparison typically focuses on cases or problems, making self-explanation potentially more flexible than analogical comparison but perhaps also less structured. In contrast, self-explanation typically focuses on one example at a time, so it may provide a greater opportunity for encoding concrete problem features compared to analogical comparison, which could result in better memory of specific problem-solving procedures.

1.7.1 Prior work comparing self-explanation and analogical comparison

A few past studies have compared directly the learning outcomes supported by analogical comparison and self-explanation. Gadgil et al. (2012) found that learners with misconceptions were more likely to undergo conceptual change when they compared their own flawed mental models to an expert model, as compared to when they were instructed to self-explain the expert model alone. While this work highlights the potential for analogical comparison to facilitate conceptual change, it may be that the conceptual change was driven by the use of a flawed mental model as one of the targets for comparison. In other words, it is unclear whether the learning was driven by the process of analogical comparison or by drawing the learners' attention to their own flawed models and misconceptions; in the case of the self-explanation condition, the learners' flawed mental models were not explicitly targeted for self-explanation. Nokes-Malach et al. (2013) compared self-explanation and analogical comparison of worked examples against worked examples paired with instructional explanations and found that analogical comparison led to less near transfer than self-explanation or instructional explanation study. Participants across conditions performed equally well on intermediate transfer measures, and the self-explanation and analogical comparison conditions demonstrated a learning advantage for far-transfer measures compared to the instructional explanations condition. This experiment was conducted in a classroom setting, however, and thus was not able to capture more fine-grained behavioral measures (e.g., solution times) that might provide insights into the types of knowledge representations constructed through each technique and the factors moderating those processes.

1.7.2 Current studies

For both techniques, the types of knowledge produced depend a great deal on the design of the materials, such as the amount of analogical comparison support provided (Gentner et al., 2003) or whether self-explanation prompts focus on filling gaps in knowledge or revising mental models (Nokes et al., 2011). Comparing the two techniques while controlling for factors such as the amount of scaffolding provided and the focus of the prompts could provide evidence for claims about the relative support each technique provides for different learning processes and outcomes.

In both of the following experiments, learners received instructional text and worked examples. Across both experiments, the prompts supporting analogical comparison or self-explanation targeted the worked examples. Prior work in both areas has prompted comparison and explanation of a variety of targets, including text and examples. However, much evidence suggests that the effectiveness of analogical comparison depends on the selection and design of the examples a learner is instructed to compare (Gentner, 1983). Therefore, simply instructing a learner to compare across blocks of text would likely prove ineffective unless those blocks of text had been carefully written and designed to highlight key structural relations. Creating content for comparison with well-aligned structural features is a simpler task with examples, as these features can be selected and manipulated more easily.

Although some work has found greater learning benefits from providing direct instruction (Schwartz & Bransford, 1998) or a statement of key principles *after* prompting learners to compare across principles (Alfieri et al., 2013), practical constraints stemming from the nature of this experiment made it better to provide direct instruction before each case comparison. Specifically, giving direct instruction on each topic before prompting participants to interact with the worked examples provided a source of prior knowledge that could be used for generating self-explanations or identifying and abstracting structural relations, depending on the condition. Thus, each condition had clear targets for the prompts with sufficient resources and scaffolding to engage in fruitful comparison or explanation of the examples. For these reasons, I determined that targeting worked examples with prompts and including instructional text that preceded the examples would be the least likely to introduce a design-based advantage for one instructional technique over the other. This design was similar to a number of past studies of analogical comparison and self-explanation, as well as other research investigating learning from worked examples.

Often learners engage in spontaneous self-explanation and analogical comparison, and prompts to engage in one may inadvertently encourage a learner to engage in the other process as well (Edwards, 2014). In both experiments I presented problems sequentially on separate pages to minimize spontaneous comparison except in the analogical comparison condition, in which examples were presented side-by-side. Instructional explanations were included along with worked examples to discourage spontaneous explanation in a third condition (Hausmann & VanLehn, 2010; Schworm & Renkl, 2006). Despite these design decisions, spontaneous comparison and explanation might occur to some degree in all conditions, just as the prompts were likely to produce great variability in the ways participants responded. Specifically, learners prompted to self-explain might also compare the worked example targeted by the prompt to their prior knowledge, the content of the instructional materials, and possibly the previous worked examples. Participants in the analogical comparison condition might also engage in some degree of explanation about why similarities or differences were meaningful. Finally, participants in the instructional explanation condition might engage in some degree of both spontaneous explanation and comparison.

However, through these experiments I sought to examine self-explanation and analogical comparison as they occur when prompted, and thus attempting to strip away all spontaneous behaviors that align with the non-prompted techniques would make the behaviors less authentic representations of what learners typically do and might remove critical elements necessary to each technique's success. Analyses of main effects target the differences between outcomes based on the learning processes supported by prompts, and not all the learning behaviors that might occur in each condition. I included self-reported measures of participants' learning processes to assess individual differences in spontaneous behaviors, and I coded the responses written by participants in the analogical comparison and self-explanation conditions to assess variability in response quality.

Although past work comparing the learning outcomes of the two techniques provides evidence of different mechanisms, this work needs to be replicated and extended with more fine-grained measures. There is strong evidence that different instructional techniques can lead to different types of processing and knowledge representations, even if those representations produce equivalent levels of accuracy (e.g., Nokes & Ohlsson, 2005; Nokes, 2009). Given that the proposed work aims to compare two instructional techniques that have both been shown to produce deep understanding (Koedinger et al., 2013; Richey & Nokes-Malach, 2015), it is particularly important to include measures capable of capturing differences in underlying processes and representations even if accuracy does not differ across conditions. For that reason, these experiments included a number of fine-grained measures of knowledge and learning processes, which I describe in greater detail in the methods sections below.

2.0 EXPERIMENT 1: SEQUENCE EXTRAPOLATION THROUGH ANALOGICAL COMPARISON, SELF-EXPLANATION, INSTRUCTIONAL EXPLANATION, AND PROBLEM-SOLVING PRACTICE

Experiment 1 compared fine-grained learning and performance measures from analogical comparison and self-explanation of sequence extrapolation problems. Sequence extrapolation problems (Figure 1) demonstrate a pattern that must first be identified and then extrapolated to continue the sequence, and they have been used in prior work to examine the relation between instruction and knowledge representations (Nokes & Ohlsson, 2005; Nokes, 2009).

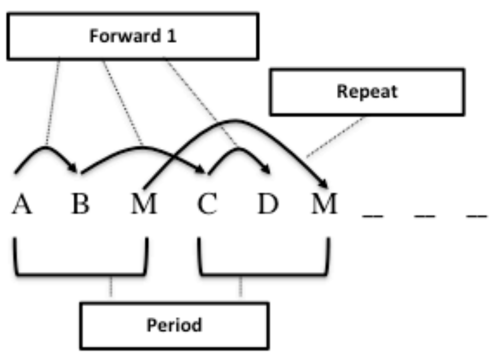


Figure 1. Sample sequence extrapolation pattern with rules labeled.

Sequence extrapolation problems are especially well suited to research aimed at differentiating the types of knowledge supported by analogical comparison and self-explanation because problems can be designed to target different types of knowledge representations. Relevant knowledge representations for completing this type of problem include declarative knowledge for describing the patterns (e.g., labels for component relations such as "forward 1"), declarative knowledge about strategies for completing the patterns that can range from concrete (e.g., “value two, B, increases one letter from value one, A”) to abstract (e.g., “identify the period of the sequence by looking for the smallest unit of the pattern”), and procedural knowledge about extrapolating patterns (e.g., knowing how to perform an extrapolation once the pattern has been identified; Simon, 1972).

Each of these knowledge representations can also be associated with different levels of transfer. I hypothesized that the type of instructional activity learners engaged in (self-explanation, analogical comparison, unprompted example study, or practice) would predict participants’ representations of knowledge, which in turn would predict the speed, accuracy, and completion patterns employed on subsequent problems. Detailed predictions are included in the methods.

3.0 METHOD

3.1 PARTICIPANTS

One hundred and eight students enrolled in an introductory psychology course were recruited to participate for course credit.

3.2 DESIGN

The experiment had a between-factors design with four conditions (self-explanation, analogical comparison, instructional explanation, and practice).

3.3 MATERIALS

3.3.1 Instructional text

All participants first received an instructional text adapted from Nokes and Ohlsson (2005) that introduced the concept of sequence extrapolation problems, discussed four types of sequence patterns, and provided strategies for identifying and extrapolating patterns. The text included diagrams of example sequence extrapolation problems along with a narrative text. At the end of each page of text, participants rated their understanding of the previous page on a 5-point Likert scale from 1 ("I don't understand at all") to 5 ("I understand completely").

3.3.2 Training and example problems

After completing the instructional text, participants received brief training in the instructional technique for their condition. Prior work has shown that self-explanation prompts are more effective when students receive explicit training in self-explanation (Bielaczyc et al., 1995), and instruction of analogical comparison principles could lead to similar effects. Training consisted of a short text that introduced the type of activity the learner would complete and emphasized that this type of activity had been shown to improve learning (Appendix A).

Following training, all participants received worked examples of sequence extrapolation problems. The problems were organized into two sets, with two worked examples in each set representing the same relational patterns but with different letters. In other words, the within-set problems differed at the intermediate level (Table 1), supporting more general abstraction across the examples. Materials were organized into two separate paper packets, with one set per packet. Each condition received different prompts to encourage the targeted instructional activity after each step of each worked example (Appendix A).

Table 1.

Levels of abstraction targeted by learning and test material problems with a simple demonstration example. For all examples, base for transfer is ABAB.

Transfer level	Characteristics	Example	Description of pattern

Near transfer	*Values*: different *Relations*: identical	RSRS	Same relations as ABAB: forward 1 and repeat
Intermediate transfer	*Values*: different *Relations*: quantitatively altered, relational structure identical	RTRT	Relations changed to forward 2 and repeat
Far transfer	*Values*: different *Relations*: quantitatively altered, different kind of relation, different relational structure	RRPP	Relations changed to backward 2 and repeat; different values related
	Values: different *Relations*: Pattern includes novel type of component relation	QRRQ	Novel mirror-flip relation

Although the examples were identical across conditions, the format and prompting differed. Participants in the analogical comparison, self-explanation, and instructional explanation conditions receive four step-by-step worked examples of sequence extrapolation problems, with each step of the example shown on a different page. Items were organized into sets of problems, with two examples and two test items associated with each set. Set membership was governed by having the same structural relations across examples, although individual examples varied in terms of the surface details and/or quantitative relations. Two additional test items were not directly associated with either set. The practice condition received the same initial and final problem states (problem and solution) on separate pages, but they did not receive the step-by-step solutions. To control for time, instructional explanation and practice conditions received two additional examples for each set. Details of each condition are described below. All example and test items are presented in Table 2. Odd-numbered examples were seen by participants in all four conditions, while even-numbered examples were given only to the instructional explanation and practice conditions to control for time. Items were adapted from prior research examining levels of transfer using sequence extrapolation problems (Nokes & Ohlsson, 2005; Nokes, 2009).

Table 2.

Examples and test items used in Experiment 1.

Item	Provided sequence	Extrapolation	Abstraction level	Transfer
Set 1				
Ex. 1	I J H K G S J K I L H T	K L J M I U L M	Initial example	N/A
Ex. 2	M N L O K W N O M P L X	O P N Q M Y P Q	All relations identical to E1	Near
Ex. 3	R T Q V P B S U R W Q C	T V S X R D U W	Some relations identical to E1, some quantitative relations changed	Intermediate
Ex. 4	N P M R L E O Q N S M F	P R O T N G Q S	Some relations identical to E1, some quantitative relations changed (all relations identical to E3)	Near (to E3)
Test 1	E F D G C O F G E H D P	G H F I E Q H I	All relations identical to E1	Near
Test 2	E H D K C O F I E L D P	G J F M E Q H K	Some relations identical to E1 or E3, some quantitative relations changed	Intermediate (quantitative)
Set 2				
Ex. 5	L M Z M L Y M N X N M W	N O V O N U O P	Initial example	N/A
Ex. 6	I J W J I V J K U K J T	K L S L K R L M	All relations identical to E5	Near
Ex. 7	E G S G E R F H Q H F P	G I O I G N H J	Some relations identical to E5, some quantitative relations changed	Intermediate
Ex. 8	M O G O M F N P E P N D	O Q C Q O B P R	Some relations identical to E1, some quantitative relations changed (all relations identical to E7)	Near (to E7)
Test 3	R S F S R E S T D T S C	T U B U T A U V	All relations identical to E5	Near
Test 4	D A Q A D P E B O B E N	G I M I G L I K	Some relations identical to E5 or E7, some quantitative relations and directions changed	Intermediate (quantitative and values)
Test 5	B C P X Y O C D N	1) Y Z M D E K Z A 2) W X M D E L V W	1) Surface similarity: new combination of familiar quantitative relations 2) Deep analogy: creation of new rule (mirror-flip alphabet) based on relations in E5-8	Far
Test 6	B A C B E D H G	L K Q P W V D C	Novel rule (forward n+1)	Far

3.3.2.1 Self-explanation

Participants in the self-explanation condition viewed the problems sequentially with one step shown at a time, accompanied by a prompt to write an explanation of the step (Figure 2). Materials primarily employed the strategy of eliciting self-explanation through direct prompting, and the language in the prompts was adapted from prior work that successfully prompted self-explanation (Hausmann & VanLehn, 2010; Wong et al., 2002). Each page provided space for participants to write their responses to the questions below the example.

Figure 2. Worked example steps formatted for the self-explanation condition.

The self-explanation prompts were expected to focus learners on the surface features and relations in each example. The degree to which learners studied surface features should be associated with time on near-transfer test problems, which have the same concrete relations shown in the examples. Self-explanation prompts also should encourage generalization about types of relations and procedures to the extent that learners used their prior knowledge to make sense of the examples. This more general knowledge should increase accuracy and reduce time on intermediate-transfer problems, though it may not help learners generate the novel relations required for the far-transfer problems. It should also support more variety in the completion patterns employed. Learners may use the examples to fill in gaps in their prior knowledge or correct errors in their understanding of general strategies, which will support accuracy across all problems, including the novel-relation problem.

3.3.2.2 Analogical comparison

Participants in the analogical comparison condition saw the parallel steps of both examples in a set at once, with instructions to identify the similarities and differences between each pair of steps (Figure 3). The examples in Figure 3 highlight the same rule, *backward* 1, across both examples (same quantitative relation), though the rule is instantiated with different letters (different surface features). Other elements of the comparison, such as the letters LM and EG at the beginning of the sequences, illustrated the same structural relation (the second letter of the pattern is *forward n* from the first letter) but instantiated it with different letters (different surface features) and different quantitative relations (*forward 1* and *forward 2*).

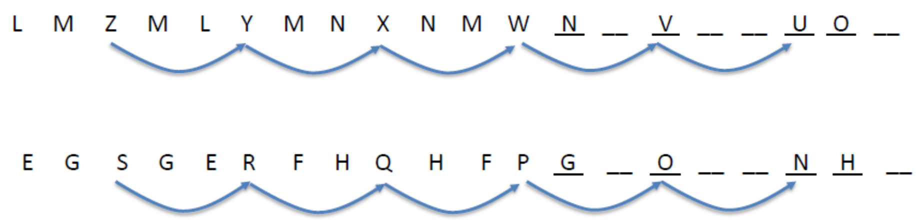


Figure 3. Worked examples steps formatted for the analogical comparison condition. Other conditions received the same steps separately, one example sequence at a time.

Participants were instructed to write their responses to the prompts in the space provided on each page. The instructional materials incorporated several strategies for encouraging structural comparison. In addition to the explicit instructions to compare, the materials used perceptual cues to emphasize alignment by presenting each pair of examples next to each other, with aligned steps directly above or below each other. Labels identifying structurally aligned steps also built on prior work investigating ways to promote comparison. The analogical comparison prompts were designed to focus learners on the abstract structure of component relations shared across the two worked examples. In other words, rather than encoding the quantitative relations (e.g., that the second value is “Forward 1” from the first value), learners were expected to encode the structural relation (e.g., that the second value is Forward X from the first value). This abstract, structural knowledge could be more easily applied to intermediate- and far-transfer problems, potentially leading to less time required on those problems and greater variety in completion patterns. The prompts also focused learners on the general strategies for identifying patterns common across the two examples (e.g., first identify the period of the pattern), which should support accuracy across all problems, including the far-transfer problems.

3.3.2.3 Instructional explanation

Participants in the instructional explanation condition viewed the problems sequentially with one step shown at a time, but instead of prompts they received instructional explanations of each step (Figure 4). Participants saw the same four problems included in the analogical comparison and self-explanation materials, but to control for time on the task, they saw an additional four problems. Each of the additional problems was isomorphic to one of the four examples all participants received (i.e., identical quantitative relations to examples included for other conditions, but different letters). Similar to the self-explanation prompts, studying the worked examples and instructional explanations was hypothesized to encourage encoding of surface values and relations, which would facilitate faster times on near-transfer problems.

Participants were expected to encode the specific pattern detection and extrapolation procedures illustrated in the worked examples. Applying identical problem-solving procedures to different problems should lead to slower and less accurate solutions compared to learners who have knowledge of more general strategies and procedures, resulting in slower times on intermediate- and far-transfer problems. Similarly, knowledge of problem-solving procedures should be easily applied only to near-transfer problems, resulting in faster times on the near-transfer problems and having no effect on solution times for of other patterns.

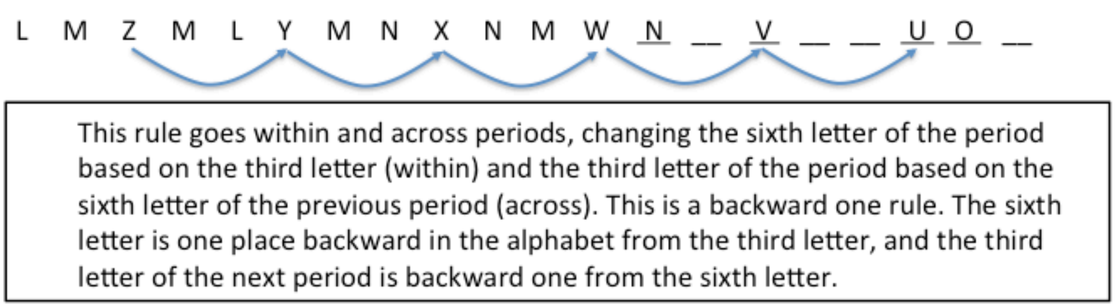


Figure 4. Sample of instructional explanation paired with worked example in the instructional explanation condition.

3.3.2.4 Practice

Participants in the practice condition received each example problem without solution steps, with the solution to each problem provided on the page immediately after the problem. Participants were instructed to solve each sequence extrapolation problem to the best of their ability, and then turn to the next page to study the solution. As with the instructional explanation condition, participants in the practice condition received an additional four problems, each of which was isomorphic to one of the examples seen by all conditions.

Participants were expected to encode the procedural features of the examples and the connections between the steps. This knowledge will result in high accuracy and speed on near-transfer problems. Practicing sequence extrapolation and studying solutions should lead to less knowledge about structural features and relations compared to conditions that saw, compared, or explained worked example steps. Consequently, practice should not perform as well on intermediate- and far-transfer items. Practice should also have little impact on the revision of errors, resulting in less error-correcting behavior.

3.3.3 Test materials

The test consisted of sequence extrapolation problems requiring varying levels of transfer including near transfer, intermediate transfer, and far transfer (Table 2). Each set of examples was associated with two test problems: one near-transfer problem with quantitative relations identical to the initial problem examples but different letter values, and one intermediate-level transfer problem with quantitative relations that were altered from the examples (e.g., a "forward 1" relation was changed into a "forward 2" relation) but the same underlying structure (e.g., values in positions 1, 3, and 5 have a forward or backward quantitative relation). Two far transfer problems relied on a new combination of familiar quantitative relations (test item 5) and a new rule that participants had not practiced (test item 6). Test item 5 had multiple possible solutions, including one that applied familiar relations illustrated in many of the examples and one that abstracted a relation illustrated in set 2 to generate a new variation on the relation (i.e., mirror flip to mirror-flip alphabet).

The two near-transfer items differed from the example problems in the same ways (identical relations, new surface features) and were analyzed together as a single measure of near transfer. The two intermediate-transfer items were analyzed separately because they differed from the example problems in different ways (item 2 had different quantitative relations but in the same direction, i.e., “forward 1” vs. “forward 3,” while item 4 had different quantitative and directional relations, i.e., “forward 1” vs. “backward 3”). The two far transfer items were also analyzed separately because they differed from example items in different ways (item 5 required a novel combination of familiar quantitative relations, while item 6 required the discovery of a novel rule).

Test materials were coded and analyzed based on accuracy (i.e., how many of the eight correctly extrapolated letters participants produced for each test item), solution time (i.e., how long participants spent from when they advanced to a new problem to when they submitted their response), and completion patterns (i.e., the position order in which they extrapolated the sequence). The most basic completion pattern is 12345678, in which the learner extrapolates each value in order from left to right. However, more effective patterns often involve extrapolating one rule at a time (see Figure 3; this example has been extrapolated in the order of positions 1736). In other words, if learners are extrapolating patterns based on one rule at a time (e.g., follow “Forward 1” to fill in every three letters, and then follow “backward 1” to fill in every five letters), they will extrapolate letters out of order. If they apply the surface or structural relations found in the prior examples (near or intermediate transfer), they may use more common extrapolation patterns by filling in one practiced rule at a time. If they are reasoning about relations by applying conceptual knowledge to construct new rules, they may use more unique extrapolation patterns. Finally, participants' metacognitive evaluation behaviors were analyzed based on the amount of time they spent between entering the final letter of each solution and submitting their responses. Their error-correcting behaviors will be assessed based on the number of times they change any letters in each response. Table 3 presents hypotheses regarding the different measures.

Table 3.

Targeted knowledge and predictions for each measure in Experiment 1.

Measures	Predictions	Knowledge targeted
Near transfer		**Surface-level abstraction**
Accuracy	AC=SE=WE=P	Declarative knowledge of example pattern relations; procedural knowledge for detecting and extrapolating patterns
Solution time	P<SE=WE<AC	
Intermediate transfer		**Intermediate-level abstraction**
Accuracy	AC=SE>WE>P	Declarative knowledge of example structures; procedural knowledge for detecting and extrapolating patterns
Solution time	AC=SE<WE=P	
Far transfer		**Structural-level abstraction**
Accuracy	AC=SE>WE>P	Declarative knowledge of abstract relations and strategies
Solution time	AC<SE<WE<P	
Use of Completion pattern variety	AC=SE>WE>P	
Error correction		**Metacognitive monitoring**
Changes to solution	SE>AC=WE>P	Declarative knowledge of monitoring/error-correction strategies
Checking time	SE>AC=WE>P	

3.3.4 Questionnaires

Participants responded to a series of questionnaires targeting their motivation and learning processes throughout the experiment. Each instructional condition was expected to influence the types of goals participants experienced during the learning phase, such that their goals would align with the types of knowledge emphasized by each technique. Conditions were also expected to influence the learning processes participants reported using, with participants' self-reports of learning processes aligning with the mechanisms underlying the technique they were prompted to use. Predictions are discussed in greater detail below.

3.3.4.1 Task-framed Achievement Goals Questionnaire

Elliot and colleagues' Achievement Goals Questionnaire-Revised (AGQ-R; Elliot & Murayama, 2008) was modified slightly to assess participants' self-reported goals at the task level. Specifically, the instructions asked participants to consider their feelings about the activities they completed during the session. The questionnaire assessed participants' self-reported endorsement of task-level mastery-approach (.75), mastery-avoidance (.76), performance-approach (.91), and performance-avoidance goals (alpha). Responses were recorded on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). Given the emphasis that explanation places on understanding, I expected participants in the self-explanation condition to report greater levels of mastery goals. Since practice emphasized arriving at correct solutions, I expected participants in the practice condition to report greater levels of performance goals.

3.3.4.2 Learning processes questionnaire

Participants responded to a 10-item learning processes questionnaire about the different processes they reported using during the learning and test phases (Table 4). Participants responded on a 7-point Likert scale ranging from 1 (“Unsure”) to 7 (“Strongly agree”). Items framed around the learning phase were designed to capture the range of behaviors directly encouraged by the prompts (e.g., explanation, comparison) as well as mechanisms thought to result from such behaviors (e.g., inference generation, error correction). Items framed around the test phase were designed to assess whether learners reported continuing to engage in the behaviors encouraged during the learning phase once the prompts were removed.

Table 4.

Learning processes questionnaire items and the level at which they were framed.

Framing	Prompt	Prediction
Learning phase	I compared what I saw in the problem packets to what I already knew from the tutorial.	SE > AC = WE = P
	I focused on the procedures I was learning.	P = WE > SE > AC
	I explained the material to myself.	SE > AC = WE = P
	I compared ideas or examples to one another.	AC > SE = WE = P
	I generated inferences about the material.	SE = AC > WE = P
	I corrected errors in my thinking.	SE > WE > AC = P
	I tried to identify the critical features of the problems.	SE = AC > WE > P
Test phase	I tried to recognize patterns I had already seen.	WE = P > SE > AC
	I compared the test problems to what I had seen in the learning packets.	SE = AC > WE = P
	I explained the problems to myself.	SE > AC = WE = P

3.4 PROCEDURE

Participants completed the experiment individually in sessions of three to five students at a time. All participants received the instructional text and were given six minutes to complete it. Participants then received the learning materials for their conditions, which were distributed in two packets with each set of examples in a separate packet. Participants were given 11 minutes to complete each packet and were encouraged to use the entire time to study and complete the materials. To decrease the likelihood of spontaneous comparison, participants were not permitted to flip back once they had completed a page in the learning packet. The test phase began after the completion of both learning packets and was presented on computers using PsyScope software (J. D. Cohen, MacWhinney, Flatt, & Provost, 1993). Participants first completed a basic practice test problem to familiarize themselves with the computer interface. They were given four minutes to complete each test problem.

4.0 RESULTS

Participants' accuracy, time, and completion patterns for each test problem were assessed. Performance was assessed using averages for near-transfer and intermediate-transfer problems; the two far-transfer problems were analyzed separately, as they involved different levels of abstraction and novel-rule generation. Error-correcting behaviors were analyzed based on the average number of times participants changed a response once they had entered a value and the average amount of time recorded after participants completed each pattern but before submitting it. I set the alpha level at .05 for main effects, interactions and planned comparisons, used Tukey's HSD test for post hoc comparisons, and report marginal effects for *p* values between .05 and .10 (Keppel & Wickens, 2004). I report effect sizes (Cohen's *d* or partial eta squared, η_p^2) for all significant main effects, interactions, and planned comparisons and interpret effects as small when $\eta_p^2 < .06$ or $d < .2$, medium when $.06 < \eta_p^2 < .14$ or $.3 < d < .8$, and large when $\eta_p^2 > .14$ or $d > .8$ (see Cohen, 1988; Olejnik & Algina, 2000).

4.1.1 Learning materials completion and accuracy

I examined participants' completion of learning materials across several measures to assess whether they felt they understood the materials, had sufficient time to complete them, and demonstrated understanding. First, I assessed participants' self-reports of understanding the instructional text, which were collected on a 5-point Likert scale at the end of each page of the text for a total of nine reports. Participants' average reports of understanding were calculated across individual prompts. Participants reported experiencing a high level of understanding (M = 4.71, SD = .44), with mean understanding falling just below a rating of "I understand completely." There were no differences in self-reported understanding across conditions, $F(3, 105) = 1.08$, $p = .36$, $\eta^2_p = .030$.

Participants were prompted to interact with the examples in different ways depending on condition (i.e., practice condition solved problems, instructional explanation condition studied examples and explanations with no prompts to write anything, and self-explanation and analogical comparison were prompted to write responses), so no between-condition comparisons of example material activities were possible. Instead, I examined descriptives of behaviors within each condition.

For the self-explanation and analogical comparison conditions, in which participants received prompts to respond to each step of each example, I assessed the average proportion of prompt responses for each condition. Participants in the self-explanation condition responded to the majority of the prompts in the example packets (M = .93, SD = .12). Of the 27 participants in the self-explanation condition, 10 left the final page of the first packet blank and three left the final page of the second packet blank, suggesting that most participants in the self-explanation condition were able to complete all materials within the allotted time. Participants in the analogical comparison condition responded to the majority of the prompts in the example packets (M = .94, SD = .088). Of the 28 participants in the analogical comparison condition, seven left the final page of the first packet blank and three left the final page of the second packet blank, suggesting that most participants in the analogical comparison condition were able to complete all materials within the allotted time.

Participants in the instructional explanation condition were not prompted to respond to any of the example content, and on average they made marks or notes on a very small portion of the example pages (M = .06, SD = .17). Eight out of 27 participants made at least one mark or note in the example packet, meaning the majority did not mark up the materials at all. Finally, participants in the practice condition were instructed to solve each example to the best of their ability before turning to the next page and viewing the solution. Extrapolations for each problem were coded based on the number of correct extrapolations made out of the eight letters they were asked to extrapolate for each problem. On average, participants correctly extrapolated about half of the letters across problems (M = 4.41, SD = 2.53).

4.1.2 Test accuracy

Participants received accuracy scores based on the number of correct extrapolations made out of eight possible extrapolations. A one-way ANOVA was conducted examining the effect of condition on accuracy across both near-transfer problems (Figure 5). There was a medium effect of condition on near-transfer problem accuracy, $F(3, 105) = 2.77$, $p = .045$, $\eta^2_p = .073$; a Tukey post hoc analysis revealed a significant difference only between the instructional explanation (M = .87, SD = .24) and practice conditions (M = .65, SD = .35; p = .042). There were no differences between instructional explanation and self-explanation conditions (M = .77, SD = *.27;* p = .64), instructional explanation and analogical comparison conditions (M = .69*,* SD = .33; p = .14), self-explanation and analogical comparison conditions (p = .77), self-explanation and practice conditions (p = .45), or analogical comparison and practice conditions (p = .95).

A one-way ANOVA assessing the effect of condition on accuracy on the intermediate-transfer problem with quantitative relation changes revealed no effect, $F(3, 105) = 1.82$, $p = .15$, $\eta^2_p = .049$. There was a marginal effect of condition on accuracy on the intermediate-transfer problem with quantitative and directional relation changes, $F(3, 105) = 2.15$, $p = .098$, $\eta^2_p = .058$; a Tukey post hoc analysis no differences between any of the conditions (ps > .12).

The far-transfer problem that required a new combination of familiar structural relations had multiple solutions, including one that involved deep analogy to a prior example (transformation of the mirror-flip rule into a mirror-flip cross-alphabet rule) and one that involved closer surface similarity to example problems (application of the "forward X" and "backward X" rules in a new pattern). There were no effects of condition on accuracy based on the surface-similarity solution, $F(3, 105) = .40$, $p = .75$, $\eta^2_p = .011$. Only four participants attempted the solution based on a deep analogy to prior examples (two in the instructional explanation condition, one in practice, and one in analogical comparison), so no further analyses were conducted using the deep analogy solution as the criterion for accuracy. There was no effect of condition on accuracy on far-transfer problem that required generation of novel relations (the "forward n + 1" rule), $F(3, 105) = 1.03$, $p = .38$, $\eta^2_p = .029$.

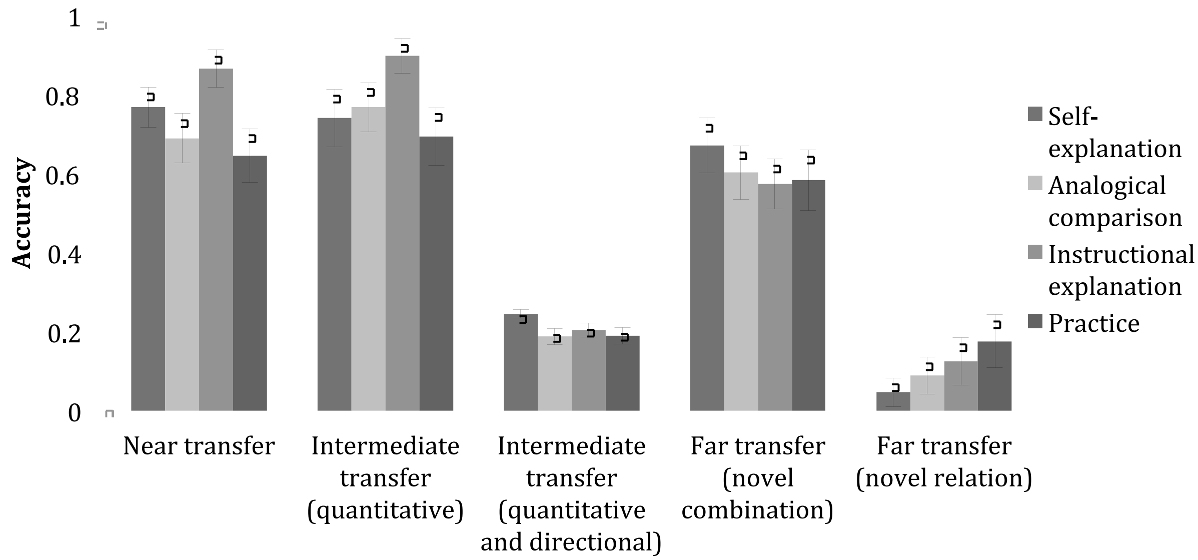


Figure 5. Accuracy by problem type across conditions.

4.1.3 Solution times

A series of one-way ANOVAs was conducted to examine the effect of condition on the time in seconds that participants spent on each problem (Figure 6). There were no differences among conditions in time on near-transfer problems, $F(3, 103) = .88$, $p = .46$, $\eta^2_p = .025$, the intermediate-transfer problem with quantitative relation changes, $F(3, 104) = .58$, $p = .63$, $\eta^2_p = .016$, or the intermediate-transfer problem with quantitative and directional relation changes, $F(3, 103) = 1.44$, $p = .24$, $\eta^2_p = .040$. There was a medium effect of condition on time on the far-transfer problem that required a new combination of familiar structural relations, $F(3, 104) = 3.81$, $p = .012$, $\eta^2_p = .099$; a Tukey post hoc analysis revealed a significant difference only between the instructional explanation ($M = 180$, $SD = 38.8$) and practice conditions ($M = 128$, $SD = 45.4$; $p = .006$). There were no differences between instructional explanation and self-explanation conditions ($M = 151$, $SD = 50.5$; $p = .26$), instructional explanation and analogical comparison conditions (M = 158, SD = 82.2; p = .48), self-explanation and analogical comparison conditions (p = .97), self-explanation and practice conditions (p = .43), or analogical comparison and practice conditions (p = .21).

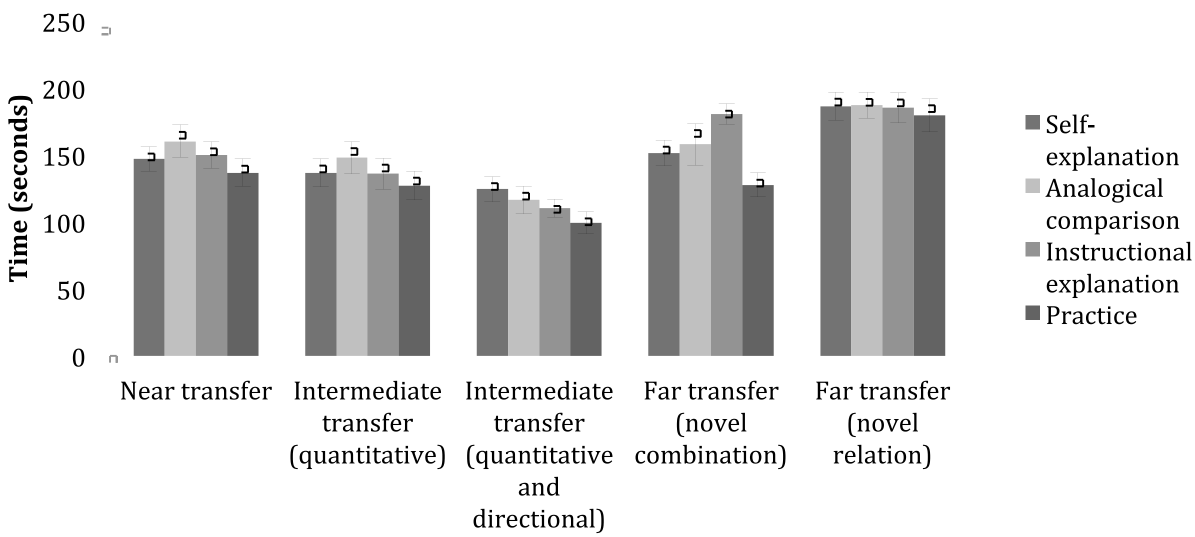


Figure 6. Solution times in seconds by problem type across conditions.

4.1.4 Detection times

A series of one-way ANOVAs were conducted to examine the effect of condition on the time in seconds that participants spent detecting the patterns for each problem (Figure 7). There was a marginal difference among conditions in detection time on near-transfer problems, $F(3, 105) = 2.48$, $p = .065$, $\eta^2_p = .066$; a Tukey post hoc analysis revealed a marginal difference only between the instructional explanation (M = 33.1, SD = 24.8) and analogical comparison conditions (M = 50.5, SD = 30.5; p = .059). There was a marginal difference among conditions in detection time on the intermediate-transfer problem with quantitative relation changes, $F(3, 105) = 2.33$, p =

.078, η^2_p = .062; a Tukey post hoc analysis revealed a marginal difference only between the instructional explanation (*M* = 22.1, *SD* = 20.6) and practice conditions (*M* = 44.1, *SD* = 42.3; *p* = .059). There was a marginal difference among conditions in detection time for the intermediate-transfer problem with quantitative and directional relation changes, *F*(3, 105) = 2.43, *p* = .069, η^2_p = .065; a Tukey post hoc analysis revealed a marginal difference only between the instructional explanation (*M* = 18.9, *SD* = 14.1) and analogical comparison conditions (*M* = 30.9, *SD* = 15.8; *p* = .050). There was a large effect of condition on time on the far-transfer problem that required a new combination of familiar structural relations, *F*(3, 105) = 5.86, *p* = .001, η^2_p = .14; a Tukey post hoc analysis revealed a significant difference between the instructional explanation condition (*M* = 65.4, *SD* = 38.1) and all other conditions (analogical comparison, *M* = 39.5, *SD* = 32.6, *p* = .008; self-explanation, *M* = 38.4, *SD* = 24.5, *p* = .006; practice, *M* = 36.2, *SD* = 19.6, *p* = .002). There was no effect of condition on detection time for far transfer problem that required generation of a novel rule, *F*(3, 105) = 1.16, *p* = .33, η^2_p = .032.

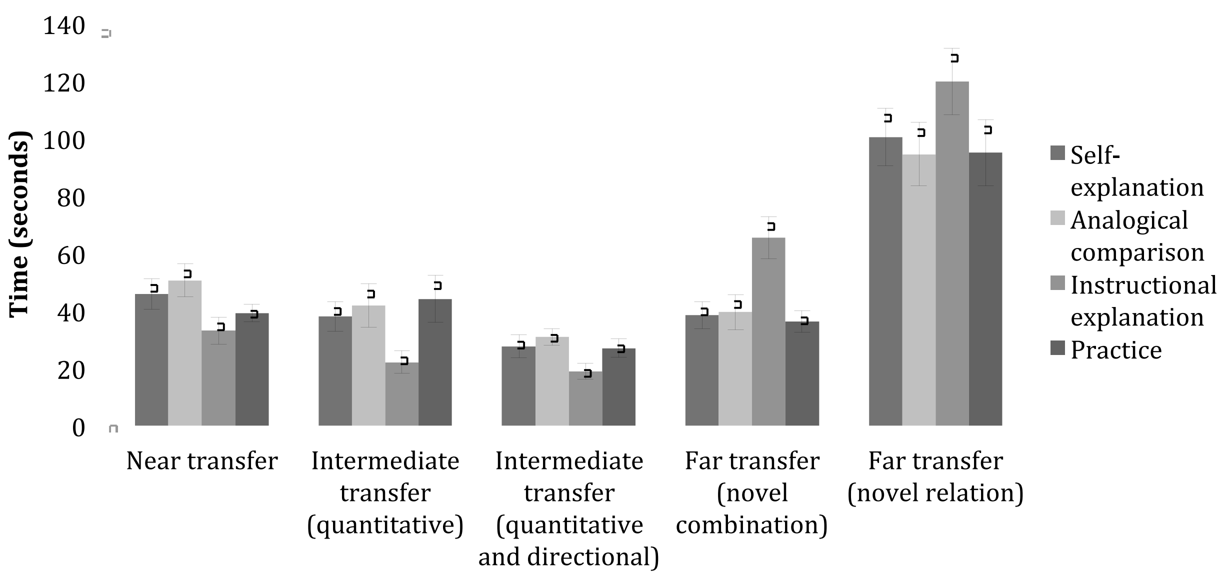


Figure 7. Detection times in seconds by problem type across conditions.

4.1.5 Extrapolation times

A series of one-way ANOVAs was conducted to examine the effect of condition on the time in seconds that participants spent extrapolating the patterns for each problem (Figure 8). There was a medium effect of condition on extrapolation time on near-transfer problems, $F(3, 105) = 3.22$, $p = .026$, $\eta^2_p = .084$; a Tukey post hoc analysis revealed differences between the instructional explanation condition ($M = 118.4$, $SD = 41.8$) and both the self-explanation condition ($M = 86.7$, $SD = 49.1$; $p = .050$) and the practice condition ($M = 85.5$, $SD = 41.1$; $p = .039$). There were no effects of condition on extrapolation time on the intermediate-transfer problem with quantitative relation changes, $F(3, 105) = 1.71$, $p = .17$, $\eta^2_p = .042$, the intermediate-transfer problem with quantitative and directional relation changes, $F(3, 105) = 1.24$, $p = .30$, $\eta^2_p = .034$, the far-transfer problem that required a new combination of familiar structural relations, $F(3, 105) = 1.06$, $p = .37$, $\eta^2_p = .029$, or the far transfer problem that required generation of a novel rule, $F(3, 105) = 1.07$, $p = .36$, $\eta^2_p = .030$.

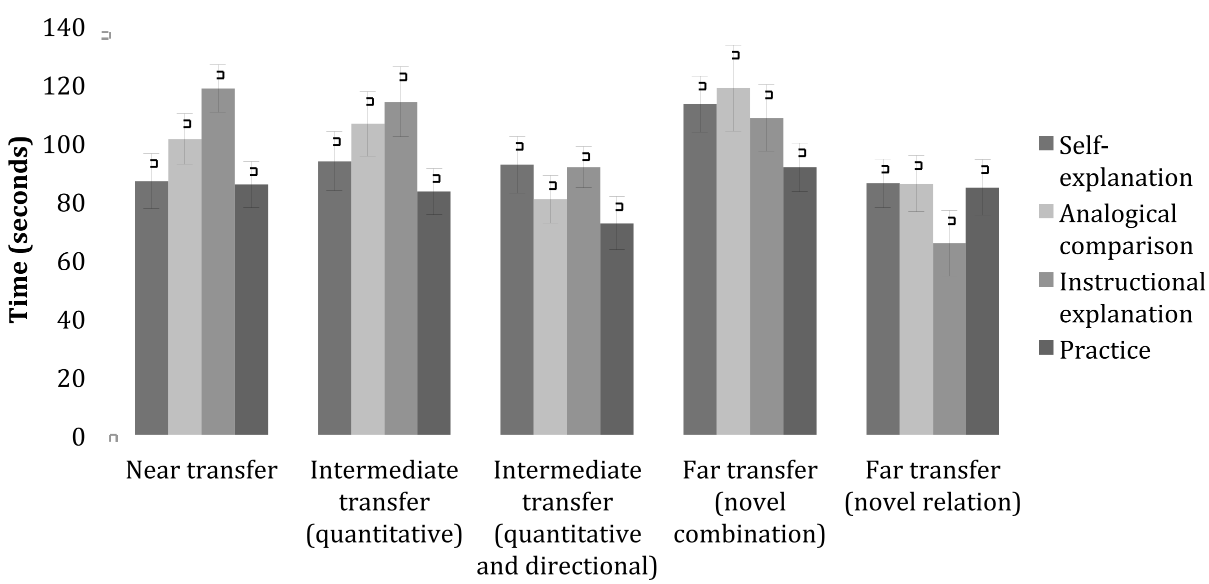


Figure 8. Detection times in seconds by problem type across conditions.

4.1.6 Completion patterns

Patterns were classified either as basic (completing the pattern from left to right, i.e., 12345678), common (someone else in the same condition used the same completion pattern), or unique (no one else in the same condition used the same pattern; Nokes & Ohlsson, 2005). I examined the completion patterns participants employed when solving each problem as a measure of whether they were applying knowledge of surface or structural relations (common pattern, near or intermediate transfer) or applying knowledge of principles to generate new knowledge of relations (unique pattern, far transfer). Completion patterns were coded only if the participant entered values for all eight extrapolations on a given problem. I conducted a series of chi-squared tests to assess whether participants' use of the three types of completion strategies differed for each problem by condition. In the case of any significant tests, I examined adjusted standardized residuals in accordance with MacDonald and Gardner (2000) and identified any cells with residuals greater than + / - 3 as having a greater or smaller number of participants using that pattern than would be expected by chance. There was a marginal difference in strategies across conditions for the first near-transfer problem, $X^2(6, N = 102) = 11.27$, $p = .080$; however, no cells had an adjusted standardized residual greater than + / - 3. There were no differences across conditions for the second near-transfer problem, $X^2(6, N = 106) = 5.00$, $p = .54$, the first intermediate-transfer problem, $X^2(6, N = 101) = 3.77$, $p = .71$, the second intermediate-transfer problem, $X^2(6, N = 107) = 2.80$, $p = .83$, or the first far-transfer problem, $X^2(6, N = 101) = 8.87$, $p = .18$. There was a significant difference across conditions for the second far transfer, $X^2(6, N$

= 92) = 27.99, $p < .01$, which required generation of a novel relation. The cell for participants in the practice condition who used a common pattern had an adjusted standardized residual of 3.9, indicating that more participants than would be expected by chance used a common pattern in the practice condition. The cell for participants in the practice condition who used a unique pattern had an adjusted standardized residual of -3.7, indicating that fewer participants than would be expected by chance used a unique pattern in the practice condition.

4.1.7 Error correction

To test whether self-explanation led to greater emphasis on checking and correcting errors, I examined the number of times each participant changed a response during the test phase (response change; Figure 9) and the amount of time spent after entering the final letter of the sequence before submitting the response (check time). A series of one-way ANOVAs revealed no differences between conditions in the number of response changes for near-transfer problems, $F(3, 105) = .56$, $p = .64$, $\eta^2_p = .016$, the first intermediate-transfer problem, $F(3, 105) = .58$, $p = .63$, $\eta^2_p = .016$, the second intermediate-transfer problem, $F(3, 105) = 1.30$, $p = .28$, $\eta^2_p = .036$, the first far-transfer problem, $F(3, 105) = 1.27$, $p = .29$, $\eta^2_p = .035$, or the second far-transfer problem, $F(3, 105) = 1.19$, $p = .32$, $\eta^2_p = .033$.

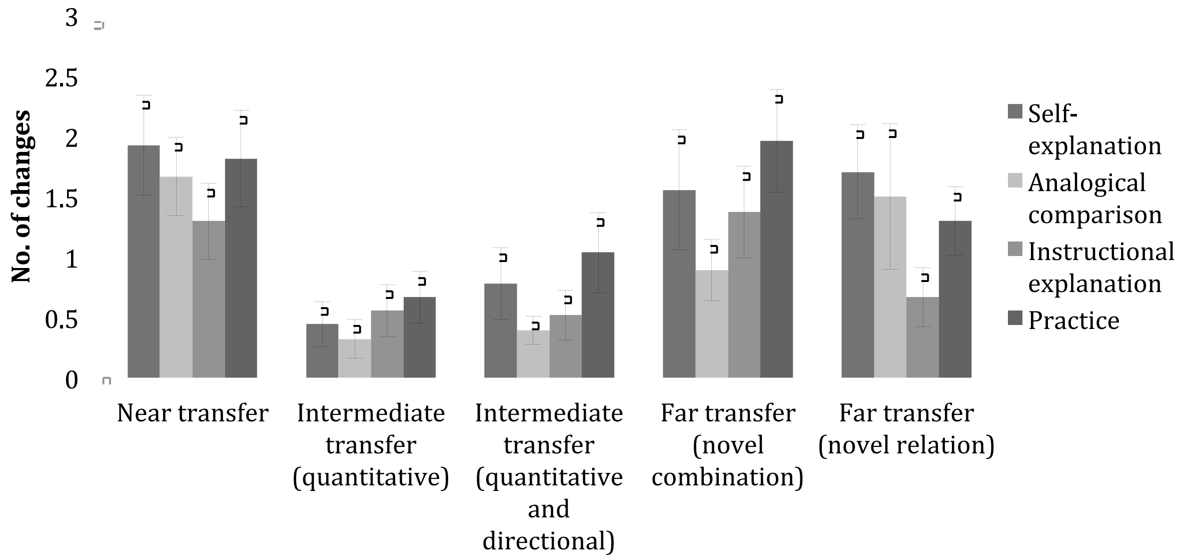


Figure 9. Mean number of changes to final extrapolation pattern by condition

The check time analyses examined check times only using items for which a participant submitted a complete response. On items where participants did not submit a complete response, they likely ran out of time and therefore did not have an opportunity to check their answers. An average check time per problem was calculated for each participant, using only check times for completed patterns. A one-way ANOVA revealed no effects of condition on average check time across problems, $F(3, 105) = .30$, $p = .82$, $\eta^2_p = .009$.

4.1.8 Achievement goals

To assess whether different learning conditions prompted participants to endorse different task-framed achievement goals, I conducted a series of one-way ANOVAs (Figure 10). Analyses revealed a medium effect of learning condition on participants' endorsement of mastery-avoidance goals, $F(3, 105) = 3.41$, $p = .020$, $\eta^2_p = .089$; a Tukey post hoc analysis revealed a significant difference only between the practice ($M = 4.17$, $SD = 1.37$) and analogical comparison conditions, (M = 5.13, SD = 1.08; p = .040). There were no differences between practice and self-explanation conditions (M = 4.93, SD = 1.16, p = .16), practice and instructional explanation conditions (M = 4.31, SD = 1.61, p = .98), analogical comparison and self-explanation conditions (p = .94), analogical comparison and instructional explanation conditions (p = .10), or self-explanation and instructional explanation conditions (p = .32).

There was also a medium effect of condition on endorsement of performance-approach goals, $F(3, 105) = 3.00$, $p = .034$, $\eta^2_p = .079$; a Tukey post hoc analysis revealed a significant difference only between the practice (M = 5.17, SD = 1.40) and instructional explanation conditions, (M = 3.88, SD = 1.74, p = .020). There were no differences between practice and self-explanation conditions (M = 4.33, SD = 1.84, p = .23), practice and analogical comparison conditions (M = 4.40, SD = 1.43, p = .30), analogical comparison and self-explanation conditions (p > .99), analogical comparison and instructional explanation conditions (p = .62), or self-explanation and instructional explanation conditions (p = .73).

There were no effects of condition on mastery-approach goal endorsement, $F(3, 105) = .31$, $p = .82$, $\eta^2_p = .009$, or on performance-avoidance goal endorsement, $F(3, 105) = 2.05$, $p = .11$, $\eta^2_p = .055$.

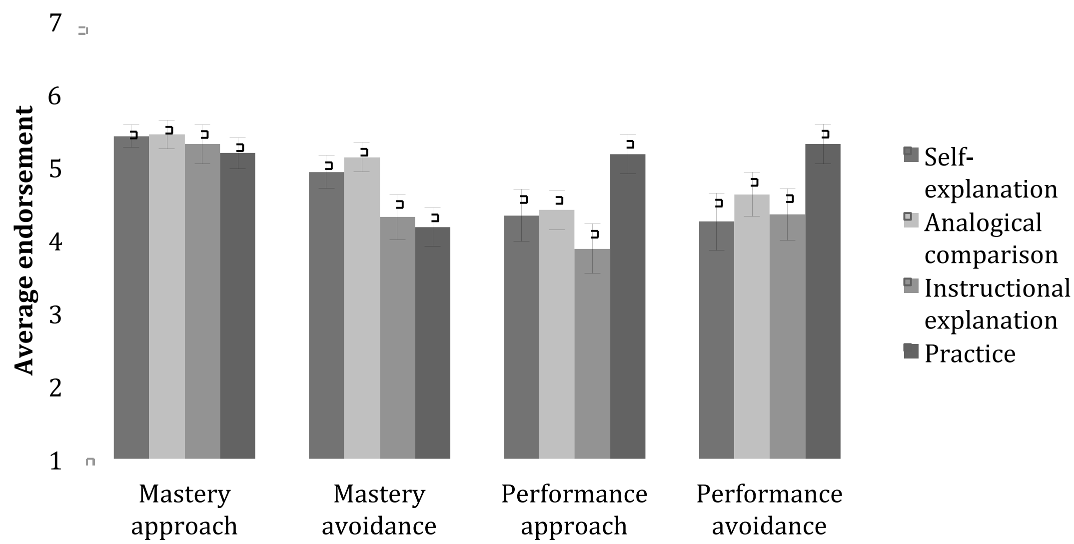


Figure 10. Condition effects on achievement goal endorsement.

Although each item on the learning and test processes questionnaire targeted a different behavior, all processes were expected to be correlated. To test the effect of condition on the sum of variables, I conducted a multivariate analysis of variance (MANOVA). Results indicated a marginal multivariate effect of condition, Wilks' $\lambda = .67$, $F(3, 105) = 1.39$, $p = .090$, $\eta^2_p = .13$. Follow-up univariate analyses were conducted using a Bonferroni adjusted alpha levels .005 per test (.05/10). Results indicated a large effect on the learning phase item, "I compared what I saw in the problem packets to what I already knew from the tutorial," $F(3, 105) = 7.24$, $p < .001$, $\eta^2_p = .17$. A Tukey post hoc analysis revealed a significant difference between the practice condition ($M = 5.07$, $SD = 1.94$) and the self-explanation ($M = 6.37$, $SD = .79$; $p = .001$), analogical comparison ($M = 6.36$, $SD = .78$; $p = .001$), and instructional explanation conditions ($M = 6.26$, $SD = .98$; $p = .003$), with the practice condition reporting lower levels of comparison. There were no differences between the self-explanation and analogical comparison conditions ($p > .99$), the self-explanation and instructional explanation conditions ($p > .99$), or the analogical comparison and instructional explanation conditions ($p > .99$). There was no effect of condition on any other questionnaire item, Fs < 2.31, ps $> .080$.

5.0 DISCUSSION

Analyses of the learning materials showed that in general, participants across conditions had sufficient time to complete the materials and made some effort to respond when prompted to do so. The absence of a performance measure collected during the learning phase for three of the four conditions meant analyses could not be focused on only those participants who had successfully learned from the instruction (e.g., the top two-thirds of learning performance), as was done in some prior work (Nokes & Ohlsson, 2005). Most instruction includes some form of practice, so adding practice to the experimental design would also be pragmatically relevant for instructional practices.

Overall, few condition effects were detected, with no differences between analogical comparison and self-explanation. The only difference in test accuracy was detected between the instructional explanation and practice conditions on near-transfer problems, with the instructional explanation condition achieving a greater level of accuracy. Although this was not consistent with predictions, prior work has shown that instructional explanations can support learning, particularly when learners are unable to identify critical concepts to focus on when engaging self-explanations or comparison (Berthold & Renkl, 2009; Renkl, 2002). The instructional explanations focused learners on key strategies and underlying structures in the sequence extrapolation problem examples. The examples supplied a great deal of surface information that varied across problems, and learners who did not receive instructional explanations may have struggled to see past the surface information when generating explanations or comparisons.

There was also a difference between the instructional explanation and practice conditions on time required to solve the first far-transfer problem. The practice condition solved the problem more quickly, which was consistent with expectations that participants in the practice condition would acquire more procedural knowledge that supports quicker problem solving. However, it was not expected that such procedural knowledge would be useful on far-transfer items, and it is surprising that no such difference was detected on near-transfer items. Accuracy of participants in the practice condition was not greater than in the other conditions on the first far-transfer problem, so while they were solving more quickly, they may have just been giving up faster or not evaluating their work as carefully.

Completion pattern results were also somewhat inconsistent with hypotheses. Participants in the practice condition, which was expected to promote greater procedural knowledge, used more common patterns and fewer unique patterns than expected on the second far-transfer problem, which required generation of a novel rule. Common completion pattern use could be driven by knowledge of surface or structural relations for many of the other problems; however, as the second problem depended on discovery of a novel relation, knowledge of previous structural relations would not be much help unless that knowledge were highly abstracted. It is possible that participants in the practice condition acquired more proceduralized knowledge that caused them to use more of the same completion patterns even when those completion patterns were less efficient, such as on a far-transfer problem; however, if that is the case, it is surprising that the trend of the practice participants using more common patterns was not detected for any other problems

Analyses of achievement goals assessed at the task level suggested that learning condition affected participants' endorsement of several achievement goals. Participants in the practice condition endorsed performance-approach goals more strongly than participants in the instructional explanation condition, suggesting that training focused on getting answers to problems promoted more of an emphasis of performing well relative to others than training focused on studying solution steps. Participants in the analogical comparison condition endorsed mastery-avoidance goals more strongly than participants in the practice condition, suggesting that training focused on comparing and abstracting across examples promotes more of an emphasis on avoiding a failure to deeply understand content compared to training focused on getting answers to problems.

Finally, analyses on learning process endorsement suggested that participants in the practice condition were less likely to report comparing example problems to the content of the tutorial. This is not surprising, as all other instructional conditions called attention to the relation between tutorial information and the examples, either in the form of instructional explanations that made those connections for the participant or prompts that encouraged the participant to apply that knowledge.

It is possible that participants were unsure of how to respond to the analogical comparison and self-explanation prompts, thus reducing distinctions between the types of knowledge representations supported by each process. Much prior work examining self-explanation has employed more extensive training and modeling to increase the degree to which participants engage in productive self-explanation when prompted. Such training may increase the strength of the intervention and improve the likelihood of detecting differences if they exist.

Finally, the design choice to use an artificial problem-solving domain presented both opportunities and challenges. Sequence extrapolation problems allowed tighter control over the relation between examples and test items, permitting carefully manipulation of the levels of transfer across problems. Using such a domain also likely reduced variability in prior knowledge, since there is relatively little relevant prior knowledge and this sort of problem is not typically discussed in school. However, there are several major shortcomings that may have made such problems ill-suited for examining the present research questions. First, the issue of there being little relevant prior knowledge could impede two hypothesized self-explanation mechanisms: the revision of errors and gaps in prior knowledge and the building of connections between prior knowledge and new information. Second, the types of principles most useful in learning to do sequence extrapolation problems may not have been well suited to the mechanisms targeted by self-explanation and analogical comparison. Both self-explanation and analogical comparison are instructional techniques thought to support the identification of abstract principles that apply to novel content; in a domain where few abstract principles exist, engaging in processes that aim to identify such principles might not be a particularly fruitful activity. Finally, sequence extrapolation problems are far removed from the type of academic content typically studied in the classroom, and thus any conclusions drawn from this work may have limited applicability to the way students learn academic content in the classroom through self-explanation or analogical comparison. Thus, while the domain choice was appropriate for fine-grained questions about knowledge representations and levels of transfer, it may not have been an ideal choice for addressing the primary question of how learning outcomes supported by self-explanation and analogical comparison might differ. I now describe a second experiment designed to extend these findings in a rich scientific domain.

6.0 EXPERIMENT 2: SELF-EXPLANATION AND ANALOGICAL COMPARISON OF SCIENCE MATERIALS

Experiment 1 showed inconsistent results, with relatively few distinctions emerging between self-explanation and analogical comparison conditions. As discussed above, results may have reflected a number of particular details related to the experimental design and measures, including the lack of training in each technique and the choice of content domain. Alternatively, the absence of differences between self-explanation and analogical comparison conditions might indicate that both support the same types of knowledge and performance. To assess whether the lack of differences in results stemmed from poor training implementation and other experimental design shortcomings or a genuine lack of difference in knowledge supported by the two techniques, I conducted a second experiment. Experiment 2 aimed to apply and extend results to a conceptually rich content domain, which may provide a better reflection of the types of knowledge and learning processes activated by different instructional techniques. Specifically, I conducted an experiment in which learners studied text about electricity and electric circuits, and engaged in self-explanation, analogical comparison, or study of instructional explanations illustrating relevant concepts. For example, students likely have more prior knowledge regarding physics content compared to sequence extrapolation problems. Therefore the mechanisms of building connections to prior knowledge and identifying gaps or contradictions in one's knowledge likely will be much more relevant.

While Experiment 1 assessed achievement goal and learning process self-reports, Experiment 2 aimed to improve understanding of those results by introducing more robust measures. A domain-framed achievement goal questionnaire was included in addition to the task-framed questionnaire to better identify the extent to which condition assignment changed participants' goals relative to their general goals for the domain. The learning process items employed in Experiment 1 aimed to capture a wide range of learning and test strategies with just one item per construct. Experiment 2 employed a series of learning process questionnaires developed in other work (Zepeda & Nokes-Malach, 2015) as an alternative to protocol analysis for quantifying the degree to which different individuals engaged specifically in processes of self-explanation and analogical comparison. Most studies of self-explanation involve extensive analysis of learners' written or verbal protocols, constraining the study of self-explanation to research environments or tasks developed for the purpose of collecting protocols. A questionnaire of self-explanation could be deployed more easily across a variety of academic settings and might improve understanding of why some students are more successful in acquiring concepts and revising misconceptions than others. Although analogical comparison research has employed protocol analysis less frequently, much prior work has assessed analogy use through observational data (e.g., Richland, Holyoak, & Stigler, 2004) and experimental manipulations (e.g., Gick & Holyoak, 1983). As with the study of self-explanation, developing unobtrusive measures of analogical comparison could also improve understanding of how frequently learners engage in analogical comparison, and it could help explain why some students learn and transfer deep concepts more successfully than others. No prior work that I know of has attempted to create a questionnaire assessing students' use of self-explanation or analogical comparison. The questionnaire also targeted metacognitive processes of monitoring, control, and evaluation. The processes of self-explanation in particular are closely related to certain metacognitive processes, but the relation has not been extensively examined in prior work. Including metacognitive questionnaire items will permit closer examination of whether instructions to self-explain lead to greater reports of metacognitive behaviors.

Research investigating both analogical comparison and self-explanation has highlighted the importance of prior knowledge (Gadgil et al., 2012; Hausmann & VanLehn, 2007; Rittle-Johnson, Star, & Durkin, 2009). In the case of analogical comparison, the amount of prior knowledge a learner possesses can determine what features should be highlighted through comparison to promote optimal learning (Rittle-Johnson et al., 2009), and in at least some cases a certain amount of relevant prior knowledge is necessary for conceptual learning from comparison (Star & Rittle-Johnson, 2009). In the case of self-explanation, two of the primary mechanisms through which it supports learning are filling gaps in prior knowledge and prompting the revision of inaccuracies in prior knowledge (Nokes et al., 2011). While prior knowledge supports learning in both cases, the mechanisms of self-explanation seem to depend more on activating and modifying prior knowledge relative to analogical comparison. Understanding the role and relative importance of prior knowledge for both instructional techniques could provide useful conditions for when learners are more likely to benefit from one instructional technique than the other. For this reason, I included a pretest measuring students' prior knowledge in the domain of electricity and electric current, and I examined whether the relation between pretest and posttest scores differed across conditions, which might suggest that prior knowledge played a different role depending on the instructional condition.

For both processes, the types of knowledge produced depend a great deal on the design of the materials, such as the amount of analogical comparison support provided (Gentner et al.,

2003) or whether self-explanation prompts focus on filling gaps in knowledge or revising mental models (Nokes et al., 2011). I therefore aimed to control factors such as the amount of scaffolding provided (introduction to the process, modeling, and prompting) and the target of the prompts to self-explain or compare (worked examples). Controlling for such factors could provide clearer evidence about the relative support each process provides for different learning outcomes. Aside from the change of content domain and addition of questionnaires, conditions and measures were designed to mirror those used in Experiment 1 as closely as possible. I hypothesized that the type of instructional activity learners engaged in (self-explanation, analogical comparison, reading instructional explanations) would predict participants' representations of knowledge, which in turn would predict the levels of transfer they demonstrated on a posttest. I also expected participants to report engaging in the activities prompted by their condition instructions more than participants in other conditions. Finally, I expected prompted instructional activities to change participants' achievement goals as they related to the learning tasks and the degree to which they reported engaging in metacognitive behaviors. Detailed predictions are included in the methods.

7.0 METHOD

7.1 PARTICIPANTS

One hundred and one college students enrolled in an introductory psychology course at the University of Pittsburgh participated in the study. Participants received three credits toward a research participation requirement associated with the course. The majority of participants were in their first (n = 68) or second (n = 20) year of college, with only a few in their third (n = 7), fourth (n = 4), or fifth years (n = 2). Eighty-eight of 101 participants reported having taken a physics course either in high school or college.

7.2 DESIGN

The experiment had a between-subjects design and participants were randomly assigned to one of three conditions: self-explanation, analogical comparison, or instructional explanations.

7.3 MATERIALS

All participants completed the same basic learning materials, tests, and questionnaires. Each condition received a different introductory page of the learning materials that highlighted the benefits of analogical comparison, self-explanation, or instructional explanation study, and different instructions targeting a series of worked examples at the end of each learning booklet, which modeled and prompted either analogical comparison, self-explanation, or study of instructional explanations. I describe these differences in greater detail below.

7.3.1 Learning materials

Four booklets of instructional materials were adapted from a prior study by Richey and Nokes-Malach (2013) and covered concepts related to electricity and electric circuits. Most college students have had prior exposure to these concepts yet still hold a number of misconceptions about the topic (Slotta & Chi, 2006). Thus, the topic was well suited for examining types of transfer and included features, principles and relations that could be identified through analogical comparison or self-explanation. The topic was also appropriate for measuring near transfer, intermediate transfer, far transfer, and PFL transfer.

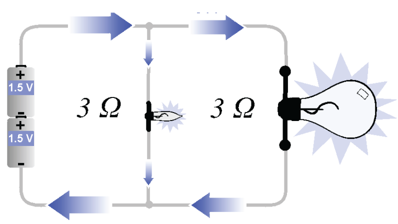
All booklets were printed on paper, and participants were encouraged to mark up the booklets or make notes. Each booklet contained several pages of instructional text followed by worked examples and practice problems related to the preceding text. Specifically, the first booklet contained two pages of instructional text, two worked examples, and one practice problem. The second booklet contained two pages of instructional text, four worked examples, and two practice problems. The third booklet contained three pages of instructional text, two worked examples, and one practice problem. The fourth booklet contained two pages of instructional text, six worked examples, and two practice problems.

Booklets differed across conditions in the introductory instructions participants received and the prompts included before each worked example at the end of each booklet (Appendix A). Conditions were reinforced with a brief description of the utility of the targeted process at the beginning of the first booklet and an example response to the prompts after the first worked example. After the first worked example prompt in the first booklet, the self-explanation condition received written modeled explanations illustrating monitoring, bridging, and elaborating statements about the worked example. Modeling self-explanation has been shown to improve the quality and effectiveness of responses (McNamara, 2004). After studying the modeled responses, participants in the self-explanation condition were prompted after the next worked example to write their own monitoring, bridging, and elaborating statements. The analogical comparison condition received written modeled comparison statements identifying similarities across the first two worked examples and discussing the conceptual significance of those similarities. After studying the modeled response, participants were prompted to elaborate by identifying additional similarities. The instructional explanation condition saw the same prompt included before every example in that condition with no additional modeling, as the prompts did not require a written response.

Worked examples were created in pairs with surface dissimilarities (e.g., different values, cover stories) and either the same or contrasting relations. Each pair of worked examples also had a corresponding practice problem with surface dissimilarities but the same relations. Worked example pairs were presented side-by-side on the same page in the analogical comparison condition. To suppress spontaneous comparison, they were presented on sequential pages in the self-explanation and instructional explanation conditions. Instructional explanations focused on concepts related to each example step and were similar to elaborative explanations participants in the self-explanation condition were expected to generate on their own (Schworm & Renkl, 2006). They were included to suppress spontaneous self-explanation in the instructional explanation condition and control the amount of information reviewed across conditions while manipulating the processes (reading, self-explaining, comparing). Figure 11 shows a worked example from the instructional explanations condition; the self-explanation and analogical comparison conditions saw the same example with the step-by-step solution (right column) but without the instructional explanations (left column). The analogical comparison condition saw the example side-by-side with the next example, which asked the same question and included a diagram of a series circuit with two 3-ohm light bulbs.

7.3.2 Test materials

A five-item pretest and 36-item posttest were administered to measure participants' learning. The posttest consisted of both multiple-choice and short-answer questions and contained 13 near-transfer items ($\alpha = .33$), 17 intermediate-transfer items ($\alpha = .61$), 12 far-transfer items ($\alpha = .44$), and nine preparation for future learning (PFL) transfer items ($\alpha = .53$). Items were coded for transfer level based on the definitions of transfer based on Barnett and Ceci's (2002) and Nokes-Malach et al.'s (2013) models of transfer and their relationship to the materials in the instructional text and worked examples. For example, the worked example in Figure 11 corresponded to several test problems, including a near-transfer question asking about current in a two-branch parallel circuit with new values for resistance (identical problem-solving procedure); an intermediate-transfer question asking about resistance in a three-branch parallel circuit (problem-solving procedure requires abstraction to be applied to circuit with a different number of branches and to solve for a different variable); and a far-transfer question asking how total current changes in a parallel circuit when additional branches are added (problem-solving procedure requires inference and reasoning to determine abstract relationship between number of branches and total current). An additional learning resource about power was embedded in the test and provided information for answering the PFL transfer items, which targeted how well participants were prepared to learn from a new instructional resource about a related topic (Bransford & Schwartz, 1999).



5. What type of circuit is this?

Calculate the current in each branch based on the information given about the circuit, as well as the total current in the circuit.

Solution

General principle applied: Identify the type of circuit. A series circuit has one path and current is the same at every point. A parallel circuit has multiple paths and current can differ across paths.	*This is a parallel circuit. Therefore, we must calculate the current through each path of the circuit.*
Define values and relations: Voltage is the same across each branch because all the branch points are on the same wire.	*This is a parallel circuit, so voltage is the same across each branch. Voltage is 3 V in each branch.*
Define values and relations: Resistance (R) measures how difficult it is for electrons to move in a circuit, or the opposition to the movement of current (I). Resistance can differ across branches.	*Branch 1 has resistance of 3 Ω and branch 2 has resistance of 3 Ω. Use Ohm's law to calculate current in each branch and Kirchhoff's current law for total current.*
Solve based on values and principle: Set up separate equations solving for current in each branch and total current.	Branch 1 current: $I_1 = V \div R_1$ Branch 2 current: $I_2 = V \div R_2$ Total current: $I_T = I_1 + I_2$
Solve based on values and principle: To solve for current in each branch current, divide voltage by resistance. Total current is the sum of current across all branches.	$I_1 = 3\ V \div 3\ \Omega = 1$ A $I_1 = 3\ V \div 3\ \Omega = 1$ A $I_T = 1\ A + 1\ A = 2\ A$

Figure 11. Worked example with instructional explanations.

Two independent coders coded all short-answer items using a rubric, discussed any differences, and reached 100 percent agreement for all items. Instructions to self-explain or compare worked examples were expected to support greater intermediate, far, and PFL transfer than studying worked examples with instructional explanations, as both self-explanation and analogical comparison support the identification of deep, conceptual features and generation of abstract principles that more easily transfer to new situations. Analogical comparison may reduce learning of the concrete procedures that support problem-solving performance.

7.3.3 Questionnaires

A series of questionnaires assessed participants' self-reported use of the learning processes targeted in this experiment, their metacognitive behaviors, and their achievement goals as they relate to science in general and the particular learning tasks completed in this experiment.

7.3.3.1 Open-ended learning processes question

Participants responded to a question asking them to describe in as much detail as possible the learning processes they used while studying the learning materials. Participants' responses, which were generally several sentences long, were used to capture the degree to which learners described engaging in behaviors consistent with self-explanation, analogical comparison, and studying instructional explanations. It was also included to capture additional strategies or behaviors that might provide a more detailed picture of individual differences in learning behaviors and unexpected behaviors (i.e., those not intentionally encouraged through the learning prompts) that might nevertheless lead to greater learning. While participants' descriptions of their learning processes were expected to show some consistencies with their condition assignments (e.g., participants assigned to the self-explanation condition were expected to use more language related to the processes associated with self-explanation), some individuals might also discuss other processes associated with better learning outcomes.

7.3.3.2 Learning processes questionnaire

Zepeda and Nokes-Malach (2015) created the learning-processes questionnaires by identifying critical features of self-explanation and analogical comparison. Ten items examined students' use of self-explanation, e.g., "During the activity, as I solved a problem I would explain to myself what concepts were being applied and why" ($\alpha = .83$), and 11 items examined analogical comparison, e.g., "During the activity, I often compared and contrasted one part to a previous part of the text or problem" ($\alpha = .90$). All items were framed at the task level, and participants rated how much they agreed or disagreed with each item on a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree). I expected instructions to self-explain or compare worked examples during the learning phase to lead to greater self-reports of engaging in self-explanation and analogical comparison, respectively. Additionally, I predicted that self-reports of engaging in self-explanation or analogical comparison will predict test performance above and beyond the differences explained by condition assignment, with analogical comparison leading to worse performance on near-transfer questions and both analogical comparison and self-explanation supporting better performance on intermediate-, far-, and PFL-transfer items.

7.3.3.3 Metacognitive questionnaire

Zepeda and Nokes-Malach (2015) selected items for the metacognitive questionnaire from a number of existing measures of metacognitive behaviors; the questionnaire was designed to target three components of metacognition: monitoring, e.g., "During the activity, I tried to determine which concepts I didn't understand well" ($\alpha = .87$), control, e.g. "During the activity, I changed strategies when I failed to understand the problem" ($\alpha = .79$), and evaluation, e.g., "During the activity, I reviewed what I had learned" ($\alpha = .82$). All items were framed at the task level, and participants rated how much they agreed or disagreed with each item on a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree).

7.3.3.4 Achievement goals questionnaires

Elliot and colleagues' Achievement Goals Questionnaire-Revised (AGQ-R; Elliot & Murayama, 2008) was used to assess participants' self-reported, domain-level achievement goals framed around science class. The questionnaire assessed participants' endorsement of mastery-approach ($\alpha = .91$), mastery-avoidance ($\alpha = .76$), performance-approach ($\alpha = .89$), and performance-avoidance goals ($\alpha = .93$). Responses were recorded on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree).

A second questionnaire was modified slightly to assess participants' self-reported goals at the task level. Specifically, the instructions asked participants to consider their feelings about the activities they completed during the learning phase. The questionnaire assessed participants'' self-reported endorsement of task-level mastery-approach ($\alpha = .80$), mastery-avoidance ($\alpha = .58$), performance-approach ($\alpha = .94$), and performance-avoidance goals ($\alpha = .91$). Responses were recorded on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree).

7.4 PROCEDURE

Participants completed the experiment individually in sessions of three to five students at a time; they were not permitted to talk to each other during the experiment. Following completion of a brief pretest, participants began working individually through the self-paced learning booklets. Participants were notified of a time limit for each booklet of the learning materials (15 minutes for the first, 20 for the second, 25 for the third, and 30 for the fourth) and booklets were distributed one at a time. While participants could flip back or ahead within each booklet, they could not go back to a previous booklet and could not move ahead until everyone in the room had finished the current materials. Participants could not move on to the next booklet until everyone in a session had completed the prior booklet. Upon completing the learning booklets, participants responded to the learning processes questionnaires, as well as a metacognition questionnaire and task-framed achievement goals questionnaire. Participants then were given 55 minutes to complete the posttest, followed by the domain-framed achievement goals questionnaire. Most sessions used the majority of the time allotted, and there were no effects of condition on learning time, $F(2, 98) = 1.40$, $p = .25$, $\eta_p^2 = .028$, or test time, $F(2, 98) = .25$, $p = .78$, $\eta_p^2 = .005$.

8.0 RESULTS

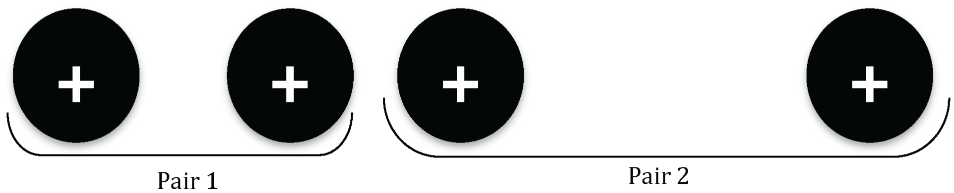
Analyses focused on testing the effect of learning condition on performance at each level of transfer assessed by the posttest as well as responses on the learning processes and achievement goal questionnaires. I also examined the relations between participants' self-reported learning processes and posttest performance. Posttest performance is reported as a proportional accuracy representing the number of correct items out of the total number of items for each level of transfer.

8.1.1 Assessment of learning quality

To better understand what cognitive processes were triggered by the learning packet prompts, I coded responses for all participants in the self-explanation and analogical comparison conditions. In the self-explanation condition, responses were coded on a scale of zero to four, ranging from no response at all, to a response that merely paraphrased content already provided in the example, to a response that went beyond what the example stated (e.g., through connections to other content in the instructional packets, references to prior knowledge or experiences, statements monitoring understanding). This coding is consistent with how Hausmann and VanLehn (2010) differentiated between paraphrasing and self-explaining. Table 5 includes self-explanation responses that received a range of scores targeting the worked example shown in Figure 12.

Worked examples:

3. Which pair of charges has a greater electrical potential energy between them?



Solution

All objects in the question are charged so they all have electric fields.
The charges are pairs of the same kind, so they should repel one another.
The charges in Pair 1 have been moved in the opposite direction of the electrical force more than the charges in Pair 2.
The charges in Pair 1 have greater electrical potential energy between them.

Figure 12. Worked example that served as the target of prompt responses included in Table 5.

Table 5.

Rubric parameters and examples of prompt responses directed at the worked example shown in Figure 12 for scores 1-4.

Scale	Rubric	Participant example
1	*Wrote very short notes about content of WE – maybe naming principle or describing briefly*	Pair 1 is the same charge +. Close together so greater PE.
2	*Wrote more but exclusively about content of WE, maybe put WE content together in different way; no mention of ideas from instructional materials, prior knowledge, monitoring, etc OR mentioned other ideas but in superficial way with no new insights.*	Electrical potential energy depends on charge and distance. Similar charge will repel; pair one will repel but have higher electrical potential energy because they repel but are close to one another.
3	*Wrote primarily about ideas presented*	The figure titled “Electrical Potential Energy

	directly in WE, with minor mention of ideas either from instructional materials, prior knowledge, info about understanding. New inferences or insights clearly present.	and Reliable Position” shows that like charged particles have more potential energy when they are closer together. Thus, Pair 1 has a greater electrical potential energy between them.
4	*Wrote primarily about ideas not presented directly in WE (e.g., from instructional text, prior knowledge, metacognitive reflections). New inferences or insights clearly present.*	Pair 1 – Repulsion: 2 of the same charge. Less distance between particles → more work required to keep particles in position (gets more of electric field) → greater electric potential energy. Pair 2 – Repulsion: 2 of the same charge. More distance between particles → Less work required to keep particles in position (gets less of electric field) → less electric potential energy

The example receiving a score of one briefly restates details from the worked example (that both charges are positive, that Pair 1 has greater electrical potential energy), while the example receiving a score of two more elaborately summarizes the points in the worked example (that charge and distance are both factors determining electrical potential energy). These are important details, but they are specific to the worked example and do not generalize toward any principles relating electrical charge, distance, or potential energy, nor do they relate the details in the examples to participants’ prior knowledge or beliefs about the topic. Since the responses do not demonstrate any of the mechanisms hypothesized to support learning and knowledge revision through self-explanation, I would not expect participants generating these types of response to experience any learning advantages as a result of receiving self-explanation prompts; moreover, participants producing such responses would likely learn less than those who received more detailed instructional explanations, which lack a constructive element but highlight more general principles.

The example receiving a score of three makes a generalization (“like charged particles have more potential energy when they are closer together”), though this generalization is not a principle that could be applied to other types of charges. The example receiving a score of four identifies a more general principle (the pair that requires more work to keep the charges in position has greater electrical potential energy) that could be applied to any two pairs of charges to determine their relative electrical potential energy. Both examples demonstrate some of the generalization process thought to support greater learning through self-explanation, so I would expect participants' generating such responses to demonstrate greater learning relative to participants in the same condition receiving lower scores as well as participants in the instructional explanations condition, which received but did not generate principles. The process of abstracting generalized information from the worked examples and the generalized information itself should support test performance.

For participants in the analogical comparison condition, I coded responses on a scale of zero to four, ranging from no response, to listing a similarity or comparison with no elaboration on its significance, to listing similarities and differences with discussion of why those similarities or differences were meaningful. Table 6 includes analogical comparison responses that received a range of scores targeting the worked example shown in Figure 13.

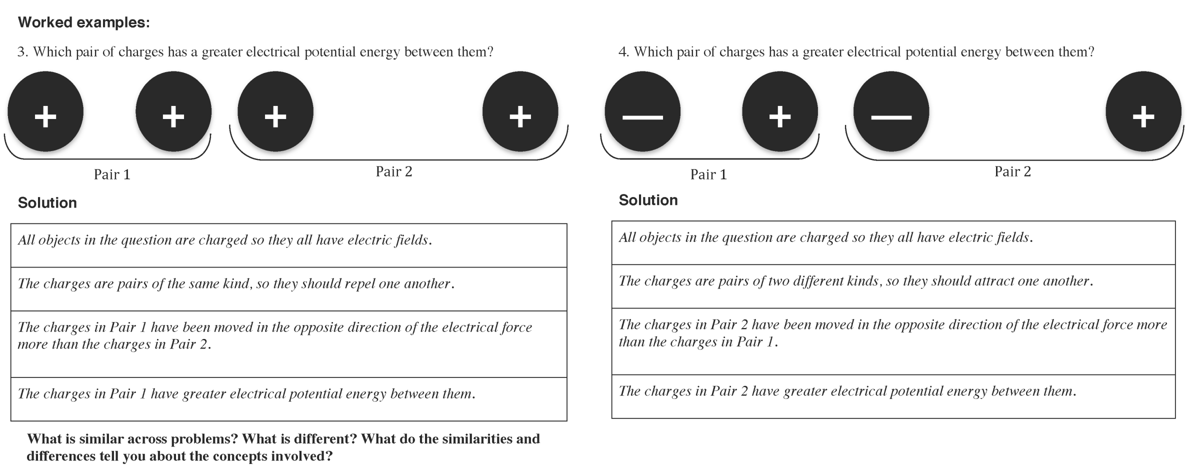


Figure 13. Worked example that served as the target of prompt responses included in Table 6.

Table 6.

Rubric parameters and examples of prompt responses directed at the worked example shown in Figure 13 for scores 1-4.

1	*Identified one similarity OR difference without explaining or discussing meaning*	One pair repels each other, the other pair attracts
2	*Identified one similarity AND one difference without explaining or discussing meaning*	Both problems give 2 different pairs of charges and ask which one has the greater electrical potential energy. In both examples, similar steps are taken and very similar qualities of charges and electrical forces are identified. However, the final answers are different because different pairs in each example are used.
3	*Identified one similarity OR one difference and elaborated on its significance, meaning, etc*	The first problem deals with like charges while the second deals with unlike charges. If the charge is the same, less distance means higher potential energy. If the charges are different, less distance means less potential energy.
4	*Identified one similarity AND one difference and elaborated on their significance, meaning, etc.*	Similar – In both problems all objects are charged and have electrical fields. They all exert forces on each other. Different – In 1 they are like charges so they repel each other meaning 1 has more PE. In 2 they are different charges so they attract meaning 2 has more PE. Concept – PE relies on both distance and charge. In like charges the distance is small for high PE in different charges distance is large for high PE.

The example receiving a one states a difference between the two examples, but it does not identify why this difference is relevant to the determination of electrical potential energy. The example receiving a two states a number of similarities and differences between the steps but again fails to relate those similarities and differences to an underlying principle. Analogical comparison is hypothesized to promote learning through the recognition of key principles highlighted through similarities or differences between examples; thus, participants producing these types of responses are not expected to demonstrate any learning advantages as a result of receiving analogical comparison prompts. Further, because participants in the instructional explanation condition received principles accompanying each example, they would be expected to perform better on the posttest than participants receiving mostly ones or twos on their prompt responses.

The example receiving a three identifies a key difference relating types of charge, distance, and electrical potential energy. This difference suggests a principle that could be applied to new cases and goes beyond what is stated in the worked examples. The example receiving a four identifies the same principle relating types of charge, distance, and electrical potential energy, but it also identifies key similarities across the examples. Both of these responses identify deep principles through the act of comparing examples, one of the underlying mechanisms thought to promote learning through analogical comparison. Consequently, participants who create such responses are more likely to exhibit the learning benefits expected from engaging in analogical comparison and should perform better on the posttest than those receiving instructional explanations.

To examine the effect of quantity and quality separately in analyzing prompt responses, I also coded the number of words participants wrote on each worked example page. Table 7 shows correlations between posttest performance (total accuracy and by transfer level), average prompt response score, and average word count per prompt. Correlations were run separately for each condition, as the content of the prompts differed across conditions. For the self-explanation condition, quality of prompt response was significantly, positively correlated with near-transfer performance and marginally correlated with total test performance. Number of words written was marginally, negatively correlated with intermediate-transfer and total posttest performance. No other correlations were significant for either condition.

Table 7.

Correlation (r) between prompt response scores, average word count, and test performance by condition.

	Word count	Posttest	Near	Intermediate	Far	PFL	*M*	*SD*
Self-explanation								
Prompt score	-.10	.31†	.38*	.015	.24	.22	1.74	1.30
Word count	—	-.34†	-.26	-.32†	-.099	-.20	29.38	13.61
Analogical comparison								
Prompt score	-.002	.16	.081	.23	.047	.039	1.66	1.32
Word count	—	.15	.17	.014	.17	.14	38.83	15.86

8.1.2 Condition effects on learning

I conducted a series of one-way analyses of variance (ANOVAs) to assess the effect of condition on the pretest and posttest. Pretest performance was calculated as an accuracy score out of one by dividing the number of correct pretest items by the total number of items. Learning packet performance was calculated as the total number of points awarded for correct answers to learning packet problems; each problem was worth 1 point except the problem in the third packet, which had two possible answers and thus participants could receive a total of 2 points if they wrote both correct answers. No condition effects were found for pretest, $F(2, 98) = .64$, $p = .53$, $n_p^2 = .013$, or learning packet accuracy, $F(2, 98) = 0.078$, $p = .93$, $n_p^2 = .002$, and thus neither measure was used as a covariate in posttest analyses. Descriptives for both are included in Table 8.

Table 8.

Mean and standard deviation of pretest and learning packet accuracy by condition.

	Pretest			**Learning packets**		
	N	*M*	*SD*	*N*	*M*	*SD*
Self-explanation	34	.49	.23	34	5.96	.84
Analogical comparison	34	.42	.21	34	5.94	.72
Instructional explanation	33	.46	.27	33	5.88	1.06
Total	101	.46	.24	101	5.93	.88

Participants received a point for each posttest item answered correctly; for quantitative questions, they received a point for the numerical value and a point for the units. Accuracy was calculated as the total number of points a participant received divided by the total number of points possible, which was 51. Accuracy was also calculated separately for near-transfer, intermediate-transfer, far-transfer, and PFL-transfer items. A one-way ANOVA revealed a medium effect of condition on total posttest accuracy, $F(2, 98) = 3.22$, $p = .044$, $n_p^2 = .062$ (Figure 10). Post hoc comparisons using the Tukey HSD test indicated that the mean score for the instructional explanation condition ($M = .63$, $SD = .10$) was significantly different from the analogical comparison condition ($M = .56$, $SD = .12$), $p = .034$, $\eta_p^2 = .062$. However, the self-explanation condition ($M = .59$, $SD = .094$) did not significantly differ from the instructional explanation ($p = .39$) or analogical comparison conditions ($p = .43$).

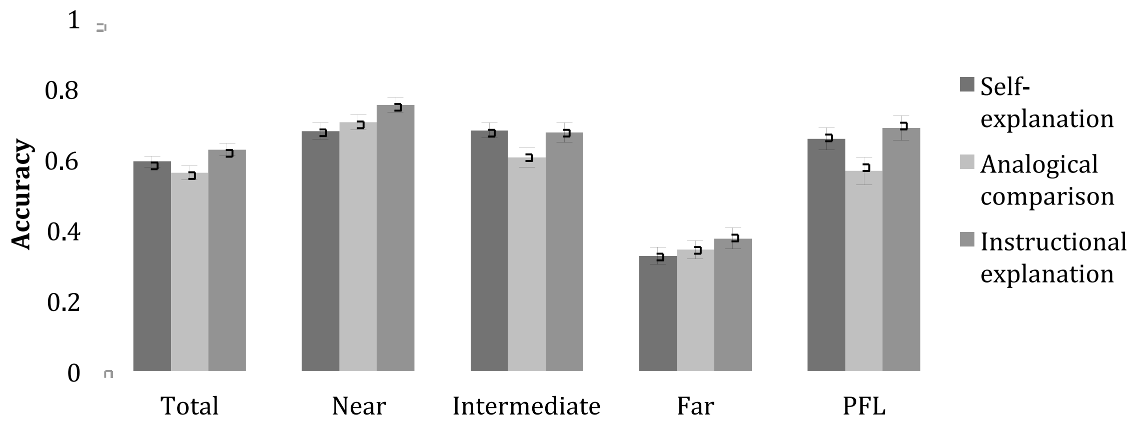


Figure 14. Posttest accuracy on subsets of items classified by transfer distance.

Looking at subsections of the posttest based on transfer level, there was a marginal effect of condition on near-transfer accuracy, $F(2, 98) = 2.91$, $p = .059$, $\eta_p^2 = .056$. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the instructional explanation condition ($M = .75$, $SD = .12$) was marginally different from the self-explanation condition ($M = .68$, $SD = .14$), $p = .051$. However, the analogical comparison condition ($M = .70$, $SD = .12$) did not significantly differ from the instructional explanation ($p = .26$) or self-explanation conditions ($p = .70$).

There was also marginal effect of condition on intermediate-transfer accuracy, $F(2, 98) = 2.89$, $p = .060$, $\eta_p^2 = .056$. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the analogical comparison condition ($M = .60$, $SD = .16$) was marginally different from the self-explanation condition ($M = .68$, $SD = .12$), $p = .081$. However, the instructional explanation condition ($M = .67$, $SD = .15$) did not significantly differ from the self-explanation ($p = .98$) or analogical comparison conditions ($p = .12$). There was no effect of condition on far-transfer accuracy, $F(2, 98) = 0.83$, $p = .44$, $\eta_p^2 = .017$.

There was a medium effect of condition on PFL performance, $F(2, 98) = 3.34$, $p = .039$, $\eta_p^2 = .064$. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the instructional explanation condition (*M* = .69, *SD* = .20) was significantly different from the analogical comparison condition (*M* = .57, *SD* = .22), *p* = .040. However, the self-explanation condition (*M* = .66, *SD* = .18) did not significantly differ from the instructional explanation (*p* = .81) or analogical comparison conditions (*p* = .15).

8.1.3 Open-ended learning processes question

Participants' open-ended learning processes descriptions were analyzed using Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, & Francis, 2007) dictionaries created to assess word-use related to cognitive processes (analogical comparison, self-explanation, and worked-example study), metacognitive processes (monitoring, control, and evaluation), and achievement goals (mastery goals, performance goals, approach goals, and avoidance goals). Dictionaries for each construct were created based on the critical features and mechanisms of each process or goal dimension and are included in Appendix B.

For cognitive processes, there was a trend such that each condition reported more use of the process associated with their condition. A series of one-way ANOVAs were conducted to assess effects of condition on word use across the different dictionaries (Figure 11). There was a significant effect on example word use, $F(2, 98) = 3.44$, $p = .036$, $\eta_p^2 = .066$, and post hoc comparisons using the Tukey HSD test indicated that the mean score for the instructional explanation condition (*M* = 5.80, *SD* = .59) was different from the self-explanation condition (*M* = 3.63, *SD* = .58), *p* = .027. However, the analogical comparison condition (*M* = 4.57, *SD* = .58) did not significantly differ from the instructional explanation (*p* = .30) or self-explanation conditions (*p* = .49). There was also a marginal effect on explanation word use, $F(2, 98) = 2.40$, $p = .096$, $\eta_p^2 = .047$, and post hoc comparisons using the Tukey HSD test indicated that the mean score for the self-explanation condition (*M* = .83, *SD* = .17) was marginally different from the instructional explanation condition (*M* = .31, *SD* = .18), $p = .081$. However, the analogical comparison condition (*M* = .44, *SD* = .17) did not significantly differ from the self-explanation ($p = .26$) or analogical comparison conditions ($p = .85$). There was no effect on comparison word use, $F(2, 98) = 1.95$, $p = .15$, $\eta_p^2 = .038$.

For metacognitive process words, there was a medium effect of condition on use of monitoring words, $F(2, 98) = 4.59$, $p = .012$, $\eta_p^2 = .086$, and post hoc comparisons using the Tukey HSD test indicated that the mean score for the self-explanation condition (*M* = 2.27, *SD* = .30) was different from the instructional explanation condition (*M* = .97, *SD* = .31), $p = .009$. However, the analogical comparison condition (*M* = 1.53, *SD* = .30) did not significantly differ from the self-explanation ($p = .20$) or analogical comparison conditions ($p = .40$). There was also a medium effect of condition on evaluation words, $F(2, 98) = 5.62$, $p = .005$, $\eta_p^2 = .10$, and post hoc comparisons using the Tukey HSD test indicated that the mean score for the analogical comparison condition (*M* = .79, *SD* = .32) was different from the instructional explanation condition (*M* = 2.30, *SD* = .32), $p = .003$. However, the self-explanation condition (*M* = 1.39, *SD* = .32) did not significantly differ from the instructional explanation ($p = .11$) or analogical comparison conditions ($p = .39$). There was no effect of condition on control words, $F(2, 98) = .235$, $p = .79$, $\eta_p^2 = .005$.

For goal words, there were no significant effects of condition on the use of mastery, $F(2, 98) = .67$, $p = .52$, $\eta_p^2 = .013$, performance, $F(2, 98) = .25$, $p = .78$, $\eta_p^2 = .005$, or approach words, $F(2, 98) = 1.82$, $p = .17$, $\eta_p^2 = .036$. No use of avoidance words was found for any participant.

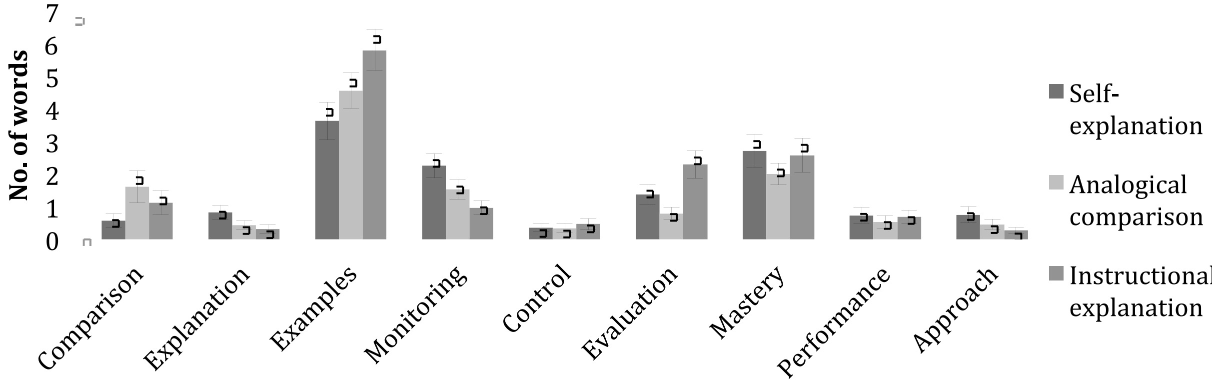


Figure 15. Use of cognitive and metacognitive processes and goal language.

Next, I examined correlations between word use in participants' open-ended learning process responses and their posttest performance (Table 9). Relatively few cognitive, metacognitive, or goal words were associated with any measure of posttest performance, and no consistent patterns emerged. There were marginal, positive correlations between comparison word use and far transfer, example word use and total accuracy, evaluation word use and intermediate transfer, and mastery goal word use and near transfer. These results are consistent with general predictions that engaging in explanation, comparison, and evaluation would promote learning and test performance, as would mastery goals. There was also a significant, negative correlation between control word use and intermediate transfer, and a marginal, negative correlation between approach goal word use and near transfer.

Table 9.

Correlations between cognitive, metacognitive, and goal word use and posttest performance.

	Total accuracy	**Near transfer**	**Intermediate transfer**	**Far transfer**	**PFL transfer**
Comparison	.16	.042	.10	.17†	.13
Explanation	.14	.073	.16	-.026	.15
Examples	.19†	.082	.14	.13	.15
Monitoring	.016	-.11	.12	-.038	.027

Control	-.091	-.022	-.21*	.024	.019
Evaluation	.073	.022	.17†	-.088	.057
Mastery	-.013	.19†	-.14	.044	-.070
Performance	.12	.038	.12	.09	.075
Approach	-.084	-.17†	-.010	-.080	-.002

8.1.4 Learning processes questionnaire

Participants' responses to each survey item were averaged within each construct to create scores for cognitive (explanation and comparison), metacognitive (monitoring, control, and evaluation), and motivation variables (mastery approach, mastery avoidance, performance approach, and performance avoidance). I examined condition effects on survey responses for each construct, the relation between survey responses and posttest performance, and the correlation between survey responses and words used in the open-ended learning processes descriptions.

A one-way ANOVA revealed a medium effect of condition on reported use of metacognitive processes, $F(2, 98) = 5.20$, $p = .007$, $\eta_p^2 = .096$ (Figure 12). Post hoc comparisons using the Tukey HSD test indicated that the mean score for the instructional explanation condition ($M = 5.22$, $SD = .96$) was significantly different from the self-explanation condition ($M = 4.51$, $SD = .92$). However, the analogical comparison condition ($M = 4.80$, $SD = .83$) did not significantly differ from the instructional explanation ($p = .14$) or self-explanation conditions ($p = .40$). There was no effect of condition on self-reported use of comparison, $F(2, 98) = .40$, $p = .67$, $\eta_p^2 = .008$, or explanation, $F(2, 98) = 1.14$, $p = .33$, $\eta_p^2 = .023$.

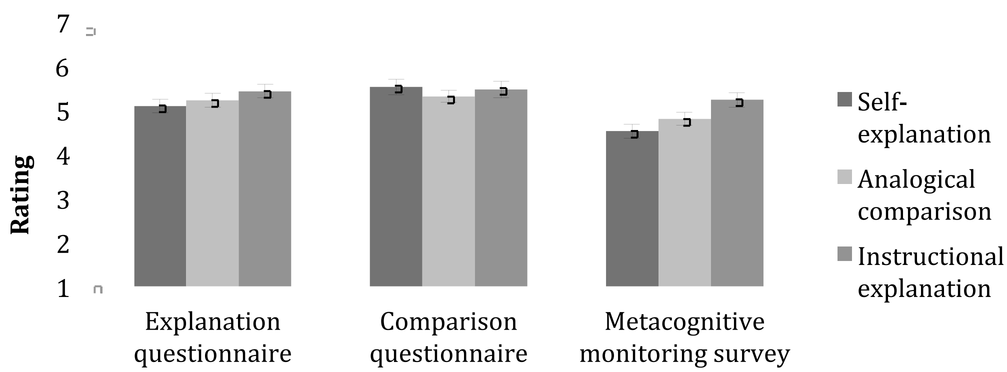


Figure 16. Endorsement of cognitive and metacognitive processes by condition.

To test the amount of variance in posttest performance explained by self-reported self-explanation, analogical comparison, and metacognition, variance due to the condition assignment was removed using hierarchical multiple regression. Condition was dummy-coded with the instructional explanation condition as the reference group. Self-reported levels of self-explanation, analogical comparison, and metacognition, were entered in a step-wise fashion into the second model with the first model containing the condition assignment variables. Table 10 reports the adjusted R^2 statistic of each model, the predictive value of each self-reported learning process independent of the condition assignment, and the effect of analogical-comparison and self-explanation condition assignment relative to instructional explanation condition assignment, controlling for self-reported processing. The models predicting overall, near-transfer, intermediate-transfer, and PFL-transfer posttest performance were significant, and within those models self-reported self-explanation was positively predictive of performance while analogical comparison was negatively predictive, controlling for condition. Self-reported metacognition was not predictive of performance in any model.

Table 10.

Summary of multiple regression analyses for self-reported learning process use predicting subsets of the posttest.

Test	**Adjusted R^2**	***F***	**Use of explanation**	**Use of comparison**	**Use of metacognition**	**Self-explanation condition**	**Analogical comparison condition**
Total	.21	6.26*	.49*	-.31*	-.070	-.080	-.27*
Near	.15	4.41*	.36*	-.31*	-.005	-0.20†	-.16
Intermediate	.15	4.46*	.46*	-.19†	-.11	.066	-.21†
Far	.019	1.37	.21	-.12	.067	-.084	-.067
PFL	.091	3.00*	.28†	-.24*	-.12	-.056	-.29*

*$p < .05$, † $p < .10$

Finally, I examined the relationship between participants' word use on the open-ended questionnaire and their self-reported ratings on the strategies questionnaire (Table 11). Contrary to predictions, there were no correlations between word use for a particular strategy and endorsement of the strategy on the questionnaire. There was a significant, positive correlation between performance goal word use and self-reported monitoring behaviors, as well as a significant, negative correlation between approach goal word use and self-reported evaluation behaviors.

Table 11.

Correlations between cognitive, metacognitive, and goal word use and self-reported responses on the strategies questionnaire.

	Survey AC	**Survey SE**	**Survey MC**	**Survey MC – Monitor**	**Survey MC - Control**	**MC - Evaluation**
Comparison	.020	.045	.041	.078	.013	-.005
Explanation	-.087	.044	-.048	-.068	-.022	-.019
Examples	.12	.080	.10	.096	.046	.11

Monitoring	.056	.030	-.081	-.057	-.11	-.071
Control	-.16	-.013	-.11	-.065	-.099	-.15
Evaluation	.069	-.032	-.038	-.064	-.075	.032
Mastery	.079	-.053	-.016	-.068	.013	.043
Performance	.13	.14	.18†	.21*	.087	.14
Approach	.081	-.088	-.14	-.083	-.097	-.20*

8.1.5 Task-framed achievement goal surveys

A one-way ANOVA revealed a marginal effect of condition on task-framed mastery-approach goal endorsement, F(2, 98) = 2.86, p = .062, η_p^2 = .055 (Figure 13). Post hoc comparisons using the Tukey HSD test indicated that the mean score for the self-explanation condition (*M* = 5.46, *SD* = 1.17) was marginally different from the instructional explanation condition (*M* = 6.01, *SD* = .87; *p* = .096). However, the analogical comparison condition (*M* = 5.47, *SD* = 1.15) did not significantly differ from the instructional explanation (*p* = .10) or self-explanation conditions (*p* > .99). There was no effect of condition on endorsement of task-framed mastery-avoidance, F(2, 98) = .025, p = .98, η_p^2 = .001, performance-approach, F(2, 98) =2.14, p = .12, η_p^2 = .042, or performance-avoidance goals, F(2, 98) = 1.23, p = .30, η_p^2 = .024.

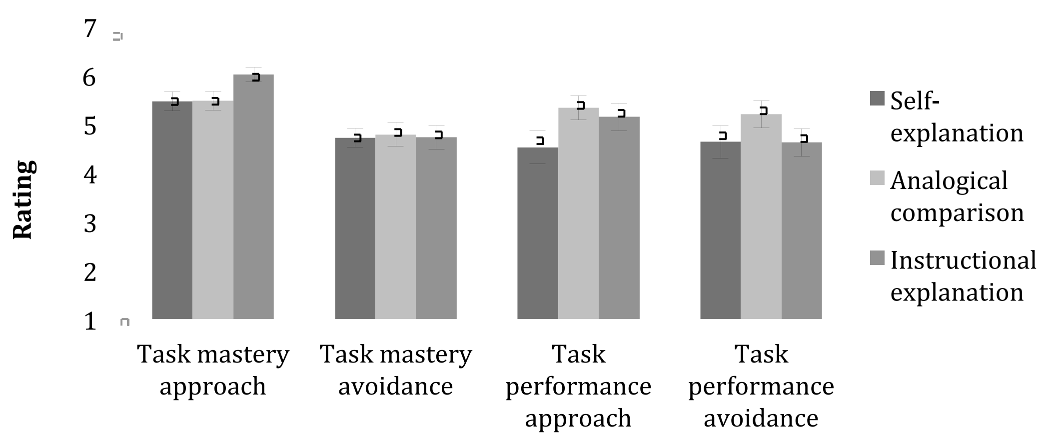


Figure 17. Endorsement of task-framed achievement goals by condition.

To test the amount of variance in posttest performance explained by task-framed achievement goals, variance due to the condition assignment was removed using hierarchical multiple regression. Condition was dummy-coded with the instructional explanation condition as the reference group. Self-reported levels of task-framed mastery approach, mastery avoidance, performance approach, and performance avoidance were entered in a step-wise fashion into the second model with the first model containing the condition assignment variables. Table 12 reports the adjusted R^2 statistic of each model, the predictive value of each task-framed goal independent of the condition assignment, and the effect of analogical-comparison and self-explanation condition assignment relative to instructional explanation condition assignment, controlling for task-framed goals. The models predicting overall, intermediate-transfer, and PFL-transfer posttest performance were significant, and within those models task-framed mastery avoidance was positively predictive of performance while performance approach was sometimes marginally predictive, controlling for condition.

Table 12.

Summary of multiple regression analyses for task-framed achievement goals predicting subsets of the posttest.

Test	**Adjusted R^2**	***F***	**Mastery approach**	**Mastery avoidance**	**Performance approach**	**Performance avoidance**	**Self-explanation condition**	**Analogical comparison condition**
Total	.14	3.61*	.11	.23*	.32†	-.26	-.067	-.24*
Near	.036	1.62	.17	-.082	.18	-.15	-.20	-.12
Inter.	.14	3.60*	.052	.32*	.21	-.19	.071	-.20†
Far	.017	1.28	.24†	-.001	.17	-.23	-.060	-.013
PFL	.13	3.42*	-.15	.32*	.30†	-.15	-.051	-.32*
Misc.	-.030	.52	-.19	.17	.059	.012	-.016	-.088

*$p < .05$, $^\dagger p < .10$

8.1.6 Science class-framed surveys

A one-way ANOVA revealed a marginal effect of condition on science class-framed mastery-approach goal endorsement, $F(2, 98) = 2.75$, $p = .069$. Post hoc comparisons using the Tukey HSD test indicated that the mean score for the instructional explanation condition ($M = 6.10$, $SD = 1.10$) was marginally different from the analogical comparison condition ($M = 5.49$, $SD = 1.29$; $p = .061$). However, the self-explanation condition ($M = 5.90$, $SD = .84$) did not significantly differ from the instructional explanation ($p = .74$) or analogical comparison conditions ($p = .27$). There was no effect of condition on endorsement of task-framed mastery-avoidance, $F(2, 98) = .14$, $p = .87$, performance-approach, $F(2, 98) = .24$, $p = .79$, or performance-avoidance goals, $F(2, 98) = 1.99$, $p = .14$ (Figure 14).

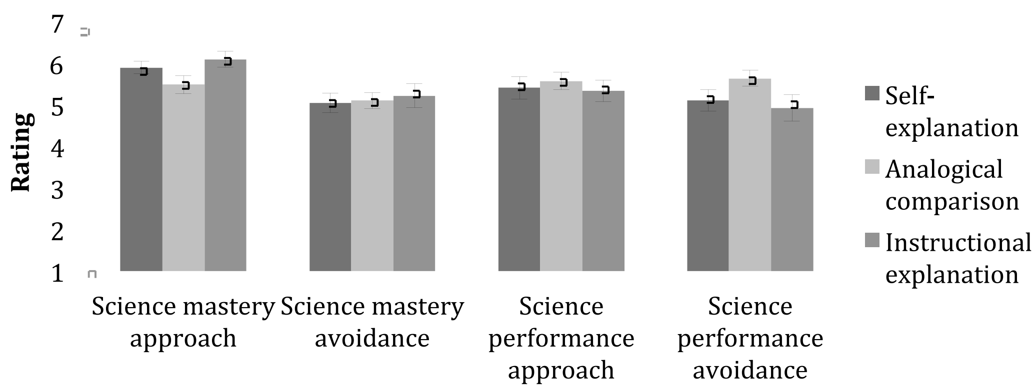


Figure 18. Endorsement of science class-framed achievement goals by condition.

To test the amount of variance in posttest performance explained by science class-framed achievement goals, variance due to the condition assignment was removed using hierarchical multiple regression. Condition was dummy-coded with the instructional explanation condition as the reference group. Self-reported levels of science class-framed mastery approach, mastery avoidance, performance approach, and performance avoidance were entered in a step-wise fashion into the second model with the first model containing the condition assignment variables.

Table 13 reports the adjusted R^2 and F statistics for each model, the predictive value of each science class-framed goal independent of the condition assignment, and the effect of analogical-comparison and self-explanation condition assignment relative to instructional explanation condition assignment, controlling for task-framed goals. The models predicting overall and PFL-transfer posttest performance were significant, and within the model predicting PFL transfer, science class-framed performance approach was marginally predictive of performance. No other science class-framed achievement goals predicted performance on any posttest measure.

Table 13.

Summary of multiple regression analyses for science class-framed achievement goals predicting subsets of the posttest.

Test	**Adjusted R^2**	***F***	**Mastery approach**	**Mastery avoidance**	**Performance approach**	**Performance avoidance**	**Self-explanation condition**	**Analogical comparison condition**
Total	.060	$2.06^{\dagger}$	.080	.15	.14	-.096	-.12	-.25*
Near	.005	1.08	.078	-.036	.064	-.046	0.27*	-.16
Inter.	.044	$2.89^{\dagger}$	.15	.12	-.001	-.008	.040	-.18
Far	.007	1.11	.078	.16	.12	-.22	-.12	-.030
PFL	.094	2.73*	-.12	.15	$.24^{\dagger}$	-.007	-.075	-.32*
Misc.	-.021	.66	-.18	.17	.095	-.053	.015	-.069

$^{*}p < .05$, $^{\dagger}p < .10$

9.0 DISCUSSION

In summary, results showed that instructing students to study worked examples with instructional explanations led to greater posttest performance compared to instructions to compare worked examples. There was no significant relation between learning condition and self-reported levels of self-explanation and analogical comparison, although instructional explanations led to greater self-reported levels of metacognition compared to self-explanation. Finally, participants' self-reports of analogical comparison were significant, negative predictors of performance on near transfer and PFL transfer posttest items and a marginal, negative predictor of performance on intermediate transfer items. Self-reports of self-explanation were significant, positive predictors of near and intermediate transfer performance and a marginal, positive predictor of PFL transfer performance.

One major factor to consider in assessing these results is the average quality of responses according to the prompt response coding. The quality of both the self-explanation and analogical comparison prompt responses averaged below a two on a four-point scale, indicating that they typically fell short of ideal responses that would be likely to entail the types of constructive processes hypothesized to support learning. This illustrates several important points. First, even when the participant sample is undergraduate students who have received instructional text, examples of self-explanations or comparisons, and prompts, they are not especially good at generating meaningful analogical comparison or self-explanation statements. Given this evidence, future work may need to emphasize a more robust training so that participants are better equipped to produce fruitful self-explanations and analogical comparison. Incorporating more scaffolded prompts and immediate feedback on the quality of explanations or comparisons learners produce might improve the effectiveness of the interventions.

Second, these results suggest the need to use caution when interpreting these results. The lack of difference between self-explanation and analogical comparison support conditions could suggest a genuine lack of differences in the knowledge outcomes facilitated by each process. This would be consistent with predictions made by with Chi's active-constructive-interactive framework (Chi, 2009), which suggests that two cognitive processes that fall into the same category should produce similar knowledge outcomes. In this case, both analogical comparison and self-explanation are constructive processes, and thus there should be no major differences in the knowledge outcomes each produces. However, the results show that participants did not generally engage in the types of analogical comparison or self-explanation processes that prior research suggests are most likely to lead to deep learning. Therefore, any conclusions about a true lack of differences in knowledge outcomes would be better supported if a follow-up study that successfully triggered more fruitful self-explanation and analogical comparison also failed to show differences. Furthermore, Chi's (2009) framework would predict that as constructive activities, both self-explanation and analogical comparison would produce greater learning than an active instructional technique like reading instructional explanations. Results indicated that instructional explanations were more effective than either of the constructive processes, suggesting that truly constructive processes did not take place in the self-explanation and analogical comparison conditions.

These results raise several important questions. First, why was there no relation between condition and self-reported levels of self-explanation and analogical comparison? Prior work has shown that self-explanation and analogical comparison are effortful and subject to a great deal of individual variation, even when explicit instructions are given to engage in these processes (e.g., Chi et al., 1994; Gick & Holyoak, 1983). Therefore, it is possible that condition assignment played a smaller role in determining the degree to which individuals engaged in different processes, compared to their spontaneous learning process tendencies. This is supported by evidence showing that self-reported use of analogical comparison and self-explanation predicted performance, suggesting these measures were meaningful. However, the lack of any relation between condition assignment and self-reported explanation and comparison behaviors suggests that some items on the questionnaire may have been misaligned to the task, or that participants had difficulty assessing the degree to which they engaged in the processes. Many students have poor awareness of their own cognitive process use and may have struggled to report what they actually did during the learning phase (Metcalfe, Eich, & Castel, 2010). Evidence from self-explanation literature is consistent with this general finding. For example, a study that compared providing learners with instructional explanations and prompting them to construct explanations found that the self-explanation prompts supported greater learning, but participants felt they were learning best from provided the instructional explanations (Schworm & Renkl, 2006). Consequently, it may be that learners' responses to the learning process questionnaires reflected something related to their performance (e.g., their desire to appear engaged with or effortful in their use of the learning materials) but not what they actually did. Prior research also shows that not all self-explanations or analogical comparisons lead to robust knowledge. The items focus on frequency of use but may not capture variability in the *quality* of self-explanations or analogical comparisons. In other words, participants who reported engaging in greater levels of comparison or self-explanation may not necessarily have engaged in more comparison of structural features or self-explanation of principles. Finally, participants might interpret the scale differently, such that one participant could have a different standard for what constitutes engaging in a lot of self-explanation. If participants' actual behaviors systematically affected their interpretations of the scales, such that those who put more effort into following the prompts also developed higher criteria for describing their efforts as "a lot," it could create a confounding relationship between participants' learning processes and their rating judgments.

Second, why did analogical comparison lead to worse performance, regardless of whether it was assigned (through condition) or spontaneous (as reflected in self-reported levels)? Some prior work has found similar results on certain types of tasks. Nokes-Malach et al. (2013) found that analogical comparison of physics problems led to worse near-transfer problem-solving performance compared to self-explanation and instructional explanation conditions, although the conditions were equivalent on intermediate-transfer items and the analogical comparison and self-explanation conditions outperformed the instructional explanation condition on far-transfer items. In the context of category learning, Edwards et al. (2013) found that instructions to compare exemplars in a group were less effective than instructions to explain because the comparison prompts constrained the types of comparison learners engaged in. More broadly, prior work has found that adding scaffolding to analogical comparison activities leads to greater learning if the scaffolding identifies key features, as learners may struggle to align structural features without guidance (Gentner et al., 2003). Thus, one possible explanation for the negative relation between analogical comparison and performance in our work could be that neither the experimental manipulation to support analogical comparison nor the learners' spontaneous comparisons consistently targeted the structural relations that, when aligned, support more fruitful comparison. Further analysis of participants' prompt responses could provide a more fine-grained assessment of how well participants' responses aligned with the types of analogical comparison statements thought to support learning of deep concepts and abstract principles.

Edwards et al. (2013) also found that participants instructed to engage in explanation reported greater levels of explanation *and* comparison when asked to rate their behaviors on a single-item scale of 1 to 7. Although these results differ from our findings that neither condition reported greater levels of explanation or comparison, they are similar in showing a lack of alignment between participants' self-reported levels of each process and the processes the experimental manipulations were intended to support. These results suggest that instructions to compare or explain likely alter learners' behaviors in a broader range of ways and encourage changes (or perceived changes) in multiple cognitive processes. It may also be that self-explanation and analogical comparison prompts led to more variation in what learners did while studying the worked examples. If learners in the instructional explanation condition more consistently attended to the information in the examples, they might have better learned the basic content. Finally, while materials were designed to suppress spontaneous engagement in analogical comparison or self-explanation outside of the targeted conditions, it is possible that students still engaged in analogical comparison across pages or elaborated on provided instructional explanations.

Third, why were prompt response scores not related to intermediate or far transfer for the self-explanation condition or any levels of transfer for the analogical comparison condition? This lack of correlation could be explained in at least two ways. It could be that engaging in the cognitive processes supported by self-explanation or analogical comparison prompts, as coded by higher rubric scores, did not lead to greater learning. While this would contradict much prior work suggesting that self-explanation and analogical comparison support deep, conceptual learning, there is also evidence that content being explained or compared and the type of knowledge tested affect whether engaging in high-quality self-explanation or analogical comparison leads to greater performance.

Alternatively, it may be that the rubric scoring levels were not optimally aligned with the key behaviors and cognitive processes responsible for deep learning through self-explanation and analogical comparison. For the self-explanation rubric, generating content that was not already stated in the worked examples was the primary difference between low scores and high scores on the rubric; however, this approach may have missed another important factor of how closely the participants' responses approached the key principle or principles being communicated in the example. That is, some responses receiving a higher score might have included novel content that was not relevant to underlying principles, while some lower-scoring responses might have stated the key principle clearly but not introduced any novel ideas. The worked examples were designed to explain key features of the problem, and thus writing a response that included novel content might have indicated in many cases a focus on *less* important features of the problem.

In the case of the analogical comparison condition, the coding rubric assigned lower scores to responses identifying similarities and/or differences without discussing their significance and higher scores that included the significance or concepts highlighted by similarities and/or differences. Again, however, this distinction might have missed a factor of relevance, such that a critical similarity or difference identified without discussion of significance would receive a lower score than an irrelevant similarity or difference identified with discussion of its significance. The rubric also gave a higher score if participants identified both a similarity and difference instead of only a similarity or a difference. While there were relevant similarities and differences to discuss in each example, there may have been more value in identifying a single, critical similarity or difference than in identifying less critical similarities and differences.

For both self-explanation and analogical comparison, revised rubrics that captured how well participants' responses identified the key principle or principles being illustrated in each worked example might produce prompt response scores that are more predictive of posttest performance. Additionally, some prior work examining self-explanation has focused on classify the content of self-explanations rather than scoring its quality, and results from this work suggest that certain kinds of content are more effective for promoting the type of conceptual learning that facilitates transfer (Ainsworth & Burcham, 2007; Berthold & Renkl, 2009; Hausmann et al., 2009; Renkl, 1997). An alternative approach of classifying the types of self-explanations and comparisons thus might be another fruitful path for better understanding the relationship between how participants responded to prompts and how they performed across different levels of transfer on the posttest.

Results relating to achievement goals were more consistent with hypotheses. While mastery-avoidance goals are often considered to be negative based on their avoidance valence, they also entail an emphasis on mastery, or understanding the material relative to what is possible or relative to one's own prior understanding. In prior work using very similar materials, I found that mastery-avoidance goals were associated with greater posttest performance (Richey & Nokes-Malach, 2013). These results were replicated in the present work, with task-framed mastery-avoidance goals being correlated with greater intermediate and PFL transfer. Since participants in both experiments were college students who mostly had previously taken a physics course in high school, it is possible that a mastery-avoidance goal directed at avoiding the loss of prior understanding can be productive when re-learning material. While it was unexpected that instructional explanations would lead to greater endorsement of mastery-avoidance goals, this may offer further evidence that the self-explanation and analogical comparison prompts failed to encourage the kinds of deep, constructive learning behaviors associated with mastery goals. It may be that reading thorough explanations led to greater mastery because the content emphasized deeper understanding, even without thee constructive element of the other instructional techniques.

10.0 CONCLUSION

Across both experiments, few posttest differences emerged between self-explanation and analogical comparison conditions. Difference in posttest performance generally involved the instructional explanation condition performing better than other conditions, including the practice condition (near transfer, Experiment 1) and the analogical comparison condition (overall accuracy and PFL transfer, Experiment 2). The only difference between self-explanation and analogical comparison conditions was a marginal difference on intermediate-transfer items in Experiment 2, on which the self-explanation condition performed better than the analogical comparison condition. Thus, results suggest a weak effect such that studying worked examples with instructional explanations may have increased learning relative to comparing worked examples.

While theories of instructional explanation and self-explanation generally suggest that self-explanation will support deeper learning through its constructive processes, there are several caveats. First, self-explanations vary in quality even when they are prompted, and only certain types of self-explanations may actually support learning (Ainsworth & Burcham, 2007; Berthold & Renkl, 2009). While self-explanations do not necessarily need to be complete or accurate to promote learning, much evidence suggests they must at least be effortful and involve the activation of prior knowledge. Although the coding of prompt responses revealed a range from the lowest to greatest number of points possible, even responses earning the largest number of points possible may have lacked the effortful focus on constructing knowledge and identifying principles that supports learning from self-explanation.

It may be that effort is more important for learning from self-explanation than from instructional explanations; if that is the case and participants across all conditions employed relatively little effort, it could explain why instructional explanations led to more learning. The learners may have felt they understood the worked examples sufficiently without engaging in effortful self-explanation or analogical comparison. Additionally, some content in the instructional text and on the test was not addressed in the worked examples, and participants may have skimmed over this content without realizing that it was also important. These possibilities are supported anecdotally in participants' open-ended responses, in which many discussed assessing their understanding by whether they were able to solve the practice problems. Most participants expressed confidence that they understood the problems, and their performance on the practice problems is consistently high; however, many of those participants nevertheless struggled with the posttest, which targeted a wider variety of concepts covered in the learning materials as well as the worked examples. Future work should include more challenging conceptual examples that will produce more failure, thus prompting students to engage more deeply in responding to prompts and studying instructional text to understand the worked examples.

Although most participants reported taking a physics course in the past, very few had completed a college course. Even with the instructional text provided before each set of worked examples, it might be that participants lacked sufficient prior knowledge to make fruitful explanations or to engage in revising their prior knowledge. There is less evidence that prior knowledge is important for engaging in fruitful analogical comparison (Alfieri et al., 2013), although a learner who lacks sufficient prior knowledge to understand key features of examples might be unable to see the structural alignment across cases. In both cases, lacking sufficient prior knowledge could have made receiving instructional explanations more effective than trying to construct explanations or comparisons.

An alternative explanation focuses not on the failures of the experiment to promote effective self-explanations, but rather on the strength of the instructional explanations condition. Instructional explanations have been shown to improve learning from worked examples under some conditions (Wittwer & Renkl, 2010), and their primary hypothesized advantage over self-explanation is that they are complete and accurate. The instructional explanations provided across both experiments focused on key principles targeted in the practice and test problems. Few participants were able to construct information in the self-explanation or analogical comparison conditions that described key principles as completely or accurately as what was provided in the instructional explanations condition. Consequently, the benefits of receiving complete, accurate instructional explanations might have outweighed the benefits of being prompted to construct individualized knowledge through self-explanation or analogical comparison prompts. Instructional explanations may be particularly useful when participants lack either the prior knowledge or effort necessary to construct explanations or comparisons focused on important principles.

Self-reports of learning processes generally did not align with condition assignments. It is possible that this means there was more individual variability in the processes learners engaged in compared to variability between conditions. Further investigation of the additional strategies participants reported might show other patterns in individual differences outside the processes targeted by the learning prompts. It is possible that engagement these other processes might vary by condition, or that they would predict performance on the test. Alternatively, this result may reflect learners' generally poor metacognitive skills when it comes to assessing their own learning processes. Despite the lack of alignment between condition assignment and self-reported strategies, several self-reported strategies were associated with learning outcomes in expected ways, including metacognitive behaviors and self-explanation.

Future work should continue to investigate how analogical comparison and self-explanation operate individually and through interactions to promote different types of learning. Questionnaires capturing specific types and sub-processes of analogical comparison and self-explanation might improve understanding of how each facilitates different types of learning. By improving understanding of similarities and differences between analogical comparison and self-explanation, I hope to identify concrete recommendations for when and how instructors can support each process based on their instructional goals.

APPENDIX A

INSTRUCTIONS AND PROMPTS BY CONDITION

A.1 EXPERIMENT 1

A.1.1 Self-explanation

a. Instructions

Now we will move on to some more examples of sequence extrapolation problems. Your task is to *find* the pattern and *continue* it so the subsequent letters follow the same pattern. You will read through some examples of sequence extrapolation problems that have been solved for you. As you read through each example, you will be prompted to explain each step. Please think carefully about each step and write your explanation in the space provided before moving on to the next step. Studying these examples carefully will help you learn how to solve this type of problem.

b. Prompts

What does this step mean to you? Could you restate or summarize the step in your own words? What else does this step tell you?

A.1.2 Analogical comparison

a. Instructions

Now we will move on to some more examples of sequence extrapolation problems. Your task is to *find* the pattern and *continue* it so the subsequent letters follow the same pattern. You will read through some examples of sequence extrapolation problems that have been solved for you. You will see steps for two separate examples at once, and you'll be prompted to compare across the two steps. Please think carefully about each comparison and write your response in the space provided before moving on to the next pair of steps. Comparing these examples carefully will help you learn how to solve this type of problem.

b. Prompts

What is similar about this step of the two problems? What is different? What do the similarities and differences tell you?

A.1.3 Instructional explanation

a. Instructions

Now we will move on to some more examples of sequence extrapolation problems. Your task is to **find** the pattern and **continue** it so the subsequent letters follow the same pattern. You will read through some examples of sequence extrapolation problems that have been solved for you. Please think carefully about each step. Studying these examples carefully will help you learn how to solve this type of problem.

b. Instructional explanation

Example: This step identifies the period of the sequence. Each period is made up of the same set of rules. Some letters within a period are determined by their relations to letters in the previous period. After identifying the length of the pattern, you can look across periods to identify the rules that determine letters in the following period to continue the pattern. You can also focus on just one complete pattern sequence, from the start to the end of the period, to identify the rules within the period. This period is six letters long.

A.1.4 Practice

a. Instructions

Now we will move on to some more examples of sequence extrapolation problems. Your task is to ***find*** the pattern and ***continue*** it so the subsequent letters follow the same pattern. You will read a series of sequence extrapolation and try to solve them one at a time. Please think carefully about each problem. Trying to solve these problems will help you learn how to solve this type of problem. After you have gotten an answer or solved as much as you can, you can turn the page and see the solution to the problem. You can study the solution for as long as you need, but please do NOT turn to the solution before you have finished solving the problem.

b. Prompt

Try your best to solve this problem. When you have reached the answer or completed as much as you can, proceed to the solution on the next page.

A.2 EXPERIMENT 2

A.2.1 Self-explanations

a. Instructions

You will read through a text about electricity and solve some problems. You will also read through some examples of problems that have been solved for you. As you read through each example, think carefully about each step. You will be prompted to explain the examples after you have read each one and write your explanation in the space provided.

This process employs a learning strategy called **self-explanation.** Self-explanation involves reading a text or example and explaining to yourself what the text means. Past research has shown that self-explanation is a powerful tool for learning, but it works best if you take your time and explain things as thoroughly as you can.

There are many types of self-explanation that can be useful, including **monitoring** your comprehension**,** making **bridging** inferences to link separate ideas in the text or examples, and **elaborating** by using prior knowledge and logic to understand the text. Different types may be more appropriate at different times depending on what the content is and how it relates to what you already know and what your goals are.

In the following pages, try to apply these different tools to self-explain as you read the text and study worked examples. Take notes in the margins of the text when it is helpful. You'll be prompted to write out your self-explanations when you encounter worked examples. For the first few examples, you will receive specific instructions to help you get started.

b. Modeling

Here are some examples of self-explanations that you might generate in response to the worked example:

Monitoring: Thinking about your comprehension and identifying what you don't understand

Example: "This problem seems to depend on understanding what 'uncharged' means. I'm not sure I remember reading that term. What does "uncharged' mean?"

Bridging: Linking to other ideas in the text

Example: "So 'uncharged' here might mean the same thing as 'electrically neutral,' which is when the number of electrons and protons are equal. This means uncharged substances have the same number of protons and electrons, and when they lose or gain electrons they become charged."

Elaboration: Making connections between the text and things you already know (e.g., creating examples based on prior knowledge), or making logical inferences to understand the text

Example: "The text said that hair and a balloon are attracted to each other after electrons transfer from the hair to the balloon. In this case, electrons transfer from the rod to the silk, so they should be attracted to each other, too."

c. Prompts

Self-explain the reasoning or justification for this solution. Write out words to describe any symbols, and provide conceptual justifications and principled reasoning to explain the solution.

A.2.2 Analogical comparison

a. Instructions

You will read through a text about electricity and solve some problems. You will also read through some examples of problems that have been solved for you. As you read through each example, think carefully about each step. You will be prompted to compare the examples after you have read each one and write your comparison in the space provided.

This process employs a learning strategy called **analogical comparison.** Analogical comparison involves aligning features of different pieces of information or examples and mapping the similarities between them. Past research has shown that analogical comparison is a powerful tool for learning, but it works best if you take your time and make productive comparisons.

As you make comparisons, try to focus on features that are critical to the concepts or solution principle. Focusing on such features helps learners to identify and understand critical information. In contrast, focusing on features that are not critical to the concepts being used or the problem-solving procedures may cause you to remember unimportant or superficial details instead of important, structural information.

It is often useful to notice important differences as well as similarities between cases. As you read the text and study the examples, try to identify key features to compare ideas in the text and worked examples. Take notes in the margins of the text when it is helpful. You'll be prompted to write out your analogical comparisons when you encounter worked examples. For the first few examples, you will receive specific instructions to help you get started.

b. Modeling

Remember, analogical comparison involves aligning features of different pieces of information, and it is most useful if you focus on features that are critical to the concepts or solution principle. Here is an example of a comparison of critical information.

Similarities: "Both worked examples deal with the idea of electron transfer. In both cases, one substance is losing electrons and the other is gaining electrons. In both, the substance that loses electrons becomes positively charged and the substance that gains electrons becomes negatively charged."

It is also good to abstract the details of the problems and think about the general principle demonstrated across problems.

Abstraction: "These problems illustrate that electrons can move within or between substances, and the creation of charged objects results from this electron movement."

Try NOT to focus on superficial similarities and differences, such as the detail that one example deals with silk and glass and the other example deals with a balloon and sweater. These features are not critical to the concepts or solution principles being applied.

Can you identify any other similarities that are relevant to the concepts applied in these problems? Write them here:

c. Prompts

What is similar across problems? What is different? What do the similarities and differences tell you about the concepts involved?

A.2.3 Instructional explanation

a. Instructions

You will read through a text about electricity and solve some problems. You will also read through some examples of problems that have been solved for you. As you read through each example, think carefully about each step. You will also receive conceptual justifications for each step.

This process employs a learning strategy called **worked example** study. Worked examples demonstrate correct problem-solving steps to help you solve similar problems. While simply seeing a solution to a problem can be helpful, it is often more beneficial for learners to see the procedures and rules applied at each step to solve a problem, so they learn not only what the correct answer is but also how to get it. Past research has shown that worked examples are a powerful tool for learning, but they work best if you take your time and study each example as thoroughly as you can.

In the following pages, read the text and take your time to study each worked example. Take notes in the margins of the text when it is helpful.

b. Instructional explanations

Example:

General principle applied: The relationship between protons and electrons determines the charge of the atom.

Define values and relations: A material usually cannot gain or lose protons, but it can gain or lose electrons. The number of protons AND the number of electrons must be known to determine charge.

Solve based on values and principle: If there are more protons than electrons, the material is positively charged. If there are more electrons than protons, the material is negatively charged.

c. Prompts

Remember to take your time and study each worked example carefully. Studying worked examples can improve your learning of the concepts and solutions.

APPENDIX B

DICTIONARIES CREATED FOR LIWC ANALYSIS

Table 14.

LIWC dictionary words for cognitive processes.

Compare	Explain	Example study
abstract*	describe*	answer*
across	elaborate*	problem*
align*	example*	procedure*
alike	explain*	read
analog*	explan*	re-read
between	infer*	solution*
common	justif*	step*
compar*	myself	studi*
differ*	past	Study
map*	previous*	
oppos*	prior	
relate*	reason*	
relating	restat*	
same	why	
share*	self	
similar*		

Table 15.

LIWC dictionary words for metacognitive processes.

Monitor	**Control**	**Evaluate**
knew	adjust*	answer*
know*	change*	evaluat*
monitor*	changing	outcome*
notice*	control*	response
noticing	plan	review*
progress	plann*	sense
recogni*	plans	solution*
test	rethink*	
tested	strateg*	
testing		
tests		
think*		
thought*		
understand*		
understood		
unknown*		
unsure*		

Table 16.

LIWC dictionary words for achievement goals.

Mastery	**Performance**	**Approach**
capab*	appear	accomplish*
competen*	correct	acquir*
ideal*	demonstrate	achiev*
knew	error	attain*
know*	incorrect	beat
knowledg*	grade	best
learn*	look	better
master*	mistake*	excel*
past	others	gain*
perfect*	perform*	good
possible	right	improve*
potential*	score	improving
previous*	show	succeed
prior	wrong	success
self		top

understand*	well
	win
	winn*
	wins
	won

BIBLIOGRAPHY

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

Ainsworth, S., & Loizou, A. T. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science*, *27*, 669–681. doi:10.1016/S0364-0213(03)00033-8

Aleven, V. A. W. M. M., & Koedinger, K. R. (2000). The need for tutorial dialog to support self-explanation. In C. P. Rose & R. Freedman (Eds.), *Building Dialogue Systems for Tutorial Applications, Papers of the 2000 AAAI Fall Symposium* (pp. 65–73). Menlo Park, CA: AAAI Press.

Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, *26*(2), 147–179. doi:10.1016/S0364-0213(02)00061-7

Aleven, V. A. W. M. M., Koedinger, K. R., & Popescu, O. (2003). A tutorial dialog system to support self-explanation: evaluation and open questions. In U. Hoppe, F. Verdejo, & J. Kay (Eds.), *Proceedings of the 11th International Conference on Artificial Intelligence in Education* (Vol. 2003, pp. 39–46). Amsterdam: IOS Press.

Aleven, V. A. W. M. M., Popescu, O., & Koedinger, K. R. (2002). Pilot-testing a tutorial dialogue system that supports self-explanation. In S. A. Cerri, G. Gouardères, & F. Paraguaçu (Eds.), *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems, ITS 2002* (pp. 344–354). Berlin: Springer Verlag.

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: a meta-analytic review. *Educational Psychologist*, *48*(2), 87–113. doi:10.1080/00461520.2013.775712

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: students' learning strategies and motivation processes. *Journal of Educational Psychology*, *80*(3), 260–267. doi:10.1037/0022-0663.80.3.260

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, *94*(2), 416–427. doi:10.1037/0022-0663.94.2.416

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. doi:10.1037/0022-0663.95.4.774

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612–637. doi:10.1037//0033-2909.128.4.612

Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology*. doi:10.1037/a0013247

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*, *13*(2), 221–252. doi:10.1207/s1532690xci1302_3

Blanchette, I., & Dunbar, K. N. (2000). How analogies are generated: The roles of structural and superficial similarity. *Memory & Cognition*, *28*(1), 108–124. doi:10.3758/BF03211580

Boekaerts, M., & Corno, L. (2005). Self-Regulation in the Classroom: A Perspective on Assessment and Intervention. *Applied Psychology*, *54*(2), 199–231. doi:10.1111/j.1464-0597.2005.00205.x

Brand, S., Reimer, T., & Opwis, K. (2003). Effects of metacognitive thinking and knowledge acquisition in dyads on individual problem solving and transfer performance. *Swiss Journal of Psychology*, *62*(4), 251–261. doi:10.1024/1421-0185.62.4.251

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: a simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of Research in Education* (Vol. 24., pp. 61–100). Washington, D.C.: American Educational Research Association.

Brown, D. E., & Clement, J. (1989). Overcoming misconceptions via analogical reasoning: abstract transfer versus explanatory model construction. *Instructional Science*, *18*(4), 237–261. doi:10.1007/BF00118013

Burns, B. D., & Vollmeyer, R. (2002). Goal specificity effects on hypothesis testing in problem solving. *The Quarterly Journal of Experimental Psychology.*, *55A*(1), 241–61. doi:10.1080/02724980143000262

Catrambone, R. (1998). The subgoal learning model: creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General*, *127*(4), 355–376. doi:10.1037/0096-3445.127.4.355

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293–332. doi:10.1207/s1532690xci0804

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* (Vol. 5, pp. 161–238). Lawrence Erlbaum Associates.

Chi, M. T. H. (2009). Active-constructive-interactive: a conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*, 73–105. doi:10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, *13*(2), 145–182. doi:10.1207/s15516709cog1302_1

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting Self-Explanations Improves Understanding. *Cognitive Science*, *18*(3), 439–477. doi:10.1207/s15516709cog1803_3

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates. doi:10.1234/12345678

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: a new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, and Computers*, *25*(2), 257–271. doi:10.3758/BF03204507

Collins, J., & Osborne, S. (2001). Pupils' views of the role and value of the science curriculum: a focus-group study. *International Journal of Science Education*, *23*(5), 441–467. doi:10.1080/09500690119975

Conati, C., & VanLehn, K. A. (2000). Toward computer-based support of meta-cognitive skills: a computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education*, *11*, 389–415.

Crowley, K., & Siegler, R. S. (1999). Explanation and generalization in young children's strategy learning. *Child Development*, *70*(2), 304–316. doi:10.1111/1467-8624.00023

Dellarosa, D. (1985). *Abstraction of problem-type schemata through problem comparison (Tech. Rep. No. 146)*. Boulder.

Edwards, B. J. (2014). *Explanation and Comparison Interact to Support Learning*. Northwestern University.

Edwards, B. J., Williams, J. J., Gentner, D., & Lombrozo, T. (2014). Effects of comparison and explanation on analogical transfer. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 445–450). Austin, TX: Cognitive Science Society.

Edwards, B. J., Williams, J. J., & Lombrozo, T. (2013). Effects of explanation and comparison on category learning. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 406–411). Austin, TX: Cognitive Science Society.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: the MASRL model. *Educational Psychologist*, *46*(1), 6–25. doi:10.1080/00461520.2011.538645

Elliot, A. J., & McGregor, H. A. (2001). A 2 X 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*(3), 501–519. doi:10.1037//0022-3514.80.3.501

Elliot, A. J., McGregor, H. A., & Gable, S. (1999). Achievement goals, study strategies, and exam performance: A mediational analysis. *Journal of Educational Psychology*, *91*(3), 549–563.

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: critique, illustration, and application. *Journal of Educational Psychology*, *100*(3), 613–628. doi:10.1037/0022-0663.100.3.613

Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General*, *120*(1), 34–45. doi:10.1037/0096-3445.120.1.34

Fonseca, B., & Chi, M. T. H. (2011). The self-explanation effect: a constructive learning activity. In R. Mayer & P. Alexander (Eds.), *The Handbook of Research on Learning and Instruction* (pp. 296–321). NY: Routledge Taylor and Frances Group.

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. doi:10.1016/j.learninstruc.2011.06.002

Gentner, D. (1983). Structure-mapping: a theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170. doi:10.1016/S0364-0213(83)80009-3

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: a general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393–405. doi:10.1037/0022-0663.95.2.393

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*(2-3), 263–297. doi:10.1207/s15516709cog1003_2

Gentner, D., & Namy, L. L. (1999). Comparison in the Development of Categories. *Cognitive Development*, *14*(4), 487–513. doi:10.1016/S0885-2014(99)00016-7

Gerjets, P., Scheiter, K., & Schuh, J. (2008). Information comparisons in example-based hypermedia environments: supporting learners with processing prompts and an interactive comparison tool. *Educational Technology Research and Development*, *56*(1), 73–92. doi:10.1007/s11423-007-9068-z

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38. doi:10.1016/0010-0285(83)90002-6

Gick, M. L., & Paterson, K. (1992). Do contrasting examples facilitate schema acquisition and analogical transfer? *Canadian Journal of Psychology*, *46*(4), 539–550. doi:10.1037/h0084333

Graham, S. A., Namy, L. L., Gentner, D., & Meagher, K. (2010). The role of comparison in preschoolers' novel object categorization. *Journal of Experimental Child Psychology*, *107*(3), 280–290. doi:10.1016/j.jecp.2010.04.017

Hausmann, R. G. M., & Chi, M. T. H. (2002). Can a computer interface support self-explaining? *Cognitive Technology*, *7*(1), 4–14.

Hausmann, R. G. M., Nokes, T. J., VanLehn, K. A., & van de Sande, B. (2009). Collaborative dialog while studying worked-out examples. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence in Education* (pp. 596–598). IOS Press. doi:10.3233/978-1-60750-028-5-596

Hausmann, R. G. M., van de Sande, B., & VanLehn, K. A. (2008). Shall we explain? Augmenting learning from intelligent tutoring systems and peer collaboration. In B. P. Woolf, E. Aïmeur, R. Nkambou, & S. Lajoie (Eds.), *Intelligent Tutoring Systems* (Vol. 5091, pp. 636–645). Berlin: Springer-Verlag. doi:10.1007/978-3-540-69132-7

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

Hausmann, R. G. M., & VanLehn, K. A. (2007). Explaining self-explaining: a contrast between content and generation. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *13th International Conference on Artificial Intelligence in Education* (pp. 417–424). Amsterdam: IOS Press.

Haworth, C. M. A., Dale, P., & Plomin, R. (2008). A twin study into the genetic and environmental influences on academic performance in science in nine-year-old boys and girls. *International Journal of Science Education*, *30*(8), 1003–1025. doi:10.1080/09500690701324190

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, *15*(4), 332–340. doi:10.3758/BF03197035

Howell, A. J., & Watson, D. C. (2007). Procrastination: Associations with achievement goal orientation and learning strategies. *Personality and Individual Differences*, *43*(1), 167–178. doi:10.1016/j.paid.2006.11.017

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology*, *102*(4), 880–895. doi:10.1037/a0019506

Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science (New York, N.Y.)*, *326*(5958), 1410–2. doi:10.1126/science.1177067

Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin*, *136*(3), 422–49. doi:10.1037/a0018947

Jarvis, T., & Pell, T. (2001). Developing attitude to science scales for use with children of ages from five to eleven years. *International Journal of Science Education*, *23*(8), 847–862. doi:10.1080/09500690117652

Jenkins, E. W., & Nelson, N. W. (2005). Important but not for me: students' attitudes towards secondary school science in England. *Research in Science & Technological Education*. doi:10.1080/02635140500068435

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, *57*, 227–54. doi:10.1146/annurev.psych.57.102904.190100

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th editio.). Upper Saddle River, NJ: Pearson Education International.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, *342*(22), 935–937. doi:10.1126/science.1238056

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, *36*(5), 757–98. doi:10.1111/j.1551-6709.2012.01245.x

Kolodner, J. L. (1997). Educational implications of analogy: a view from case-based reasoning. *American Psychologist*, *52*(1), 57–66. doi:10.1037/0003-066X.52.1.57

Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences*, *10*(4), 417–446. doi:10.1207/S15327809JLS1004new_2

Leinhardt, G. (2010). Introduction: explaining instructional explanations. In M. K. Stein & L. Kucan (Eds.), *Instructional explanations in the disciplines* (pp. 1–5). Boston, MA: Springer US. doi:10.1007/978-1-4419-0594-9

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. doi:10.1002/(SICI)1098-2736(199909)36:7<837::AID-TEA6>3.0.CO;2-U

Linnenbrink-Garcia, L., Tyson, D. F., & Patall, E. A. (2008). When are achievement goal orientations beneficial for academic achievement? A closer look at main effects and moderating factors. *International Review of Social Psychology*, *21*(1-2), 19–70.

Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin Review*, *6*(4), 586–597. doi:10.3758/BF03212967

Lyons, T. (2006). Different countries, same science classes: students' experiences of school science in their own words. *International Journal of Science Education*, *28*(6), 591–613. doi:10.1080/09500690500339621

MacDonald, P. L., & Gardner, R. C. (2000). Type I Error Rate Comparisons of Post Hoc Procedures for I j Chi-Square Tables. *Educational and Psychological Measurement*, *60*(5), 735–754. doi:10.1177/00131640021970871

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*(4), 431–467. doi:10.1006/cogp.1993.1011

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

Middleton, M. J., & Midgley, C. (1997). Avoiding the demonstration of lack of ability: An underexplored aspect of goal theory. *Journal of Educational Psychology*, *89*(4), 710–718. doi:10.1037/0022-0663.89.4.710

Muñoz, B., Magliano, J. P., Sheridan, R., & McNamara, D. S. (2006). Typing versus thinking aloud when reading: implications for computer-based assessment and training tools. *Behavior Research Methods*, *38*(2), 211–217. doi:10.3758/BF03192771

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, *131*(1), 5–15. doi:10.1037/0096-3445.131.1.5

Neuman, Y., & Schwarz, B. (1998). Is self-explanation while solving problems helpful? The case of analogical problem-solving. *British Journal of Educational Psychology*, *68*, 15–24. doi:10.1111/j.2044-8279.1998.tb01271.x

Nokes, T. J. (2009). Mechanisms of knowledge transfer. *Thinking & Reasoning*, *15*(1), 1–36. doi:10.1080/13546780802490186

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

Nokes, T. J., & Ohlsson, S. (2005). Comparing multiple paths to mastery: what is learned? *Cognitive Science*, *29*(5), 769–96. doi:10.1207/s15516709cog0000_32

Nokes-Malach, T. J., VanLehn, K. A., Belenky, D. M., Lichtenstein, M., & Cox, G. (2013). Coordinating principles and examples through analogy and self-explanation. *European Journal of Psychology of Education*, *28*(4), 1237–1263. doi:10.1007/s10212-012-0164-z

Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(3), 398–415. doi:10.1037/0278-7393.17.3.398

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: applications, interpretations, and limitations. *Contemporary Educational Psychology*, *25*(3), 241–286. doi:10.1006/ceps.2000.1040

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434. doi:10.1037/0022-0663.84.4.429

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of Educational Psychology*, *86*(1), 122–133. doi:10.1037/0022-0663.86.1.122

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal Teaching of Comprehension-Fostering and Comprehension-Monitoring Activities. *Cognition and Instruction*, *1*(2), 117–175. doi:10.1207/s1532690xci0102_1

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic Inquiry and Word Count: LIWC 2007. Austin, TX: LIWC. Retrieved from www.liwc.net

Pintrich, P. R. (2003). A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Contexts. *Journal of Educational Psychology*, *95*(4), 667–686. doi:10.1037/0022-0663.95.4.667

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. doi:10.1037//0022-0663.82.1.33

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond Cold Conceptual Change: The Role of Motivational Beliefs and Classroom Contextual Factors in the Process of Conceptual Change. *Review of Educational Research*, *63*(2), 167–199. doi:10.3102/00346543063002167

Renkl, A. (1997). Learning from worked-out examples: a study on individual differences. *Cognitive Science*, *21*(1), 1–29. doi:10.1207/s15516709cog2101_1

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

Renkl, A. (2005). The worked-out-example principle in multimedia learning. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 229–246). New York: Cambridge University Press.

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, *38*(1), 1–37. doi:10.1111/cogs.12086

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: the effects of example variability and elicited self-explanations. *Contemporary Educational Psychology*, *23*(1), 90–108. doi:10.1006/ceps.1997.0959

Richey, J. E., & Nokes-Malach, T. J. (2013). How much is too much? Learning and motivation effects of adding instructional explanations to worked examples. *Learning and Instruction*, *25*, 104–124. doi:10.1016/j.learninstruc.2012.11.006

Richey, J. E., & Nokes-Malach, T. J. (2015). Comparing four instructional techniques for promoting robust knowledge. *Educational Psychology Review*, *27*(1), 181–218. doi:10.1007/s10648-014-9268-0

Richland, L. E., Holyoak, K. J., & Stigler, J. W. (2004). Analogy use in eighth-grade mathematics classrooms. *Cognition and Instruction*, *22*(1), 37–60. doi:10.1207/s1532690Xci2201

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

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology*, *99*(3), 561–574. doi:10.1037/0022-0663.99.3.561

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology*, *101*(4), 836–852. doi:10.1037/a0016026

Ross, B. H., & Kilbane, M. C. (1997). Effects of principle explanation and superficial similarity on analogical mapping in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 427–440. doi:10.1037//0278-7393.23.2.427

Roussel, P., Elliot, A. J., & Feltman, R. (2011). The influence of achievement goals and social goals on help-seeking from peers in an academic context. *Learning and Instruction*, *21*(3), 394–402. doi:10.1016/j.learninstruc.2010.05.003

Schraw, G., Dunkle, M. E., Bendixen, L. D., & Roedel, T. D. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology*, *87*(3), 433–444. doi:10.1037/0022-0663.87.3.433

Schwartz, D. L., & Bransford, J. D. (1998). A time For telling. *Cognition and Instruction*, *16*(4), 475–5223. doi:10.1207/s1532690xci1604_4

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1–51). CT: Information Age Publishing.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: the hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, *22*(2), 129–184. doi:10.1207/s1532690xci2202_1

Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: when instructional explanations reduce self-explanations. *Computers & Education*, *46*(4), 426–445. doi: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. doi:10.1037/0022-0663.99.2.285

Simon, H. A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review*, *79*(5), 369–382. doi:10.1037/h0033118

Slotta, J. D., & Chi, M. T. H. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction*, *24*(2), 261–289. doi:10.1207/s1532690xci2402_3

Star, J. R., & Rittle-Johnson, B. (2009). It pays to compare: an experimental study on computational estimation. *Journal of Experimental Child Psychology*, *102*(4), 408–26. doi:10.1016/j.jecp.2008.11.004

Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review*, *84*(4), 353–378. doi:10.1037/0033-295X.84.4.353

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. doi:10.1207/s15516709cog1202_4

Van Gog, T., Paas, F. G. W. C., & van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction*, *18*(3), 211–222. doi:10.1016/j.learninstruc.2007.03.003

VanLehn, K. A., & Jones, R. M. (1993). (1993b). What mediates the self-explanation effect? Knowledge gaps, schemas or analogies? In M. Polson (Ed.), *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 1034– 1039). Hillsdale, NJ: Erlbaum.

Veenman, M. V. J., Elshout, J. J., & Meijer, J. (1997). The generality vs domain-specificity of metacognitive skills in novice learning across domains. *Learning and Instruction*, *7*(2), 187–209. doi:10.1016/S0959-4752(96)00025-4

Veenman, M. V. J., & Verheij, J. (2001). Technical students' metacognitive skills: relating general vs. specific metacognitive skills to study success. *Learning and Individual Differences*, *13*(3), 259–272. doi:10.1016/S1041-6080(02)00094-8

Ward, M., & Sweller, J. (1990). Structuring Effective Worked Examples. *Cognition and Instruction*, *7*(1), 1–39. doi:10.1207/s1532690xci0701_1

Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational Psychologist*, *30*(4), 173–187. doi:10.1207/s15326985ep3004_2

Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: a framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, *43*(1), 49–64. doi:10.1080/00461520701756420

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

Wolters, C. A., & Pintrich, P. R. (1998). Contextual differences in student motivation and self-regulated learning in mathematics, English, and social studies classrooms. *Instructional Science*, *26*(1-2), 27–47.

Wong, R. M. F., Lawson, M. J., & Keeves, J. (2002). The effects of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction*, *12*(2), 233–262. doi:10.1016/S0959-4752(01)00027-5

Zepeda, C. D., Richey, J. E., Ronevich, P., & Nokes-Malach, T. J. (n.d.). Direct instruction of metacognition benefits adolescent science learning, transfer, and motivation: an in-vivo study. *Journal of Educational Psychology*. doi:10.1037/edu0000022

Zimmerman, B. J. (2000). Attaining self-regulation: a social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press.

Zimmerman, B. J. (2011). Motivational sources and outcomes in self-regulated learning and performance. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 49–64). New York: Routledge.