UNIVERSITY OF CALIFORNIA

Los Angeles

An Analogical Approach to STEM Education

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy in Psychology

by

Maureen Gray

2021

ABSTRACT OF THE DISSERTATION

An Analogical Approach to STEM Education

by

Maureen Gray

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2021

Professor Keith Holyoak, Chair

STEM education is a persistent problem in the United States. Analogy offers a potential tool for improving educational outcomes because analogical comparison increases attention to the structural-relational information that characterizes experts' conceptual representations. The current project investigated analogy-inspired instruction in two lab studies using UCLA undergraduates and one naturalistic classroom study. In Study 1, UCLA undergraduates learned about STEM concepts from lecture videos using analogical principles or control videos, and performance was assessed with an immediate posttest. Performance was similar across both instructional conditions, which may be attributable to the high-ability sample. In Study 2, UCLA undergraduates learned how to solve equation construction problems from videos that represented relational information explicitly in a geometric format, in a carefully-matched symbolic format, or in an adaptation of the gold standard of instruction for this topic, JUMP Math. While all lessons improved performance, the geometric and symbolic lessons were most effective. As in Study 1, the high-ability sample demonstrated an ability to learn from all types of instruction. The classroom study investigated the efficacy of analogical instruction in an online class environment in the context of cognitive load theory. UCLA students enrolled in Life Sciences 30A: Quantitative Concepts for Life Scientists (in Winter quarter 2021) learned topics through a structured teacher-directed approach to analogical instruction or a less-structured student-directed approach, and exam performance was measured. Students benefitted from the teacher-directed approach and the benefit was especially pronounced for low-performing students. Implications for designing educational interventions for students with lower abilities, and for successful researcher-practitioner collaborations, are discussed.

The dissertation of Maureen Gray is approved.

Patricia Cheng

Catherine Sandhofer

James Stigler

Keith Holyoak, Committee Chair

University of California, Los Angeles

2021

Contents

List of Tables

List of Figures and Equations

Acknowledgments

A dissertation is an immense individual endeavor, but by no means was it accomplished by me alone. This work was possible only with the gracious support of a great number of people who have lifted me up both professionally and personally. I do not intend to pare these relationships down by simply committing a fragment of them to paper. They are in essence indefinable, complex, and larger than an “Acknowledgments” section could contain. With the understanding that this section is fundamentally limited in scope and that my gratitude extends deeper than words can express, I would like to thank the following people.

My academic advisor Keith Holyoak has supported me for many years and has taught me innumerable things, among them the value of expertise. I recall reading Gick and Holyoak (1980) for the first time as an undergraduate at the University of Illinois at Urbana-Champaign and thinking to myself, “Wow, it would be a dream to work with Keith Holyoak someday when I go to graduate school.” While my time as a graduate student has not been uniformly dreamlike, I can say with certainty that it has been a privilege to be guided through this process by Keith. He granted me the freedom to pursue and develop my own interests and my work and intellectual curiosity have benefitted from his continued excitement for psychology. It is a great joy to watch Keith puzzle over a problem— his considerable breadth of knowledge has never failed to surprise me and his deep understanding of cognitive psychology has enriched my own.

Before I ever set foot at University of California, Los Angeles, graduate school was just a dream. My undergraduate mentor, John Hummel, helped me to make that dream a reality. John taught me that science is fun and he ignited my passion for analogical reasoning. My memories of John’s lab meetings are some of the fondest I have. Although I was an undergraduate, he treated my ideas with utmost respect and genuine joy. I would not be here today without John.

I would also like to thank some members of the UCLA community. First, I am grateful to my committee members for their support and valuable insights. Dr. Patricia Cheng's thorough and careful approach to psychological science is matched by her enthusiasm. I am inspired by her admiration for the beauty of scientific explanation and by her belief that science can change the world if people can see it clearly and connect to it. Dr. Jim Stigler very helpfully advised me on conducting classroom research. I didn't have the resources to conduct a randomized controlled experiment that takes place *within* a single classroom (like those that come out of Jim's lab) but it has been eye-opening to learn from him. He has broadened my perspective of research in education and helped me to draw connections between researchers and practitioners. Jim encouraged me to fit my work into a wider context and to think of classrooms as they are: messy, open systems. Dr. Cathy Sandhofer's expertise in cognitive development has been a valuable asset to me as well. Her graduate seminar is one of the best classes I have taken in graduate school and she exposed me to a novel developmental view of learning. On the administrative side, psychology graduate students are incredibly fortunate to have Lisa Lee looking out for us. Her kindness and conscientiousness know no bounds.

I am loathe to be melodramatic, but it is not an overstatement to say that my life changed for the better the day I walked into Dr. Jane Shevtsov's office in March of 2019. I would like to thank Jane for a great many things and what follows is merely a subset. Working with Jane has demonstrated to me the excitement and productive capability of truly interdisciplinary work. In short, this dissertation grew out of the connection between Jane and I. Before I met Jane, I believed emphatically that I was not a "math person" and that math would never make sense to me; it just wasn't meant to. Jane showed me that there really is no such thing as a "math person". It may come easier to some, but I know now that math is a logically consistent set of interrelated

concepts and that excellent teaching can uncover its beauty. Jane's expertise in biology and mathematics and her genuine curiosity for cognitive psychology threw my own developing expertise into sharp relief and this was instrumental in my development as an expert in my own right. Jane has mentored me both as a researcher and as a teacher. Her rigorous approach to science and statistics has improved my understanding of both and the current work is much improved as a result. She advised me on statistical approaches appropriate for non-normal data and assisted with material development throughout the duration of the project. I greatly admire her devotion to science and her students—she is a paragon of a researcher-practitioner. She is committed to discovering empirically-supported educational interventions and working them into her own teaching. Our meetings often wander, but they are productive. They often go long, we laugh a lot, and I would not have it any other way.

Outside of academia, my family has played an immeasurable role in lifting me up. At the start of the Covid-19 pandemic in March 2020, I came home to spend the short spring break with my family, seven shirts stowed away in my suitcase. I never moved back to LA and I am grateful every single day that I get to spend with these incredible people. My family has taught me that relationships are what make life meaningful and they have imbued untold meaning to mine. I am so fortunate to have brothers that I can count among my best friends. Doug and G are great company and spending time with them kept me grounded and sane when it felt like the world was falling apart around us. Purposefully or not, Doug has reminded me that getting older doesn't mean you have to stop playing. Doug will always be able to make me laugh and I can't thank him enough for that. For the rest of my life, I will fondly remember the night that he and Marin helped "translate" *Kung Fu Yoga* for G and I. I admire his commitment to enjoying his life and I hope to emulate it when I am finished with graduate school. G has been a boundless source of emotional support. They are kind, sensitive, and empathetic beyond belief. G understands how difficult life can sometimes be and I can always go to them for unconditional support and compassion. As an added bonus, G is stupidly fun to be around and has endless patience for my inability to play video games.

I have been putting off writing about my parents because these relationships are particularly difficult to distill into words. One expects their parents to be a solid support system and mine are no exception. Less expected is that my parents are my best friends. I can talk to them about anything and I love spending time with them. They have instilled in me a lifelong love of learning and have pushed me to accomplish everything I set my mind to.

My mom is a model of both perseverance and love. She has overcome considerable challenges and I greatly admire her resilience. She is constantly on my side and I will always remember with fondness the long phone conversations we shared while I ambled to and from campus. She is giving, compassionate, and amazingly friendly – the special kind of person that makes you feel like you're her friend even though you've just met. When I think of her I see a vast web of the social connections that she has forged and thoughtfully maintained. She is at the center of a network that she can draw upon for support because she has put forth effort to keep it strong. To be a part of that network is something wonderful and to know her and love her is a joy.

My dad is, in short, a problem solver. "For what kind of problem?", you might ask. Why, any problem, any problem at all! Need a haircut? He's got you— just grab a butter knife. What's that? You've cut your knee and need to get your 19 stitches removed and the doctor's office will charge you $194? Don't worry—he'll do it for $193. How about an athletic coach, teaching mentor, music recommendation generator, and career advisor? He's got you covered. Some of his ideas are better than others (I'll let you decide where "disposable clothes" falls on that spectrum) but I am constantly grateful for his perspective, his sense of humor, and his high Need for Cognition. He is a master of applying psychological principles, looking past surface features to see and pull apart the deep structure of a situation. The desultory stream-of-consciousness conversations we have on long runs or drives to and from campus are some of the highlights of my week. He pushes me to achieve my goals as an athlete, an academic, an educator, and as a human being. My dad is my role model and his support is instrumental to my success.

My friendships have sustained me and throughout my time in graduate school. My oldest friends remain some of my closest to this day. In third grade, Lisa moved in down the street, promptly broke her arm on the swing set at my house, and became my best friend. We have come a long way since then (e.g., Lisa no longer hits me on the head with her cast) and we have not lived in the same place for more than a month since 2012. We have grown and changed separately from one another for the majority of our adult lives but the people we have grown into are just as compatible. Lisa's forceful opinions on politics and societal issues are carefully constructed and well-reasoned. Her ability to uplift me and the women around her is admirable and she is a force for good in this world. In high school, cross country facilitated the formation of several other lifelong connections. Dale, Sofia, and Catherine and I were inseparable then and these people have remained important to me. Dale is and always will be a superstar. She is eminently capable and I am constantly in awe of her achievements and level-headed approach to everything that comes her way. She has kept me grounded and she always makes time for her friends because she wants us to know how important we are to her. Dale's support and validation have been a lifeline through this experience and it is a great privilege to be her friend. Sofia's incredible resilience and her commitment to pursuing her goals and self-expression is an

inspiration to me. Plus, she gives good book recommendations. Cat is a delight to be around and I will always value the time we spent squashed together on her couch.

I still cannot quite believe my luck in meeting Anika Guha during my first year of graduate school. She has been a constant source of support and I can count on her for silly tv, sound advice on things as wide-ranging as cooking techniques or mental health, and an alternative viewpoint on any problem I might have. She has encouraged me to be self-reflective and to take care of myself. It has been a great pleasure to be a part of her life as her expertise grows and I am a better person for having met her. Mary Flaim is about as cool as cool gets in my book and her willingness to be herself has made me more confident in doing the same. She has excellent taste in horror movies and books, bakes like a champion, and has lent a (very cool gauged) listening ear to my laments about some of the more unpleasant aspects of graduate school. Amalia Ionescu made the miserable moments more bearable with her devastating sense of humor and exposed me to a very good Romanian salad and some high quality film and tv. Nick Ichien is a talented creative thinker— listening to him think aloud and participating in thought-provoking discussions with him has been a delight in my graduate experience. Finally, I would like to thank my partner, Charlie Brady. At times throughout this process, it felt as if his unwavering belief in me kept me going and I cannot thank him enough for his love and support.

Vita

EDUCATION

University of California, Los Angeles
MA in Cognitive Psychology — 2017
University of Illinois at Urbana-Champaign
BS in Psychology and Statistics — 2016
Graduated with Highest Distinction and Honors in Psychology, High Distinction in Statistics

EXPERIENCE

Adjunct Faculty, College of DuPage; Glen Ellen, IL — August 2021 - present
Courses taught: General Psychology, Abnormal Psychology

- Designed and taught full-semester courses on General Psychology and Abnormal Psychology using engaging lecture videos supplemented by interactive lectures and group activities during face-to-face meetings

Instructional Design and Technology Assistant, Online Teaching and Learning Initiative, UCLA; Los Angeles, CA — April 2021 - June 2021

- Assisted with development and delivery of a workshop designed to guide UCLA faculty through designing high-quality online courses
- Collaborated with UCLA faculty and senior instructional designers to adapt in-person courses for delivery in an online setting
- Designed and proofed course websites in Canvas and Moodle

Adjunct Faculty, Los Angeles Valley College; Los Angeles, CA — June 2019 - June 2020
Courses taught: Introductory Psychology, Foundations of Statistical Reasoning, Statistics for Behavioral Sciences

- Designed and taught full-semester and summer session courses on introductory psychology and statistics
- Developed engaging online activities and fair assessment practices to support student learning during remote instruction

Graduate Instructor of Record, UCLA; Los Angeles, CA — June 2020 - July 2020

- Applied for and received a competitive grant to independently design and teach a section of Introductory Psychology
- Responsible for design of all assignments, lectures, and assessments

Curriculum Consultant, UCLA; Los Angeles, CA — March 2019 - September 2019

- Analyzed student error patterns in biological mathematical modeling course to diagnose problems with student learning and understanding
- Reviewed relevant literature to give context to observed problems
- Recommended empirically validated and theoretically motivated improvements to course design

PUBLICATIONS

Gray, M. E., & Holyoak, K. J. (2021). Teaching by Analogy: From Theory to Practice. *Mind, Brain, and Education.* https://doi.org/10.1111/mbe.12288

Gray, M. E., & Holyoak, K. J. (2020). Individual differences in relational reasoning. *Memory & Cognition*, *48*(1), 96–110. https://doi.org/10.3758/s13421-019-00964-y

Gray, M. E., & Holyoak, K. J. (2018). Individual differences in relational reasoning. In C. Kalish, M. Rau, J. Zhu & T. T. Rogers (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Gray, M. E., DeWolf, M., Bassok, M., & Holyoak, K. J. (2018). Dissociation between magnitude comparison and relation identification across different formats for rational numbers. *Thinking & Reasoning*, *24*(2), 179-197. https://doi.org/10.1080/13546783.2017.1367327

TALKS

Panelist, *Increasing Engagement in and out of the Classroom: Slack and Discord as Community-Building Tools*. College of DuPage Instruction Committee. (2021, October). Virtual.

Gray, M.E. *Incorporating Slack into your course.* UCLA Online Teaching and Learning Initiative. (2021, January). Virtual.

Gray, M.E. *Using Slack to Promote Meaningful Discussion.* UC Psychology Teaching and Learning Conference. (2020, September). Virtual.

Panelist, *Summer Workshop in Online Instruction* presented by the Online Teaching and Learning Initiative in cooperation with the Provost's Office and Graduate Division. (2019, July). University of California, Los Angeles; Los Angeles, CA.

Gray, M. E., Holyoak, K. J. *Individual differences in relational reasoning.* (2018, October). Cognitive Forum, University of California, Los Angeles; Los Angeles, CA.

Gray, M. E. *Preparing for and applying to graduate school.* (2016, June). College of DuPage, Glen Ellen, IL.

SERVICE

Volunteer Research Group Co-Facilitator, College of DuPage — Fall 20-present
Advise students on various stages of the research process, including literature review, experiment design, data analysis, and interpretation of results.

Cognitive Area Lecture Series Coordinator, UCLA — Fall 18-Winter 19
Coordinated and publicized a weekly lecture series in which speakers from UCLA and other universities give talks on various topics related to cognitive psychology. Directed a team that completed individual tasks relevant to the lecture series.

Psychology Graduate Student Association student representative on TA/Fellowship committee, UCLA — Fall 17 - Fall 20
Represent interests of graduate students at quarterly TA/Fellowship meetings, communicating with various faculty members and department staff.

SELECTED HONORS AND AWARDS

Shepherd Ivory Franz TA Award, UCLA Psychology Department, June 2021
Distinguished Teaching Award Nominee, UCLA Graduate Division, June 2021
Teaching Practicum Program, UCLA Psychology Department, 2020
Graduate Research Fellowship Program Honorable Mention, National Science Foundation, 2017
Dean's Award, University of California, Los Angeles, 2016
Janet Tritsch Award for Outstanding Thesis, University of Illinois at Urbana-Champaign, 2016
Outstanding Undergraduate Student Award, University of Illinois at Urbana-Champaign, 2016
Edmund J James Scholar Honors Student, University of Illinois at Urbana-Champaign, 2012-2016

Chapter 1: Introduction

Science, technology, engineering, and mathematics (STEM) fields present learning challenges for many students, and the United States has typically lagged behind comparable developed nations on standardized tests in science and mathematics. For example, American high school students placed 38th out of 71 surveyed countries in an international mathematics assessment and 24th in science (OECD, 2018). This issue is even more pronounced for students with traditionally under-represented minority (URM) backgrounds: achievement gaps between white and URM students have persisted for decades (Gonzalez & Kuenzi, 2012).

The problem is more severe than poor performance on standardized tests; rather, it stems from a lack of conceptual understanding in the fields of math and science (Maloney, O'Kuma, Hieggelke, & Van Heuvelen, 2001; Richland, Stigler, & Holyoak, 2012). In a study investigating the mathematical knowledge of American community college students, Stigler, Givvin, and Thomson (2010) found that these students failed to understand that mathematics is fundamentally a system of logically related concepts. Instead, the students believed math to be a collection of unrelated procedures that must be memorized and applied in an inflexible manner. Undergraduate physics students similarly lack conceptual understanding, instead relying on rote application of formulas to solve physics problems (Jonassen, 2003). Given the consistent and pervasive weakness of STEM education, it is clearly an important aim to foster greater conceptual understanding in STEM fields and to improve STEM education.

Work on relational reasoning and analogy, in particular, may help to improve educational outcomes in these fields. Many concepts in STEM fields are relational in nature (i.e., defined by shared relational structure as opposed to shared features), and mathematics is formally a system of relations (Devlin, 2012). Relationally-defined categories are more challenging to learn than their feature-defined counterparts (Gentner & Kurtz, 2005) because they cohere around shared structure and do not necessarily share surface features, which makes identifying new instances more difficult. Further, much previous work has shown that novices have trouble noticing and capitalizing on shared structure of concepts. Instead, novices frequently focus on superficial characteristics, which are generally not central for problem solving (e.g., Chi, Feltovich, & Glaser, 1981; Rottman, Gentner, & Goldwater, 2012; Stains & Talanquer, 2008). For example, physics experts are able to look past superficial dissimilarities and categorize physics problems based on common principles (e.g., grouping problems together that can be solved by applying Newton's second law). Novices, on the other hand, do not notice shared principles and instead categorize problems based on superficial components of the problem (e.g., grouping together all problems containing springs) (Chi et al., 1981). The divergence in classification strategies used by novices and experts suggests that novices' representations of physics problems are not strongly linked to relevant physics concepts. This gap in conceptual knowledge is likely to manifest itself as difficulty with problem-solving. In addition to indicating a gap in conceptual understanding, inattention to relationships is a barrier to success in STEM fields.

Furthermore, it has been argued that the "end goal" of education is to foster the development of abstract relational schemas that can be applied flexibly and transferred to diverse situations (Goldwater & Schalk, 2016). For example, if a student learns how to analyze the structure of an argument in a philosophy class, but later does not apply that knowledge to political arguments, their education has in an important sense failed. Failure to transfer knowledge from the context in which it was learned to a novel context is a disappointingly robust finding (Renkl, Mandl, & Gruber, 1996). Novices' knowledge representations do not prominently feature structural-relational information that remains constant across contexts and this weakness in conceptual representation contributes to the inert knowledge problem (Fries, Son, Givvin, & Stigler, 2021).

The difficulty inherent in learning relational categories is compounded by ineffective teaching techniques that often leave the structural nature of STEM concepts and the relations between them implicit (Boaler, 2015; Richland et al., 2012; Tekkumru-Kisa, Kisa, & Hiester, 2020). Compared to K-12 instructors in Japan and Hong Kong, American instructors do not effectively highlight the conceptual structure of mathematics and connections between mathematics concepts (Richland, Zur, & Holyoak, 2007). When American teachers do present problems that encourage students to make connections, they frequently lower the cognitive demands of the task (and thus its potential benefit) by focusing on accuracy of procedures rather than reasoning processes (Henningsen & Stein, 1997). Further, teachers often provide inappropriate scaffolding and effectively do the problem for the students (Stein, Grover, & Henningsen, 1996). These practices effectively transform the problem into yet another opportunity to practice procedural skills.

Science education is similarly problematic; in many K-8 classrooms, science is presented as a static body of accumulated knowledge that must be memorized and not as a process of reasoning and discovery (Tekkumru-Kisa et al., 2020). One study compared 8th grade science lessons from the US and four other higher achieving countries (Japan, the Czech Republic, Australia, and the Netherlands) (Roth & Garnier, 2006). In contrast with higher achieving countries, scientific content did not play a central role in over 70% of the videotaped American lessons. Further, when American teachers did present content, it was frequently "[organized] as a collection of discrete facts, definitions, and algorithms rather than as a connected set of ideas" (Roth & Garnier, 2006, p. 20).

Given the importance of relational knowledge in STEM education, analogical reasoning is a plausible domain of research from which techniques to improve STEM education can be drawn. Analogical reasoning, which involves a reasoner comparing two structurally (i.e., relationally) similar situations that may differ in their surface features, enables the powerful capacity to use knowledge about the source domain (typically well understood) to derive inferences about a target domain (typically less well understood). In this way, analogy allows a reasoner to capitalize on prior knowledge by using a familiar situation to better understand a novel or unfamiliar situation (for a review, see Holyoak, 2012). Importantly for the current project, analogical comparison necessitates attention to relational structure.

The benefits of incorporating analogies and analogical reasoning in STEM education have been robustly demonstrated (see Alfieri, Nokes-Malach, & Schunn, 2013 for a review). For example, analogical comparison of physics concepts increases far transfer (Nokes-Malach, VanLehn, Belenky, Lichtenstein, & Cox, 2013), and several studies have shown that including analogies in science texts increases comprehension of the scientific material in question and its causal relational structure (Braasch & Goldman, 2010; Clement & Yanowitz, 2003; Jaeger & Wiley, 2015). In the domain of mathematics, analogical comparison of worked examples illustrating proportionality led to higher performance on test problems and reduced common misconceptions (Begolli & Richland, 2016).

While the learning outcomes associated with including analogies in learning opportunities are typically positive, this approach is not without its potential pitfalls. Students in any classroom are not uniform. They do not come to a learning opportunity with the same expectations, prior knowledge, or cognitive resources, and this variability in student populations undoubtedly affects the efficacy of any proposed educational interventions. It is especially important to keep the impact of individual differences in mind when considering URM students. These students are typically less academically prepared to begin college than their peers (Chen, 2005), and educational interventions designed to benefit disadvantaged children often improve educational outcomes for their advantaged peers to an even greater degree (Ceci & Papierno, 2005). Clearly, concerns regarding efficacy of educational interventions are especially relevant for these disadvantaged students.

Previous research has implicated executive functions (e.g., working memory, fluid intelligence, inhibitory control) in analogical reasoning (Gray & Holyoak, 2020; Hummel & Holyoak, 1997, 2003; Kubricht, Lu, & Holyoak, 2017; Tohill & Holyoak, 2000; Waltz, Lau, Grewal, & Holyoak, 2000) and in academic achievement (e.g., Campos, Almeida, Ferreira, Martinez, & Ramalho, 2013). One consequence of this connection is that students with fewer cognitive resources may be less able to benefit from analogies in educational settings. In fact, some research suggests that less able students may even be harmed by the presence of analogies in educational material (Jaeger & Wiley, 2015; Zook & Maier, 1994).

Since executive functions are correlated significantly with academic achievement (Campos et al., 2013), students with fewer cognitive resources at their disposal are already at a significant disadvantage in most educational settings. The learning capabilities of these students should not be discounted, however, as some studies have shown that children who attain low scores on figural analogy tests have similar potential for learning as children with higher scores (Touw, Vogelaar, Verdel, Bakker, & Resing, 2017). Furthermore, if insufficient executive functioning is the root of the problem, educational analogical interventions could be altered to lessen the cognitive load on the reasoner (e.g., Begolli & Richland, 2016; Richland &

McDonough, 2010; Richland et al., 2007). The ideal method to incorporate analogy into STEM education while keeping cognitive load in mind remains unclear.

The goal of the current project is to leverage principles from the field of cognitive psychology in general and analogical reasoning in particular to improve educational outcomes in STEM. More specifically, the aim is to improve students' conceptual representations and ability to transfer their learning to novel contexts. This problem will be investigated in a specific course in a specific domain: Life Sciences 30AB: Quantitative Concepts for Life Scientists.

The LS 30 series is not a typical math class. It is a math class designed to build quantitative skills relevant to modern biology such as modeling, computation, and simulation. Thus, topics like single- and multiple-variable calculus, differential equations, and mathematical modeling are embedded in a rich, meaningful biological context. Most students in the LS 30 sequence perform relatively well. Typical students might have difficulty with a few isolated concepts in the course, but most often these problems are addressable if the student seeks additional help in discussion section or in office hours. For a portion of struggling students, however, seeking a second explanation of a difficult concept is not enough. Discussions with these students in office hours reveal significant gaps in knowledge of course content and virtually no understanding of the interrelated structure of the material. The STEM achievement gap is present in this context – this population of struggling students largely consists of URMs.

Poorly (or un-) structured conceptual knowledge does not necessarily arise from a lack of effort on the students' parts. It may arise from students focusing on the wrong aspects of concepts and failing to notice structural similarities. American K-12 instructors fail to impart the existence and importance of structure in mathematics and science (Richland et al., 2007; Roth & Garnier, 2006), and as a result many students enter LS 30 classrooms without the expectation that there exists deep underlying structure to be discovered. This weak conceptual knowledge may manifest itself as a failure to recognize novel problems as structurally similar to learned problems on a test, or as failure to flexibly manipulate known procedures to solve a new kind of problem (i.e., transfer failure). Such problems have been informally corroborated by the LS 30 instructors (Shevtsov, 2019, personal communication).

An Analogical Approach

The suggested interventions (collectively referred to as an analogical approach) are based on a set of principles drawn from cognitive psychology and the field of analogical reasoning. The principles are summarized in Table 1. The analogical approach provides a theoretically driven and empirically supported framework for instructors to devise their own interventions to maximize the benefit of analogy in education. The current recommendations build upon previous guides to analogical teaching (see, for example Treagust, Harrison, & Venville, 1998; Vendetti, Matlen, Richland, & Bunge, 2015). Here I consolidate techniques that service the same general instructional goals, propose additional methods to incorporate analogy into classrooms, and highlight the importance of considering limitations in students' cognitive resources. The principles described here are not intended to constitute an exhaustive set, but they all serve to highlight causal relations that are crucial to learning and transfer, particularly in STEM fields.

One of the strengths of an analogical approach is its flexibility – it is not a prescription for a specific educational intervention or a call for teachers to drastically alter their teaching style. Instead, I have compiled a list of principles that instructors can draw upon as they see fit to make modest changes to the way in which they present material to their students. Although these instructional changes are fairly simple, the potential for benefits to student understanding are significant.

1. Capitalize on prior knowledge

Analogies allow learners to use prior knowledge to better understand an unfamiliar topic, and instructors would be well-advised to take advantage of this characteristic of analogy when possible. Many instances of analogical reasoning occur in situations in which there is an imbalance in knowledge: the reasoner draws on prior knowledge of a source domain to aid in understanding an unfamiliar target domain. The positive effect of prior knowledge on learning is well documented (Chiesi, Spilich, & Voss, 1979; McNamara & Kintsch, 1996). In general, prior knowledge alters the encoding of new knowledge (Boshuizen & Schmidt, 2008; Gobet & Simon, 1996; Kimball & Holyoak, 2000), and this is no less true in the specific case of analogical encoding. Grasping that some new material is analogous to prior knowledge allows the reasoner to match corresponding elements of the target material into roles and relations that are already stored in memory, thus aiding comprehension (Bean, Searles, Singer, & Cowen, 1990). The powerful ability to draw inferences can guide scientific discovery and improve understanding (Holyoak & Thagard, 1995; Yanowitz, 2001). Further, using a familiar real-world experience as a source analog may increase student motivation, as the analogy provides an example of the application and relevance of classroom content (Duit, 1991).

An important type of situation in which prior knowledge may be particularly helpful in understanding novel material arises for concepts that cannot be directly perceived – either because they are too small (like submicroscopic particles), too large (like plate tectonics), or too abstract (like the human mind). A well-conceived analogy has the ability to bring such rather un-imageable constructs into a tangible, imageable realm. In so doing, an analogy makes these constructs easier to understand and reason about.

Reviews of science textbooks show that many textbook authors are sensitive to this characteristic of analogy – approximately 90% of analogies found in a collection of college-level biochemistry textbooks related concrete source analogs to abstract target concepts (Orgill & Bodner, 2006). There are also several examples of studies in which analogies were used to improve comprehension of a difficult-to-visualize concept (Baker & Lawson, 2001; Braasch & Goldman, 2010; Jaeger, Taylor, & Wiley, 2016). One study investigated the use of analogy to aid understanding of El Niño weather systems (Jaeger et al., 2016). Participants with low spatial reasoning abilities (i.e., those who would likely have the most trouble visualizing the large-scale, hard-to-visualize weather system) showed significant improvement in comprehension when they received a text with an analogy to a small-scale, more easily imageable situation (letting the air out of a balloon) (Jaeger et al., 2016). This benefit of analogy has also been demonstrated in classroom settings. College-level genetics students showed increased comprehension of abstract genetics concepts after receiving instructional analogies to help teach the concepts when compared to students in the same course who did not receive instructional analogies (Baker & Lawson, 2001).

Although analogy can be helpful in many circumstances, a student must possess the requisite prior knowledge in order to capitalize upon it. Educators should not assume that all students have sufficient prior knowledge to benefit from an analogy discussed in class. Several studies have demonstrated that analogies are particularly helpful for individuals with a high level of prior knowledge (Braasch & Goldman, 2010; Jee et al., 2013; Rittle-Johnson, Star, & Durkin, 2009). Second, instructors should not assume that students know how to effectively utilize analogies, even if they have been identified by an instructor (Venville, Bryer, & Treagust, 1994). Simply stating a cell is analogous to a factory may leave some students uncertain about which aspects of the factory are like the cell and which aspects are not, so educators should take care to explain the source analog and its correspondence to the target concept fully. Finally, in nearly all cases the analogy will not be perfect. Some aspects of the source analog may not have corresponding elements in the target concept, and thus should not be carried over to the target concept. In order to prevent inappropriate prior knowledge from producing misconceptions, instructors should explicitly map the relevant correspondences and point out the limits of the analogy—that is, which aspects of the source analog are irrelevant. When possible, the description of the source may be selectively tailored to optimize the analogical match with the target.

2. Highlight shared structure

Several methods to increase attention to shared structure have been proposed, and instructors can use these techniques to guide students to attend to the most important aspects of to-be-learned material. Aligning and comparing analogous examples focuses attention on shared relational structure, which emphasizes the critical characteristics of relationally-defined STEM concepts. Substantial work has demonstrated that analogical comparison increases attention to relations (Catrambone & Holyoak, 1989; Gentner & Markman, 1997; Gick & Holyoak, 1983; Goldwater & Gentner, 2015; Kotovsky & Gentner, 1996), and that alignment and analogical comparison of exemplars improves learning outcomes in STEM fields (Alfieri et al., 2013; Begolli & Richland, 2016; Gentner et al., 2016; Klein, Piacente-Cimini, & Williams, 2007; Nokes-Malach et al., 2013; Richland & McDonough, 2010).

Educators with extensive domain knowledge may believe that similarities between analogous cases are obvious, but research has consistently demonstrated that this is not true for novices (Chi et al., 1981; Gick & Holyoak, 1980, 1983; Stains & Talanquer, 2008). Further, significant work has shown that analogical reasoning is a resource-intensive process. Analogical processing relies on several constructs that comprise cognitive capacity, including working memory, fluid intelligence, inhibitory control, and spatial abilities (e.g., Krawczyk et al., 2008; Viskontas, Morrison, Holyoak, Hummel, & Knowlton, 2004; Waltz et al., 2000). Individuals with weaker executive functioning capabilities will have a harder time carrying out analogical reasoning and are less likely to benefit from it in educational settings (Jaeger & Wiley, 2015; Richland & McDonough, 2010). These sources of individual variability must be taken into account when using analogy to highlight the shared structure among examples of a STEM concept. When analogical comparisons are introduced, they should be labeled as such to obviate the challenging process of noticing analogical similarity.

Comparing examples highlights shared structure, but the presentation of the examples should facilitate comparison without overloading limited capacity resources. For example, shared structure can be communicated through visuo-spatial cues in the form of static visual support (Begolli & Richland, 2016; Matlen, Vosniadou, Jee, & Ptouchkina, 2011; Rittle-Johnson & Star, 2007). Simultaneous static presentation of exemplars frees up cognitive resources to devote to the comparison process and attend to the target material, obviating the need to hold both analogs in working memory at the same time. In addition to keeping all analogs visible during comparison, analogical processing can be facilitated through the specific presentation style of the visual representations. Displaying visual representations such that corresponding elements are spatially aligned can support greater learning (Matlen et al., 2011; Richland et al., 2007). Color coding can also be used to emphasize the entities in different exemplars that play analogous roles (see Figure 1.1). Simultaneous presentation of the analogs and of the layout of the entities being compared improves the facility of the comparison process.

Visuo-spatial methods of emphasizing shared structure, while effective, are largely implicit. To maximize attention to this crucial aspect of the to-be-learned material, the correspondences between analogs should be described explicitly. This can be accomplished through verbal descriptions of the correspondences in which the instructor explicitly points out that two entities play the same role in analogous situations. In addition, students can be prompted to attend to this shared structure through the use of guided compare-and-contrast prompts. General prompts to compare two situations don't reliably focus attention on the most relevant dimensions of comparison, but directed compare-and-contrast instructions (e.g., to compare instances and identify their similarities) lead students to notice shared structure (Catrambone & Holyoak, 1989; Gick & Holyoak, 1983).

3. Align and connect semantic, mathematical, and graphical representations

In complex domains, many concepts can be represented in multiple formats. For example, a negative feedback loop can be represented verbally, visually in a diagram or graph, or symbolically as a set of differential equations. Exposure to multiple representations of the same concept has the potential to improve STEM learning (Cheng, 2000; Nistal et al., 2009) and practicing making connections between representations can improve statistics education (Fries et al., 2021). However, students must see and understand the connections between representations in order to benefit from them. Correspondences that seem intuitive for domain experts are less obvious to novice learners. Analogy offers suggestions for guiding students' attention to these correspondences and for increasing their ability to map between representations.

For example, Bassok and colleagues (1998) introduced the construct of *semantic alignment*, which refers to the tendency to maintain systematic correspondence between the semantic relations that exist between pairs of real-world objects and mathematical relations between arguments of arithmetic operations. The alignment between real-world and mathematical knowledge guides the application of abstract mathematical knowledge by both students and textbook writers, not only in the United States (DeWolf, Bassok, & Holyoak, 2015; Rapp, Bassok, DeWolf, & Holyoak, 2015) but also in South Korea (Lee, DeWolf, Bassok, & Holyoak, 2016) and Russia (Tyumeneva et al., 2018). Semantic alignment is a case of connecting verbal semantic representations with formal symbolic representations. Importantly, the alignment itself and its influence on problem solving procedures likely remains implicit in the mind of the problem solver and is not taught explicitly in schools. While previous research has not investigated the effects of making semantic alignment more explicit, previous work showing benefits resulting from highlighting shared structure (e.g., Begolli & Richland, 2016; Richland & McDonough, 2010) leads to the prediction that emphasizing the alignment between a model of a situation and its corresponding mathematical model should facilitate understanding of the shared structure and improve students' abilities to translate verbal models into mathematical ones.

The original conception of semantic alignment is fairly limited in scope, referring to the use of intuitive real-world knowledge in guiding mathematical reasoning. However, its potential applications in LS 30 and other classrooms calls for a broader interpretation of the construct: students should be explicitly instructed in the meaning of various mathematical operations and the mapping between the various representations (semantic, mathematical, and graphical).

Solid conceptual understanding of the meaning of mathematical operations is crucial for mathematical modeling, which is one of the central topics of the LS 30 series. One commonly misunderstood operation is multiplication: many students conceive of multiplication as repeated addition, but it is better understood as scaling by some multiplicative factor. In the very first week of class, students learn how to translate a set of verbal assumptions that describe a dynamic biological system into a series of differential equations that model the system. A student may be tasked with writing a mathematical expression to represent a population of rabbits, R, that grows at a rate b proportional to the population size. The conceptualization of "multiplication as repeated addition" does not support reasoning in this situation. However, if the conceptual meaning of multiplication is addressed, it is clear that the growth of the population should be represented by its current size scaled by the growth rate, and that this relationship can be expressed mathematically as multiplication, yielding $b * R$.

Whenever possible, instructors should explicitly discuss the conceptual meaning of mathematical operations, and in so doing explain their reasoning for selecting a particular mathematical representation to model a situation. In addition to explaining the conceptual meaning of mathematical operations, these abstract concepts should be connected explicitly to real-world referents in the word problems from which they were generated. To summarize, the recommendation is to highlight the conceptual meaning of mathematical operations, and to make explicit the alignment between various representations of a situation. Explicit acknowledgement of and instruction on the reasoning behind alignment may lead to improved performance on mathematical modeling and related problems that require connecting various representations.

4. Consider cognitive load

All learning imposes some cognitive load on learners (Sweller, 2011). Some load is due to the inherent difficulty of the material being learned, termed intrinsic cognitive load. Extraneous cognitive load is due to the particular manner in which the material is presented to learners and is not inherent to the complexity of the material. Germane cognitive load refers to processing demands related to creation of abstract schemas. While it is clear that instruction should not overload a student's processing capacity, the ideal balance of the sources of cognitive load is not clear. Increasing extraneous cognitive load leaves fewer resources to devote to the target material and generally harms learning (Richland & McDonough, 2010). Increasing germane cognitive load, however, may facilitate students' acquisition of the relational structure of target concepts (e.g., Paas & Van Merrienboer, 1994). The relation between the three sources of cognitive load was traditionally considered to be additive (Sweller, 2005), though more recent conceptualizations have characterized germane cognitive load as the proportion of working memory resources that are devoted to the intrinsic load of the target information (Sweller, 2010). In both conceptualizations, cognitive load theory emphasizes that educators must consider the load placed on the learner and maximize processing related to the target material.

In the context of cognitive load theory, analogical processing may increase both extraneous and germane cognitive load. Since analogical comparison is a resource-intensive process, instructors should present content in a manner that facilitates comparison and thus reduces the extraneous load placed on the learner. As noted in the section on highlighting structure, the extraneous cognitive load incurred by analogical comparison of exemplars can be minimized through small changes to the instructional delivery. First, the exemplars that are the focus of the alignment should be represented visually and presented simultaneously whenever possible. Corresponding elements of the exemplars should be aligned spatially, and may be written in corresponding colors to further highlight that they play analogous roles in their respective situations. In addition to these visuo-spatial methods, extraneous load may be reduced through the use of external scaffolding. Instructors can explicitly guide analogical comparisons and point out correspondences. Instructional materials (e.g., worksheets) can be written to structure the learning experience for the students.

Analogy also offers an opportunity to increase germane cognitive load. Novices often do not engage in cognitive processes that facilitate schema formation without prompting (e.g., Catrambone & Holyoak, 1989) and many instances of transfer failure can be explained through novices' incomplete schemas. For example, the most coherent way to sort physics problems is to use schemas for underlying physics principles. Physics novices, when left to their own devices, sort problems based on surface similarities (Chi et al., 1981). Redirecting the learner's attention to structural/relational information through use of analogical comparison may aid in schema construction. As long as these modifications to increase germane cognitive load do not overload the capacity limitations of the learner, they stand to improve learning outcomes. External supports, like instructors or worksheets, may guide students through analogical processing to reduce germane cognitive load if necessary.

5. Encourage generation of inferences

Generating information leads to better retention than passive study (for a review, see Bertsch, Pesta, Wiscott, & McDaniel, 2007), and the so-called generation effect has been demonstrated in educational settings (Metcalfe & Kornell, 2007). The effect of generating the correspondences between analogs has not been explicitly investigated, but some recent work suggests it may be similarly beneficial. Vendetti, Wu, and Holyoak (2014) compared the effects of generating solutions to semantically distant four-term analogies to passively viewing and evaluating completed analogies. Effects on memory were not examined, but generating solutions led to the induction of a relational mindset that biased attention to relational information in a subsequent unrelated task. These findings suggest that attention to relational information is malleable and, further, that generating relational information may push participants toward attending to it.

When students are first introduced to a concept using analogous examples, tasking them with generating the mappings between corresponding elements of the examples would likely impose too great a cognitive load. However, this technique may be introduced later on in a lesson, when students have some familiarity with the concept. Low-knowledge learners typically need significant scaffolding (see principles 2 and 4), but these techniques may lose their efficacy for high-knowledge learners. This transition in effective instructional techniques for low- to high-knowledge learners has been termed the expertise reversal effect (Kalyuga, 2007), and the analogical approach fits within this framework. Early on, instructors should provide explicit guidance on analogical comparison to prevent the comparison from overwhelming limited cognitive resources. As learners gain expertise, however, their need for instructor guidance is reduced. Generating mappings among analogs and drawing appropriate inferences directs attention to relational structure, which gives proficient students the opportunity to practice attending to the important relational information without direction from an instructor. This reflects testing and real-world contexts in which instructor guidance is conspicuously absent. Further, previous research suggests that generating the underlying structure will lead to greater retention of that structure (Bertsch et al., 2007).

An Analogical Approach at Work: Examples Using LS 30A Concepts

The principles may be applied in tandem to improve multiple aspects of instruction. I will now review some concrete examples of how an analogical approach can be applied to LS 30A concepts.

Model writing

One of the foundational ideas in LS 30A is that the state of a dynamic biological system can be represented by values of relevant variables at specific time points. Early on in the course, students learn how to translate a set of verbal assumptions that describe a biological system into a series of differential equations that model how the system changes. Figure 1.2 shows an example of four assumptions and the differential equations that can be constructed from them.

The analogical approach to teaching model writing emphasizes two things. First, mapping between the verbal assumptions and the mathematical operations that they correspond to should be highlighted. One way to accomplish this is to use corresponding colors, as shown in Figure 1.2. For example, the first assumption, colored in blue, corresponds to the blue term in the model below. Second, the instructor should explain the significance of each of the mathematical relations at work in the mathematical model representation and connect them to their real-world interpretations.

In the model shown in Figure 1.2, there are three different mathematical concepts at work that bear mentioning and should be explained by an instructor. The first mathematical concept at work is addition. Mathematically, adding a positive term to a quantity increases the magnitude of that quantity. In this context, the population is the thing to which we are adding a positive term. A positive term conceptually represents something that makes a population grow in size, such as the birth of new animals. The second mathematical concept at work is subtraction. Mathematically, subtracting is the inverse of addition: subtracting a positive term from a quantity decreases the magnitude of the quantity. In this context, subtracting a positive term conceptually represents something that makes a population decrease in size, such as death of animals by predation or old age. Finally, multiplication is also relevant. Mathematically, multiplication scales a quantity by some factor. In this context, several quantities are being scaled. For example, hares, H, are born at a constant per-capita rate, 0.1. In other words, we expect each existing hare to increase the population of hares by a factor of 0.1. Multiplying the current population of hares by the per-capita birth rate yields a growth term that is scaled by the birth rate.

Instructors should not expect that these interpretations will come naturally to students, as many students lack conceptual understanding of even basic mathematical operations (Stigler et al., 2010). Connecting verbal descriptions with their formal counterparts and explicating that correspondence will enhance students' conceptual understanding of mathematics and help them link abstract mathematical operations with real-world meaning. An analogical approach can increase understanding of the structure of mathematics and of the reasons for selecting particular mathematical models to represent real-world situations.

Feedback loops

Feedback loops occur when a change in one variable causes a later change in that same variable. Feedback loops may be negative or positive, but I will restrict my discussion to negative feedback loops, as the analogical approach to teaching both of these concepts is similar. It is helpful to start with an example that is likely familiar to students, like the relationship between temperature and air conditioner activity. When temperature increases, this causes the air conditioner to turn on. When AC activity increases, this causes temperature to decrease. The temperature-AC activity example is an example of a negative feedback loop, which occurs when an increase in a quantity (temperature) causes a later decrease in that same quantity, or when a decrease in a quantity causes a later increase in that same quantity. The top panel of Figure 1.1 shows a diagram of the temperature-AC activity example, which should be drawn on the board to reduce extraneous cognitive load.

After introducing the concept of negative feedback loops with a familiar example to capitalize on prior knowledge, instructors may then describe and diagram a second example, such as the relationship between glucose and insulin. When a person eats, glucose levels in the bloodstream rise. This causes the pancreas to secrete insulin, which lowers the level of glucose by helping the body to metabolize glucose. When diagramming this example, instructors should apply the analogical approach to highlight the shared structure that defines negative feedback loops. Specifically, instructors should draw the diagrams so elements that correspond to one another are spatially aligned, and use color to further emphasize elements of each example that play the same role (see Figure 1.1). The correspondences can be explicitly pointed out (e.g., temperature corresponds to glucose because each of these things causes something else to increase). Each of these modifications to instruction serves to highlight shared structure and to reduce extraneous cognitive load. Visually aligning the exemplars and writing corresponding components in the same color invites comparison of the exemplars and correspondingly increases the germane load of the instruction.

Next, students may be given a third example to solidify their understanding: the relationship between a population of sharks and a population of tuna. When the tuna population increases, that causes the shark population to increase because they have more prey available. When the shark population increases, this causes a decrease in the tuna population. Students may then diagram and align the example with the previous. Here, students are generating the shared structure that defines negative feedback loops. Students may also be prompted to compare and contrast the examples and identify similarities, which further focuses attention on shared structure and increases germane load.

Equilibrium points

Equilibrium is an important concept in the LS 30 series. An equilibrium point denotes a point in time at which a biological system does not change. For example, in a simple population model, the birth rate of rabbits may be completely cancelled out by an equivalent death rate so that the population overall remains steady. Two kinds of equilibrium points exist: those that are stable and those that are unstable. To teach this concept to students, LS 30 instructors use a concrete source analog that grounds the abstract concept of equilibrium in something familiar and visualizable. Stable equilibrium points and their behavior are analogous to a ball in a cup (Figure 1.3, left panel): if a ball at the bottom of a cup is pushed slightly, it will return to its resting place at the bottom of the cup. Similarly, if a biological system is perturbed away from a stable equilibrium point, the system will return to the equilibrium point. Unstable equilibrium points and their behavior are analogous to a ball balanced on top of a hill (Figure 1.3, right panel): if the ball is pushed slightly, it will roll away from its point of stability and never return. Similarly, if a biological system is perturbed away from an unstable equilibrium point, the system will move to a different equilibrium point or on a trajectory of infinite growth.

The analogical approach to teaching equilibrium points makes use of students' prior knowledge. To ensure that students use their prior knowledge effectively, instructors should be sure to fully explain the source analog and the correspondence of the target concept to the source (e.g., that the ball corresponds to the state of the system at a time point). In addition, instructors should describe the limits of the analogy and note which aspects of the source analog do not map onto the target concept. Explaining the limits of the analogy should keep students from drawing incorrect inferences about the target concept from the source domain.

The Current Project

The current project aimed to investigate the efficacy of an analogical approach to STEM education. The goal of Study 1 was to explore the effectiveness of the analogical approach in an online lab setting. In this study, undergraduate participants recruited from the UCLA subject pool watched lecture videos on topics drawn from LS 30A. Lecture videos that are designed using the analogical approach were compared to control videos that removed the key elements of the analogical approach. Learning from each of these videos was assessed with an immediate posttest.

Study 2 investigated one of the proposed mechanisms thought to underlie an analogical approach: increased attention to relational information. In this study, participants recruited from the UCLA subject pool learned how to translate verbal statements into mathematical expressions from lecture videos designed using an analogical approach, a carefully matched symbolic control, or the gold standard instruction for this topic. Of these three methods, the analogical approach is unique in that it explicitly represents the relational information that is crucial for problem solving. Learning and transfer were assessed on an immediate posttest.

Study 3 examined the analogical approach in a classroom setting. Two instructors teaching two different sections of LS 30A in Winter 2021 were recruited and their students participated in this study. The study compared the efficacy of analogical instruction that maximized germane cognitive load to analogical instruction that reduced the germane cognitive load associated with analogical processing through external scaffolds (instructor guidance and structured student worksheets). To minimize the variability due to student and instructor differences, Study 3 utilized a within-subjects crossover design in which students in both classes received both types of instruction throughout the quarter. The efficacy of the instructional materials for students at the low end of the course grade distribution were of particular interest in this study.

The classroom study was designed to assess the implementation of analogical instruction in an analogy-heavy course, but the implementation of the study deviated slightly from the original vision. The Covid-19 pandemic and ensuing changes to the structure of the course necessitated changes to the materials and procedure, which will be explained further in the Method section.

Chapter 2: Study 1. The potential benefit of an analogical approach to STEM learning in an online lab setting

The goal of Study 1 was to explore the possible benefit of an analogical approach to instruction in an online research setting. Participants watched video lessons on one of two topics drawn from LS 30A (feedback loops and functions). Learning from videos containing instructional techniques from analogical reasoning literature was compared with carefully matched control videos.

Method

Participants

Participants were 311 undergraduate students (M_{age} = 21.0, 236 female, 50 male, 3 nonbinary, 22 did not respond) recruited from the UCLA psychology subject pool. Participants received course credit in exchange for their participation.

Design and Procedure

Instructional style (analogical vs control) was manipulated between subjects. Participants were randomly assigned to view analogical (*N* = 166) or control materials (*N* = 145) for one of two topics, feedback loops (*N* = 162) or functions (*N* = 149). Participants first took a pretest followed by video lessons that contained embedded questions. Participants were instructed to treat the video lessons as they would a recorded lecture in an online class because they would answer some questions about the lessons at the end of the study. Participants were able to pause, rewind, and speed up the video lessons. After the lessons concluded, participants took a posttest and filled out a demographic questionnaire and an end-of-study survey. The study was conducted via Qualtrics Research Platform and it took approximately 30-45 minutes to complete.

Materials

Functions

Instructional materials

Instructional materials for each of the four conditions consisted of a series of lecture videos and embedded practice problems. Lecture videos were recorded using the iOS screen recording feature. An image of the instructor was superimposed over the screen (see Figure 2.1). Participants were able to see the instructor and the screen while the instructor wrote on an iPad.

An experienced LS 30A instructor provided learning objectives that students should be able to meet after learning about functions. Students should be able to (1) define and identify a function; (2) define and identify a function's domain and codomain; and (3) define and identify an instance of function composition. Based on these objectives, video lessons were recorded for each instructional style.

The lessons were comparable in length: the analogical lesson lasted 17:30 minutes and the control lesson lasted 14:22 minutes. Each lesson was broken into six segments separated by embedded practice questions. The embedded questions were designed as attention checks to ensure that participants were paying attention to the lessons. The embedded questions required straightforward recognition of the concepts in a new real-world example, and feedback was provided after each question. Four of the six embedded questions were the same in each condition. The two questions unique to the analogical instructional condition asked participants to identify corresponding elements of a function across different examples, and the corresponding questions in the control condition asked participants to identify elements of a function within a single example.

Each video lesson followed the same base script and the major topics were delivered in the same order. The lessons began with a real-world example of a function, before giving the definition and explaining how the example fit the definition of the concept. Then, the instructor explained a very common analogy used to teach functions: that of a machine that takes inputs and assigns each one to an output. The instructor then defined a function's domain and codomain and identified these concepts in the real-world example. Next, the instructor presented additional examples that were symbolic and graphical in nature. The lessons ended with a definition and two examples of function composition.

In the analogical instruction condition, the script was adapted to emphasize analogical learning techniques. Throughout the lesson, the instructor highlighted shared structure using color coding, spatial alignment, and comparing/contrasting. The instructor used a consistent color coding scheme to denote the correspondence between elements of the various examples (e.g., all inputs were written in red and all outputs were written in blue). In addition, corresponding examples were aligned spatially when possible. The instructor also explicitly compared and contrasted the examples and pointed out similarities throughout. After explaining the machine analogy, the instructor pointed out the limits of the analogy (aiming to keep participants from developing misconceptions). Finally, examples were presented simultaneously to reduce the cognitive load imposed by the comparison process.

The control instruction condition used the same examples as the analogical condition, but the instructor did not use color coding, spatial alignment, or explicit comparing/contrasting. The instructor did not explain the limit of the machine analogy, and each example was discussed and presented sequentially.

Assessment

To assess participants' knowledge of functions, a set of seven questions were selected from the Function Concept Inventory (FCI; O'Shea, Breen, & Jaworski, 2016). The FCI is designed to assess students' understanding of several key properties of the function concept (e.g., understanding the difference between functions and equations, recognizing and relating different representations of functions). These questions were supplemented with five questions written by an experienced LS 30A instructor to assess the learning objectives identified as important at the outset of the study. Two versions of each question were created and randomly assigned to set A or set B. Assignment of question set to pretest and posttest was counterbalanced across participants. The experienced instructor ranked the questions according to how well they assessed the learning objectives and questions were weighted accordingly in scoring. Partial credit was possible on some questions. Scores on each question were summed and standardized to yield a total proportion correct. After completing the posttest, participants completed a transfer test in which they generated a novel example of a function that they did not see in the lesson and provided its domain and codomain. Scores on this question ranged from 0-3.

Feedback loops

Instructional materials

An experienced LS 30A instructor provided learning objectives that students should be able to meet after learning about feedback loops. Students should be able to (1) define and identify positive and negative feedback loops; (2) modify a feedback loop so that it is the other type of loop. Based on these objectives, video lessons were recorded for each instructional condition.

The lessons for each condition were comparable in length: the analogical lesson lasted 17:52 minutes and the control lesson lasted 14:36 minutes. Similarly to the function topic, the lessons were broken into seven segments separated by embedded practice questions to ensure that participants were attending to the lessons. In the analogical condition, the questions required alignment of different examples. In the control condition, the questions required recognition of the concepts within a single example. Feedback was provided after each question.

Each video lesson was created from a single base script and covered the same topics. Both lessons contained two examples of negative feedback loops, two examples of positive feedback loops, and abstract representations of each type. The examples came from the domains of everyday experience or ecology. In order to emphasize analogy, the topics were rearranged in the analogical condition.

The analogical condition began with a real-world example of a negative feedback loop and provided a definition. The instructor diagrammed the example (see Figure 1.1 and Figure 2.1) and explained how it fit the definition of a negative feedback loop. Then, the instructor presented a second example in the same manner. Next, the instructor explicitly pointed out the corresponding elements of the examples before presenting an abstract representation of a negative feedback loop, diagramming it, and mapping it to both examples. The instruction for positive feedback loops followed the same pattern (i.e., example-example-abstraction).

Similarly to the functions topic, the analogical instruction condition highlighted shared structure using color coding, spatial alignment, and explicit comparing/contrasting. Examples from each topic were displayed simultaneously to reduce the load of the comparison process.

In contrast, the control condition began with the definition of a negative feedback loop and the abstract causal diagram. Then, the instructor sequentially presented two real-world examples, diagrammed them, and explained how each fit the definition. The instruction for positive feedback loops followed the same pattern (i.e., abstraction-example-example). The instructor did not use color coding and examples were presented sequentially without explicit prompts to compare.

Assessment

The assessment consisted of 20 short scenarios describing phenomena in one of four domains (everyday experience, ecology, physiology, and economics). For each of the four domains, the short scenarios exemplified one of five different causal-relational categories (positive feedback, negative feedback, causal chain, common cause, common effect). Twelve of the scenarios were drawn from the Ambiguous Sorting Task (Rottman, Gentner, & Goldwater, 2012) and the remaining eight were written by the experimenter. For each scenario, participants were tasked with selecting a causal-relational diagram that best depicted the scenario (see Figure 2.2).

The examples that participants saw in the lessons were drawn from the domains of everyday experience and ecology, while physiology and economics were not used in the lessons and were thus untrained domains for the participants. Pretests and posttests were created for each participant by randomly assigning one familiar domain and one unfamiliar domain to the pretest and the remaining domains to the posttest. Scores on the positive and negative feedback loop questions were coded as correct or incorrect and summed. Scores ranged from 0-4.

The transfer test required participants to extend knowledge of two-variable feedback loops they had been seen in the lessons to more complex feedback loops involving three variables. Four three-variable scenarios were written, one of each type for each familiar domain. Participants read each scenario and selected which causal-relational diagram best fit the scenario. Performance was coded as correct or incorrect. Participants were also asked to generate an example of each type of feedback loop that they did not see in the lesson. These responses were coded by independent coders on a scale from 0-3.

Demographic Survey and End of Study Questions

Participants entered their age, gender identity, race, major, and college GPA. They also reported the highest level of mathematics that they had taken in high school and math grades received, and whether or not they had taken AP Biology and AP Environmental Science in high school.

A supplemental goal of this study was to investigate how participants engaged with the study materials in an online setting. Participants in the analogical instructional conditions were informed that the video lessons contained some strategies designed to help them learn. Then, the strategies were listed (color coding, explicit compare/contrast, simultaneous presentation of examples). The participants were asked whether or not they found the strategies helpful and to explain why or why not. All participants then reported how they engaged with the lecture videos, indicating whether or not they paused/rewound the videos, the speed they watched the videos at, and use of scratch paper. Finally, participants reported if they took the study seriously or if they had difficulty paying attention.

Data Analysis

Coding

Two trained coders coded participants' responses to the open-ended questions. All disagreements were resolved by discussion among the coders. The two generation questions for the feedback loops topic were coded similarly. Two points were awarded if participants' examples illustrated a full loop in which the second quantity clearly affected the first quantity, and one additional point for generating the right kind of feedback loop.

Reliability was assessed using weighted Cohen's Kappa, a variant of Cohen's Kappa designed for use with ordinal variables where closeness of agreement or disagreement should be taken into consideration. Inter-rater reliability for the three feedback loop questions (generating a positive feedback loop, generating a negative feedback loop, and altering a feedback loop) was moderately high, indicating good agreement beyond chance (all Cohen's *k* between 0.60-0.81).

For the function topic, coders awarded one point for a clear description of the function's domain, one point for a clear description of the function's codomain, and one point for a clear description of the mapping from domain to codomain. Inter-rater reliability for the components of the functions question was lower but still indicated acceptable agreement beyond chance (all Cohen's *k* between 0.52-0.60).

Results

Demographic information and video engagement

Demographic information and video engagement was similar across topics. Generally, the sample was high-achieving academically; the average self-reported GPA was 3.71. In addition, pre-exposure to the topics covered in this study was high. Across both topics, 75% of the sample were identified as STEM majors. Further, the curriculum of AP Biology and AP Environmental Science covers feedback loops and nearly half of the sample in the feedback loops conditions (46%) had taken one or both of these classes in high school. Among participants in the functions conditions, 70% of the participants had taken calculus in high school and 72% reported getting mostly A's in their high school math classes.

Across both topics, approximately half of the participants watched the video lessons at normal speed (47%) and a quarter watched the lessons at maximum speed, which was twice the normal speed. Only 8% of the participants took notes on the lessons and 35% of participants reported pausing or rewinding the videos. Participant engagement was moderate and 75% of participants reported that they took the study seriously and were able to pay attention.

Perception of instructional strategies

Overall, participants reported that the strategies were helpful. Only 3% of participants across both analogical conditions reported that they were not helpful. Color coding was endorsed as helpful by the largest number of participants (77%) followed by leaving past examples on the board (65%). About half (48%) of the participants stated that explicitly identifying correspondences between the examples was helpful.

Feedback Loops

Of the 162 participants assigned to the feedback loop topic, 160 completed the pretest. The two participants that did not complete the pretest were dropped from the analysis. In general, participants performed well on the feedback loop pretest. Almost one third of the participants (n = 51) scored perfectly on the pretest, and the median score was 75%.

Selecting a measure to capture change from pretest to posttest involves weighing costs and benefits of various approaches. In the current project, normalized change scores were calculated for each participant using the pretest and posttest scores (Marx & Cummings, 2007). It is often subjectively more difficult to make progress at a high level of performance and normalized change scores reflect this reality. While not a perfect measure, normalized change scores allow the detection of change at this high level while avoiding a low pretest score bias. All scores are normalized and range from -1 to +1, regardless of performance on the pretest. As a result, normalized change scores are easier to interpret than other measures of learning gain. Normalized change scores (c) are calculated following Equation 1.

If the participant performs better on the posttest than the pretest, c is calculated as a gain score and represents the participant's actual improvement relative to the maximum possible improvement (Hake, 1998). If the participant scores perfectly on the pretest and posttest or if the participant receives 0 points on the pretest and posttest, they are dropped from the analysis. In these cases, the participant's performance is beyond the scope of the assessment. If the participant performs the same on the pretest and posttest, $c = 0$. Finally, if the participant performs better on the pretest than the posttest, c captures the participant's actual loss relative to the maximum possible loss.

After dropping participants according to the rules above and those that did not finish the experiment, there were 134 participants remaining for the normalized change analysis. Figure 2.3 shows the distribution of normalized change scores in each condition, which are markedly non-normal. Accordingly, nonparametric methods were used to test statistical significance. Bootstrapping procedures rely on the single assumption that the sample is representative of the population and do not make any assumptions about the underlying population. The sample is treated as a population and many "pseudo-samples" are repeatedly drawn from the original sample (Calmettes, Drummond, & Vowler, 2012). Then, observed sample statistics can be compared to the distribution of statistics computed from the pseudo-samples.

In both conditions the scores formed two clusters around 0 and 1.0. To investigate the subpopulations apparent in the dataset, the data were split based on normalized change scores. Participants were categorized based on whether or not they improved over the course of the study (i.e., whether $c > 0$). A Chi-Square test of independence was performed to assess the relationship between conditions and improvement over the course of the study. There was relationship between the variables, $X^2(1, N = 134) < .01$, $p = .99$. As a result, following comparisons were calculated after collapsing across conditions.

Across both conditions, 34% of participants ($N = 45$) achieved maximum possible improvement (i.e., $c = 1.0$). The second most common score was 0 (17.2%, $N = 23$), which indicates no change. Finally, 22 participants (16.4%) attained a score of -0.25, which indicates that these participants scored perfectly on the pretest and answered one question incorrectly on the posttest. Visual and numerical comparisons of these groups showed that they did not differ on self-reported GPA (*median* 3.80 vs 3.82), prior experience with the material in high school, time spent on the experiment, or performance on the transfer test.

Although the distributions were bimodal, they were similar in shape so the median was used to summarize the data. The median for each condition was 0. The observed difference in performance was calculated by subtracting median performance in the analogical condition from median performance in the control condition. Bootstrap/resampling methods were used to test the null hypothesis that there is no difference between the median of the analogical and control groups. Briefly, I conjoined the two data sets into a single set, and resampled (with replacement) "control" and "analogical" groups from this set. This was repeated 10,000 times, and the *p*-value of the result was calculated as the number of simulations producing results as extreme or more than the observed result, divided by 10,000. The observed difference between conditions was 0.0 with a 95% confidence interval (-0.50, 0.50). The bootstrap analysis showed that the conditions were not significantly different ($p = .56$).

A total of 155 participants completed both the pretest and transfer test and thus were subject to analysis. Figure 2.4 shows transfer scores for participants in each condition. Again, the scores are non-normal and as a result nonparametric methods were used to test statistical significance. In both conditions the median score was 11 out of 12 points possible.

The observed difference in performance was calculated by subtracting median performance in the control condition from median performance in the analogical condition, yielding an observed difference of 0. A similar bootstrap procedure was used to test the null hypothesis that there is no difference in transfer performance between the analogical and control groups. The observed difference between conditions was 0 with a 95% confidence interval (-1.0, 1.5). The bootstrap analysis showed that the conditions were not significantly different (p = .36).

Functions

Compared to the feedback loops topic, performance on the functions pretest was lower and had greater variability. Performance is reported as proportion correct. The median score was .49 and the inter-quartile range was .32. Of the 149 participants, 10 did not complete the posttest and so were dropped from the analysis. As in the feedback loops conditions, a normalized change score was calculated for each participant.

Figure 2.5 shows the distribution of normalized change scores in each condition. The distribution of scores in each condition appeared comparable and the median scores did not differ greatly (0.18 for participants in the analogical condition compared to 0.22 in the control condition).

The observed difference in performance was calculated by subtracting median performance in the control condition from median performance in the analogical condition. A similar bootstrap procedure was used to test the null hypothesis that there is no difference in transfer performance between the analogical and control conditions. The observed difference between conditions was -0.04 with a 95% confidence interval (-0.07, 0.19). The analysis showed no significant difference in average performance (p = .27).

Next, scores on the function generation question were analyzed. Fourteen participants did not complete the function generation question and were dropped from the analysis. Figure 2.6 shows that performance on the generation question was very good and 67% of participants scored perfectly. The observed difference between condition medians was 0.0 with negligible variability, resulting in a 95% confidence interval (0, 0). The analysis showed no significant difference in average performance (p = .99).

Study 1 Discussion

The aim of this study was to investigate the potential benefit of an analogical approach to STEM education in an online research setting. Instruction designed using principles from analogical reasoning was compared to control instruction. Specifically, the experimental videos used color coding, spatial alignment, and explicit comparing/contrasting to highlight shared structure. In addition, old examples remained on screen to lessen the cognitive load of comparison. These elements were removed from the control videos. Performance was assessed on an immediate posttest followed by a transfer test.

Across both topics covered by the study, learning outcomes were remarkably similar across instructional conditions. No difference in posttest performance was observed in performance or in transfer performance. While data on attention and motivation was not analyzed systematically, a significant proportion of participants reported low engagement with the lessons and this may dampen the efficacy of any instruction. In fact, a significant proportion of the sample showed no improvement at all or a slight decrease in performance. While these participants did not differ in any way detectable by the current study, it is possible that fatigue or carelessness may account for whether or not participants improved in the study.

In general, the sample was high-performing academically and had high levels of prior experience with the topics covered in the study. Previous research suggests that instructional interventions have reduced (and sometimes no) impact on students with high cognitive ability (e.g., Nesbit & Adesope, 2006) and intelligence and prior academic achievement are significant predictors of achievement in higher education. For example, one meta-analysis ranked 105 predictors by effect size and found that High School GPA and previous academic achievement as measured by admissions tests both ranked in the top ten, over variables like a teacher's enthusiasm for the subject and teaching (ranked 21) and instructional tasks that are designed to increase students' conceptual understanding (ranked 30; Schneider & Preckel, 2017). Given the overall high academic achievement of UCLA students, the present study is in line with these findings.

A great majority of participants reported that color coding and leaving the old examples visible aided their learning, while only half of participants reported that the explicit comparison was helpful. Previous research shows that students' beliefs about effective learning techniques are not always accurate (Bjork, Dunlosky, & Kornell, 2013). Further, effectiveness of instructional style does not solely determine efficacy. Students are not passive receivers of instruction; their perception of the instructional style and teacher behaviors mediate instructional efficacy (Weinstein, 1983).

Although Study 1 did not find significant effects of instructional method, conclusions about efficacy are limited by the time frame of the learning and assessment. Many researchers have drawn a distinction between immediate performance and long-term learning (Soderstrom &

Bjork, 2015). Future work should investigate whether or not these instructional manipulations confer differential longer-term benefits.

Chapter 3: Study 2. The potential benefit of an analogical approach to solving equation construction problems in an online lab setting

The goal of Study 2 was to investigate whether increased attention to relational information improves instruction on a specific topic in mathematics. Participants learned to translate verbal descriptions of proportional relationships into mathematical expressions. For example, the statement, "There are six times as many students as professors at this university" can be translated into the symbolic expression $S = 6P$ (Martin & Bassok, 2005; Simon & Hayes, 1976). These problems are notoriously difficult, with error rates ranging between 20% and 60%, even for students with considerable experience in mathematics (Christianson, Mestre, & Luke, 2012). Translating a verbal description of a situation into a mathematical representation is a key topic in LS 30A, and many weaker students have difficulty with this topic on the final exam despite the fact that it is initially introduced during the first week of instruction. The most common error on equation construction problems is to reverse the relationship between the quantities in the problem (i.e., $6S = P$), and success requires students to accurately represent and reason about the relation between the two quantities. Visual representations likely facilitate problem solving because they allow students to see relationships contained in symbolic or verbal forms represented explicitly (Larkin & Simon, 1987).

In this study, participants watched one of three lecture videos aimed at teaching strategies to solve equation construction problems. One video highlighted relations among quantities in the word problem using a geometric representation. Learning from this video was compared to a carefully-matched purely symbolic lesson that contained the same procedural steps as the experimental condition, but without explicit reference to relations, and a third condition that constitutes the current gold standard instructional method for this topic.

Method

Participants

Participants were 399 undergraduate students (M_{age} = 20.4, 264 female, 108 male, 3 nonbinary, and 24 did not respond) recruited from the UCLA subject pool. All participants received course credit in exchange for their participation.

Design and Procedure

The experiment was conducted entirely online using the Qualtrics survey platform. Participants completed the study on their own time and it took approximately 30 minutes to complete. In order to be included in the study, some prior knowledge of basic mathematical concepts was necessary. A five-question multiple choice test assessed knowledge of graphing points in a cartesian system, slope, and equations of lines. Each question was presented on its own page and participants were not permitted to go backward to questions they had already answered. Participants that answered three or fewer questions correctly (N = 51) were excluded from the study.

Participants who met the inclusionary criteria moved on to take a pretest consisting of five equation-construction problems, and were then randomly assigned to one of three instructional conditions (geometric, symbolic control, JUMP Math control). No feedback was provided on the pretest. The experimental condition highlighted the relationship between quantities in equation-construction problems using an explicit geometric referent. The first comparison condition replicated the procedural steps of the geometric condition but was purely symbolic in nature. The final comparison condition adapted lessons on equation construction from JUMP Math, which is considered by many mathematics educators to be a gold standard in mathematics instruction (e.g., Solomon et al., 2019).

In order to increase the similarity of the experiment to conditions of online learning, participants were allowed to watch the video lessons at their own pace. Participants were free to pause, rewind, and speed up the videos as they were watching. Each video lesson was cut into eight segments, with each segment presented on its own page of the survey. Once participants advanced past a segment, they were not permitted to go back and review the segment. Each lesson also contained a set of embedded questions to ensure participants were following along with the videos and two full practice problems. All questions embedded in the video included feedback.

After the video lessons, participants took a posttest with both equation construction problems and transfer problems, filled out the demographic questionnaire, and answered some debriefing questions about how they engaged with the experimental materials. Participants reported whether or not they paused or rewound the videos, note-taking behavior, and the speed at which they watched the videos. Finally, participants reported whether or not they used the strategy they learned in the videos.

Materials

Instructional materials

Instructional materials for each condition consisted of a lecture video, embedded questions, and practice problems. In each video, an instructor taught participants how to solve equation construction problems. Participants watched as the instructor wrote on a white board and solved two example problems. The same example problems were solved in each lesson. In the geometric and symbolic lessons, the instructor began with an easy example problem, then generalized the procedure, and prompted participants to solve a similar practice problem. After participants received feedback on the practice problem, the video lessons resumed and the instructor solved a more difficult example problem. Participants then solved a more difficult practice problem. The JUMP Math lessons began with a discussion of ratios before the instructor solved the easy example problem. Participants were then prompted to solve an easy practice problem before moving on to watch the instructor solve a more difficult example problem. Participants then solved the difficult practice problem. All lessons concluded with an explanation of the concept of proportionality. The geometric lesson videos totaled 18.25 minutes and the symbolic comparison was comparable at 16 minutes. The JUMP Math lesson videos were 13.5 minutes long.

Geometric lesson

The geometric lessons taught participants a strategy to represent the relationship between two quantities explicitly as the equation of a line. The strategy consists of four basic steps. The first step is to draw a set of axes to represent the quantities in the problem. For example, one problem reads, “For every banana, there are three apples. Write an equation for the number of apples.” Since the problem asks for the number of apples in terms of bananas, “number of apples” goes on the vertical axis and “number of bananas” goes on the horizontal axis (see Figure 3.1a). Second, participants are instructed to plot points on the axes that correspond to the number of bananas and apples. According to the problem text, when there is one banana, there are three apples and this corresponds to the point (1, 3) on the banana-apple axes (see Figure 3.1b). In order to draw a line capturing the relationship, one more point is necessary. The problem establishes a proportional relationship, so when there are 0 bananas there will also be 0 apples. This corresponds to the point (0, 0). After plotting the points, the next step is to connect the points with a line and find its slope. The line visually captures the relation between the number of bananas and the number of apples, which draws attention to the relational information that is crucial to solving these problems. The slope of the line is the proportionality constant that captures specifically how the number of apples varies as the number of bananas increases. Since the line goes through the origin, the line must be of the form $y = k * x$. Finally, participants are taught how to calculate the slope of the line (k) (see Figure 3.1c). Putting everything together, the equation for the line mathematically representing the relationship between the number of apples and the number of bananas is: $\text{number of apples} = 3 * number \text{ of bananas}$. The geometric representation used to find the equation for the line focuses attention on the critical relational information: the rate of change between bananas and apples.

In sum, this approach uses a physical geometric representation to reify the relation between the quantities in the problem. This lesson highlights the relations that participants must see and use to guide the translation of words into symbols. Representing the relationship visually also reduces the extraneous cognitive load imposed on the learner, as the quantities and their relationship do not need to be held in mind and manipulated.

Symbolic control lesson

The first comparison condition consisted of the same exact steps as those used in the geometric lesson, but without the explicit representation of relational information. This lesson taught a strategy with procedural steps that mirrored the geometric lesson: participants learned to set up an equation to find a quotient (q) that captures the number of apples per one banana. Then, participants learned how to use q to find the final expression (see Figure 3.2).

JUMP Math lesson

JUMP Math is a non-profit organization that prepares resources for teachers covering grades 1 to 8. Their lessons follow an evidence-based "guided discovery" approach that encourages students to work through challenges on their own with feedback and scaffolding from the instructor when necessary ("Research Supporting JUMP Math | JUMP Math," n.d.). JUMP Math outperformed classic instruction styles in a series of randomized controlled trials, improving students' math performance and attitudes (Solomon et al., 2019), and distinguished mathematics instructors consider this program to be the current gold standard in mathematics instruction ("Research Supporting JUMP Math | JUMP Math," n.d.).

The lesson was adapted from JUMP Math's Teacher Resource guide on equations, ratio problems, and the constant of proportionality ("Teacher Resource for Grade 8 - New US Edition | JUMP Math," n.d.). The lesson first taught participants how to use ratio tables to express ratios between two quantities (see Figure 3.3a), and then walked participants through the process of writing an equation from a ratio table in an example problem (see Figure 3.3b). Notably, the ratio table contains the same information (i.e., ordered pairs) captured by the line in the geometric lesson, but the relationship is not explicitly represented.

Pretest and posttest

Equation-construction problems require the translation of verbal information describing a multiplicative (i.e., proportional) relationship into a symbolic mathematical expression. Proportional problems describe scenarios in which two quantities scale with one another by a constant rate. Ten equation-construction problems were written for use in the study and randomly divided into two sets for use in the pretest and posttest. Assignment of problems to pretest and posttest was counterbalanced across participants. Each problem was scored as correct or incorrect to and scores were summed to yield a total out of 5. Reversal errors, in which participants reverse the quantities in the expression, were also counted.

In addition to the five-equation construction problems, participants completed a transfer test that consisted of six problems that assessed understanding of proportionality. Each problem was scored as correct or incorrect and scores were summed to yield a total out of 6.

Results

Demographic information and video engagement

As in Study 1, the sample was high-achieving academically; the average self-reported GPA was 3.65, 68% of participants reported taking calculus in high school and 68% reported getting mostly A's in their high school math classes.

Slightly over one-third of the participants (37%) reported watching the videos at normal speed and 31% watched the lessons at maximum speed, which was twice the normal speed. 22% of the participants took notes on the lessons and 26% of participants reported pausing or rewinding the videos.

Across the three instructional conditions, participants reported using the instructional strategies at seemingly varying rates (see Table 2). A Chi-Square test of independence was performed to assess the relationship between self-reported strategy use and instructional condition revealed a significant relationship between the variables, $X^2(4, N = 333) = 19.53, p < .001$.

Analyses

In general, participants performed well on the pretest (34.8% of participants scored four or higher and 21.9% received a perfect score). As in Study 1, a normalized change score was calculated for each participant to assess the effect of each instructional method on improving performance on equation construction problems (Marx & Cummings, 2007).

Figure 3.4 shows that participants in the geometric and symbolic conditions (*median* = 0.75 and 0.80 change scores, respectively) performed better than the participants in the JUMP condition (*median* = 0.33). In addition to the difference in average performance, the distribution of normalized change scores in the JUMP condition appears bimodal: participants are clustered around scores of 0 and 1.0.

To compare normalized change scores between conditions, an *F*-like statistic was computed. Roughly, to capture within-group variation, I computed the difference between each observation and its respective condition median and compared this to the difference between each observation and the overall grand median. Bootstrap/resampling methods were used to test the null hypothesis that there is no difference between the medians of three conditions. Briefly, I conjoined the three conditions into a single set, and resampled (with replacement) "geometric", "symbolic control" and "JUMP Math" groups from this set. From each of these pseudosamples, I computed resampled condition medians and an overall resampled grand median. I then computed a resampled *F*-like statistic. This procedure was repeated 10,000 times, and the *p*-value of the result was calculated as the number of simulations producing results as extreme or more than the observed result, divided by 10,000. The observed *F*-like statistic was 0.50 and the bootstrap analysis showed that the conditions were significantly different ($p = .002$).

To investigate which conditions were significantly different from one another, individual condition medians were compared using resampling methods. The Holm-Bonferroni method was used to control family-wise error rate (Holm, 1979). The analyses showed that the observed difference between the geometric and symbolic control conditions was not statistically different from 0 (*observed difference* = -0.05, p = .72). However, the geometric condition showed greater improvement than the JUMP Math condition (*observed difference* = 0.42, p = .04), as did the symbolic control compared to the JUMP Math condition (*observed difference* = 0.47, p = .05). In sum, the JUMP Math condition led to less improvement in performance on equation construction problems than did the other two conditions.

Next, transfer performance was examined in the full dataset (i.e., including participants who scored perfectly on the pretest and posttest and participants who scored 0 on the pretest and posttest). Figure 3.5 shows transfer performance for each condition. To compare transfer performance among conditions, an F-like statistic was computed from the dataset. A similar bootstrap method was used to test the null hypothesis that there is no difference between the mean of the three conditions. The observed F-like statistic was 0.01 and the bootstrap analysis showed that the conditions were not significantly different (p = .26).

Finally, error patterns were examined. Reversal errors occur when participants reverse the relationship between the quantities in their response. Reversal scores were computed for each participant on the pretest and posttest by subtracting the number of reversal errors from 5 (the maximum score on the pretest and posttest and thus the maximum number of reversal errors possible). This measure captures how many reversal errors a participant makes, and does not include other types of errors. Higher scores correspond to fewer reversal errors. A reversal change score was computed similarly to the normalized change score. The reversal change score is a measure of actual improvement in reversal errors relative to possible improvement. Numerically, the symbolic control condition (*median* = 1.00) performed better than both the geometric (*median* = .667) and JUMP Math condition (*median* = .333).

To compare performance among conditions, an *F*-like statistic was computed. A similar bootstrap method was used to test the null hypothesis that there is no difference between the median of the three conditions. The observed *F*-like statistic was 0.65 and the bootstrap analysis showed that the conditions were significantly different (p = .006).

Individual conditions were again compared using resampling methods and the Holm-Bonferroni method to control family-wise error rate (Holm, 1979). Although the omnibus test indicated a difference among conditions, none of the pairwise comparisons reached significance after correcting for multiple comparisons. The observed difference between the geometric and symbolic control conditions was not statistically different from 0 (*observed difference* = -0.333, p = .72). The observed difference between the geometric and JUMP Math conditions was not statistically significant (*observed difference* = 0.333, p =.43), and neither was the difference between the symbolic control and the JUMP Math condition (*observed difference* = .667, p = .11).

Study 2 Discussion

The aim of this study was to investigate the potential role of attention to relational information in mathematics. Participants watched one of three lecture videos that taught strategies to solve equation construction problems. Lessons using a relational geometric representation were compared to a carefully-matched symbolic control lesson and a lesson adapted from high-quality standard instruction for this topic. Both the geometric lesson and the symbolic control lesson outperformed standard instruction. An omnibus test revealed that lessons also differed in the effectiveness of reducing reversal errors, although none of the pairwise comparisons reached significance after correcting for multiple comparisons. No difference was observed in transfer performance. In addition, a Chi-square analysis suggested that use of strategies differed among conditions.

Gains in performance on equation construction problems were observed most prominently in the geometric and symbolic conditions. These instructional methods represent novel strategies to improve performance on equation construction problems. Previous attempts to improve performance have investigated the impact of changing word order to facilitate problem solving (Cohen & Kanim, 2005; MacGregor & Stacey, 1993) or training students to construct accurate situation models (Reed, 1987; Weaver & Kintsch, 1992). One study found that repeated practice alone improved performance for non-math majors (Christianson et al., 2012), but the intervention was highly procedural in nature and it is unclear the extent to which such instruction improved conceptual understanding of equations and proportionality.

Overall accuracy and improvement in reversal errors was indistinguishable between the geometric and symbolic instruction, and both appeared to be more effective than JUMP Math instruction. Although previous research suggests that visual referents may be particularly helpful for improving problem solving in mathematics (Larkin & Simon, 1989), it appears that the addition of a visual representation of the relational information in word problems did not aid students in the present study any more than a purely symbolic procedure.

One possible explanation for the efficacy of the purely symbolic procedural approach is that this approach to solving problems likely mirrors much of the mathematics instruction that students have already received. K-12 mathematics instruction in the US is highly procedural in nature (Richland, Stigler, & Holyoak, 2012). The Chi-square analysis suggests that more participants reported using the symbolic than the geometric and JUMP Math strategies, so it is possible that the familiarity of the strategy made it easier for participants to learn and use.

Further, students' beliefs about the efficacy of instruction mediate their learning from it (Weinstein, 1983). If students expect mathematics to be a series of procedural steps (as many students do; Stigler, Givvin, & Thomson, 2010) and the lesson matches their expectations, they may be more willing to engage in the lesson and thus to benefit from it. Students are less willing to engage in strategies that they perceive as effortful (Biwer, de Bruin, Schreurs, & oude Egbrink, 2020); to the extent that a strategy was perceived as effortful, it is less likely that students would try to adopt it.

Despite the prior research support it has received (e.g., Solomon et al., 2019), the JUMP Math condition reliably produced less improvement than the other conditions. One possibility is that this approach is less amenable to adaptation to an asynchronous online learning environment. JUMP Math lessons are built around considerable cooperation with fellow students and feedback from an instructor. These elements were absent from the current study, which may have damaged the fidelity of the JUMP Math instruction. That online instruction is different than face-to-face instruction is a truism, and efficacious instruction looks different across these different mediums. While JUMP Math has been shown to be efficacious in live classroom settings, it is possible that such strategies are less suitable for adaptation to asynchronous lecture videos.

The lack of difference in transfer performance is consistent with previous findings indicating that transfer performance is difficult to elicit (Renkl, Mandl, & Gruber, 1996). While each instructional condition contained a brief description of the concept of proportionality, the bulk of each lesson was devoted to explicit instruction in solving equation construction problems. Transfer performance is more likely to be improved by lessons that present several different problems that all embody the proportionality concept (Richland et al., 2012).

Study 2 suffered similar limitations as those noted in Study 1. Specifically, the study may have underestimated the efficacy of these instructional strategies because instructional interventions often have less impact on students with high prior achievement (Schneider & Preckel, 2017). The relational information that was explicitly represented in the geometric instruction was contained in each of the other conditions, though it was less obvious. It is possible that sufficiently high-performing students do not need the extra representational support to aid them in problem solving.

Chapter 4: Investigating an analogical approach in a classroom setting

The goal of Study 3 was to investigate how to implement analogical instruction in a naturalistic classroom setting, and specifically, how analogical processing may be utilized to increase germane cognitive load. There are two general strategies for adopting analogical instruction, which vary in the amount of germane cognitive load imposed on the learner. Student-centered interventions use strategies that ask students to take charge of their own learning to focus on relevant information (e.g., Wright, 2011). In these interventions, the onus is on the learner to direct their attention to the aspects of the target material that is identified by the instructor as important. For example, students might be tasked with identifying similarities among analogous examples of a negative feedback loop. These strategies are likely to increase germane cognitive load to a great extent, but may overload the limited capacity of novice learners.

In contrast, teacher-directed interventions change the very structure of the student experience to restrict students from engaging in processing that is irrelevant to the target material. External scaffolds (e.g., teachers, worksheets) play a larger role in these interventions. For example, a teacher may guide students through a comparison of analogous examples of a negative feedback loop. These strategies are likely to increase germane cognitive load to a lesser extent than the student-centered strategies because students receive explicit guidance on the target material. However, these strategies are less likely to overload limited working memory capacity. The goal of this study was to contrast the effects of these two methods to incorporate analogical processing into a classroom environment and examine their effects on course performance. Unexpected methodological difficulties were encountered due to the Covid-19 pandemic. These will be detailed further in the method section.

The study took place in two sections of LS 30A taught by two different instructors during Winter 2021 at UCLA. LS 30A is an introductory-level five-credit-hour course with lecture and discussion section components. The class was run using a flipped-classroom design: each week, students in each section watched lecture videos recorded by one of the instructors before they met for synchronous live sessions. Live sessions were devoted to emphasis of particularly important or difficult concepts and short group activities. The class met for two 75-minute live lectures and one 110-minute live discussion section. Professors led the class-wide lectures and teaching assistants (TAs) led the discussion sections. All class meetings were conducted live via zoom. During discussion sections, TAs spent approximately 40 minutes reviewing and reinforcing lecture content. The instructional intervention was delivered via worksheets that students completed during the content review portion of discussion section. In the remaining discussion section time, TAs assisted students with the programming components of the course, which were not covered in lecture and were not subject to the instructional intervention.

Method

Participants

Participants were 214 undergraduate students enrolled across two sections of LS 30A in Winter 2021. Lecture 1, taught by Professor A, had 83 students and Lecture 2, taught by Professor B, had 131 students.

Students in each lecture section also enrolled in one discussion section. Enrollment for all six of the discussion sections associated with Lecture 1 ranged from 7-14 students. Five of the seven Lecture 2 discussion sections had 19 or more students (the two smaller sections had 12 and 15 students). Across both lectures, 2 TAs taught two sections each and the lead TA, who also contributed to curriculum development, taught one section. At the start of the quarter, students in each discussion section were randomly divided into Learning Teams that contained 3-4 students. For the entirety of the quarter, students worked with their Learning Teams on worksheets in discussion sections and collaborative portions of the exams.

Design and Procedure

The material covered in LS 30A is divided into topics, which roughly correspond to the first two chapters of the textbook. The material covered in Chapter 1 is largely independent of Chapter 2, so topics can be taught using different instructional methodologies without substantial contamination. The instructional manipulation was delivered using a crossover design. In a crossover design, course topics are divided into two sets. One of the lecture sections received student-centered (SC) instruction on topics assessed on the Midterm (all Chapter 1 material and early Chapter 2 material) and teacher-centered (TC) instruction for the topics assessed uniquely on the Final exam (later Chapter 2 material). The other lecture section received the reverse: TC instruction for Midterm topics and SC for Final exam topics. Use of a crossover design minimizes inter-instructor and inter-classroom variability, as each student will serve as their own control (Jones & Kenword, 1989). In this design, each student received both methods of instruction, though for a different set of topics.

The Midterm was administered during Week 6 of the 10-week quarter and the Final exam was administered during Finals Week. Discussion sections for Lecture 1 received TC instruction for the material learned before the Midterm and SC instruction for the material learned after the Midterm. Discussion sections for Lecture 2 received the reverse order (SC instruction for the material learned prior to the Midterm and TC instruction for material learned after the Midterm). Assignment of instructional condition to lecture section was randomized at the start of the quarter.

Materials

Development of materials

During discussion sections, TAs were instructed to spend the first 40 minutes reviewing lecture content and the remaining time working on programming. The instructional interventions were designed to reinforce lecture content and were delivered during discussion sections in the form of worksheets. For each week of material, one set of worksheets was created by the primary researcher to utilize analogy to increase germane cognitive load. These worksheets were then adapted by an experienced LS 30 instructor to increase external scaffolding and moderate the germane cognitive load imposed by the introduction of analogy.

Worksheets consisted of a series of problems and activities that students completed either individually, in Learning Teams, or as a class (i.e., led by the TA). Instructional materials were developed in a three-stage process by the primary researcher (a domain expert in analogy), the lead TA, and an additional domain expert. Both the lead TA and the domain expert have several years of training in biology, mathematical modeling, and are experienced educators. First, the lead TA provided me with a series of sample problems and activities administered in discussion sections during previous offerings of LS 30A. Using these problems as a starting point, I adapted the problems to emphasize the application of analogy to create the student-centered analogical worksheet. Then, the SC worksheet was reviewed by the lead TA and the domain expert and the problems were refined based on their recommendations. Finally, the domain expert adapted the SC worksheet to moderate the germane cognitive load imposed by analogy and increase scaffolding to create the TC worksheet.

In general, the SC worksheets were created following five principles of analogical instruction (Gray & Holyoak, 2021; see Table 1). Several strategies were used to increase attention to the structure of course concepts and the relationships between concepts (principle 2). For example, color coding schemes were developed to emphasize correspondences between different examples of the same concept (see Figure 1.1 and Figure 4.1) or across different representations of the same concept (see Figure 1.3). In addition, some activities utilized comparing and contrasting. In some cases, students were tasked with explicitly identifying similarities and differences among examples of the same concept. This served to focus attention on shared structure among isomorphic examples. In other cases, students compared and contrasted two different concepts following specific prompts that focused on the important dimensions of comparison. Many of these compare/contrast exercises were organized in the form of tables (see Figure 4.2). In these comparison exercises, students went row-by-row down the table, comparing and contrasting each element of the target concepts. In order to connect mathematical and verbal models (principle 3), some activities explicitly directed students to generate explanations for the use of a particular mathematical term (see Figure 4.3). Finally, to encourage the generation of inferences (principle 5) students were tasked with either coming up with their own examples of course concepts or aligning previously given examples. Importantly, each of these activities was student-directed. For example, students were tasked with applying color coding schemes to highlight role-based similarity in Learning Teams or individually.

There is a limit to how much instruction in LS 30A can be "de-analogized" because the course is based upon deep structural similarities across varying domains and inherently involves connecting representations of the same ideas (verbal descriptions, mathematical models, graphical representations). Many of the same analogical principles at work in the SC worksheets remained in the TC worksheets. In general, these worksheets used the same problems as the SC worksheets, but they were adapted to lessen germane load and increase scaffolding. Although these changes sometimes happened at the expense of analogical learning strategies and thus reduced potentially helpful germane cognitive load, it is possible that students require this additional guidance.

The worksheets were adapted by adding structure to the worksheets or explicit guidance from the instructor. For some worksheets, the domain expert rearranged the order of the problems from the SC worksheet to provide a logical flow so that the TC worksheet started with easier problems and moved to more difficult problems. This meant that the worksheet began with the basic components of the topic and built up complexity. Second, the expert identified the essential aspects of each topic and stripped down the SC worksheet to contain only the information most essential to the topic; irrelevant information was withheld completely. Third, the expert added more explicit scaffolding to the activities. Sometimes, this extra guidance was embedded within the activity itself (e.g., an activity that asked students to generate an entire idea or explanation in the SC worksheet was simplified to a fill-in-the-blank activity in the TC worksheet; see Figure 4.3). Other activities were adapted from independent group work to a TA-led class-wide activity in which the TA provided explicit guidance. In the TC worksheets, students were explicitly guided through the activities with the help of the worksheet itself or the instructor. To reduce the cognitive load of the compare/contrast exercises, students were guided through completing the tables one column at a time (see Figure 4.2). In this fashion, both targets of comparison were presented separately in full before any comparison took place.

TA training and delivery of instructional manipulation

While the lead TA was involved in material development and thus had knowledge of the study and its hypotheses, the remaining TAs were not informed about the study, the manipulation, or the hypotheses. While it is possible that background knowledge on the study's aims and methods could improve TAs' abilities to deliver the instructional materials, TAs were kept blind to the study so as to reduce any potential bias. The TAs were simply informed that each class would receive slightly different worksheets.

Each week, the lead TA led an hour-long meeting in which TAs received general guidance on the week's worksheets and advice on which activities to focus on in their sections. TAs from both classes received guidance on both worksheets during this training meeting and had access to both versions of the worksheets throughout the course. However, TAs were instructed to only use the worksheet written for their own class. After the TAs received this short training, they led their own sections. After leading the section, TAs reported which activities from the worksheets they covered in each section and how the activity was completed (i.e., as a class, in learning teams, individually).

Implementation data was used to assess how the instructional manipulations were delivered in the classroom. At the end of the course, activities that fewer than half of the TAs completed with their sections were considered to be "not delivered" to the students for the purposes of the study. Of 35 total activities across the 10-week quarter, 13 activities from the SC worksheets were not delivered and 10 were not delivered from the TC worksheets. In considering the materials that were actually delivered to students and thus reasonable to analyze, the conditions contained nearly identical problems that differed only in how they were delivered. The student-centered worksheets incorporated a great deal of analogical processing to focus attention on the structural information in the target material. These worksheets required students to direct their own attention to relevant information and thus increased germane cognitive load to a great extent. The teacher-centered worksheets also incorporated a great deal of analogical processing, but the activities were externally scaffolded to guide students' attention to important structural information. The TC worksheets involved less germane cognitive load.

Dependent measure: change scores

The primary assessment of student performance were individual exams. In this offering of LS 30A, exams were administered in three stages. In the first stage, students worked with their Learning Teams to create a study guide following specific guidelines and a rubric from the instructors. In the second stage, students took an individual exam on the course learning management system. Individual exams were open-book and unproctored but students were not allowed to work together. The Midterm lasted two hours and the Final lasted three hours. After all students completed the individual exam, students worked together in their learning teams to complete a collaborative exam. The key dependent measure of student performance was performance on the individual stage of each exam.

The instructors and lead TA wrote the exams and the same exams were given to each class. The Midterm was administered during Week 6 and the Final exam was administered during Finals Week. In this offering of LS 30A, the Midterm covered all Chapter 1 material and some early Chapter 2 material, but all topics on the Midterm were taught using the same instructional method within each class. The Final Exam was comprehensive and contained problems that assessed content taught before the Midterm (i.e., with one instructional method), problems that uniquely assessed content taught after the Midterm (i.e., with the other instructional method), and problems that combined content that was taught with both instructional methods. As a result, performance on topics taught exclusively after the Midterm was used to isolate the impact of the second instructional technique.

In order to control for inter-student differences, the key dependent measure was computed by subtracting each student's performance on the Final Exam from their performance on the Midterm. This yielded a change score for each student. A positive change score indicates that the student performed better on the Midterm than the Final and a negative change score indicates that the student performed better on the Final than the Midterm.

Results

Descriptive statistics

Experienced LS 30A instructors report that exam performance is typically negatively skewed and the same pattern was observed in the Winter 2021 offering of the course. Figure 4.4 shows student performance on each exam broken down by class and Table 3 shows the five number summary for each exam broken down by class.

Changes in overall performance and in unique, non-overlapping topics were computed for each student by subtracting Final Exam scores from Midterm scores. Changes in overall performance and topics unique to each instructional condition are plotted in Figure 4.5. In general, change scores are expected to be positive because the Final Exam is typically more difficult than the Midterm and scores decrease accordingly (Shevtsov, personal communication). Change scores for topics unique to each instructional condition are reported in Table 4. If TC instruction is more effective than SC instruction, change scores for Lecture 1 students should be larger (i.e., more positive) than change scores for Lecture 2 students because Lecture 1 students received TC instruction for the Midterm topics. If, on the other hand, SC instruction is more effective, the change scores for Lecture 2 students should be larger because Lecture 2 students received SC instruction for the Midterm topics. The median change score in non-overlapping topics was greater for Lecture 1 students than Lecture 2 students (6.53% vs 2.00%), suggesting a benefit for TC instruction.

Change score analyses

Examination of QQ plots for each class's change scores show that this measure is not normally distributed, so parametric analyses are not appropriate (see Figure 4.6). To assess the efficacy of SC and TC instruction, change scores from the two classes were compared. First, the observed difference in change scores between the two classes was computed by subtracting the median change score for Lecture 1 students from the median change score for Lecture 2 students. Since change scores reflect the change in performance from the Midterm to the Final, the observed difference value captures the relative performance of the classes. The observed difference was -4.5%, which suggests a benefit for TC instruction.

Bootstrap/resampling methods were used to test the null hypothesis that there is no difference between the median change score of the Lecture 1 and Lecture 2 groups. Briefly, I conjoined the two data sets into a single set, and resampled (with replacement) "Lecture 1" and "Lecture 2" groups from this set. This was repeated 10,000 times, and the *p*-value of the result was calculated as the number of simulations producing results as extreme or more than the observed result, divided by 10,000. The observed difference of -4.5% had a 95% confidence interval (-9.5%, -3.1%). Overall, the results of this analysis show a benefit for TC instruction (p = .03).

Analysis on low-performing students

One of the key aims of the current project was to improve learning outcomes for students at the low end of the performance distribution. First, low-performing students were defined as students that performed at or below the 25th percentile on the Midterm Exam for each class (Lecture 1 25th percentile = Lecture 2 25th percentile = 80%, $N_{Lecture\ 1}$ = 17, $N_{Lecture\ 2}$ = 33). Figure 4.7 shows performance for low-performing students on the Midterm and the Final exam. Among low-performing students, it appears that Lecture 1 students performed better on the Midterm (*median* = 75%) than Lecture 2 students (*median* = 73.8%) and that Lecture 1 students performed worse on the Final Exam (*median* = 69.5%) than Lecture 2 students (*median* = 72.3%). Once again, a change score in non-overlapping topics was computed for each student.

To assess the efficacy of SC and TC instruction among low-performing students, change scores from the two classes were compared in a manner similar to the change scores for the entire classes. First, the observed difference in change scores between the two classes was computed by subtracting the median change score for Lecture 1 students from the median change score for Lecture 2 students. The observed difference was -14.1% which suggests a benefit for TC instruction.

The bootstrapping procedure detailed above was rerun on the sample of low-performing students to determine whether the observed difference was large enough to be considered statistically significant. Again, the null hypothesis is that there is no difference in the efficacy of instructional methodology and, accordingly, no difference in change scores between classes.

10,000 pseudo-samples were generated and the difference in change scores between classes were stored. The observed difference was -14.1% with a 95% confidence interval of (-26.2%, -5.1%). The analysis showed that change scores were significantly different between conditions (p = .01), which shows a benefit for TC instruction for low-performing students.

Classroom Study Discussion

Study 3 used the framework of cognitive load theory to investigate the most efficacious implementation of analogical instruction in a naturalistic online classroom environment. Students enrolled in LS 30A during Winter 2021 learned topics through a more-structured teacher-directed approach to administering analogical interventions, or else through a less-structured student-directed method. Analyses showed that students benefitted from the structured teacher-centered approach. For low-performing students, the benefit for the teacher-directed approach was larger.

These findings are in line with the general picture that emerged from the two lab studies: among higher-performing individuals, instructional manipulations had a smaller effect on performance. Since low-performing students were of particular interest in the classroom study, the efficacy of the instructional intervention was examined separately in this subpopulation, and a larger benefit was observed. These students in particular benefitted from analogical instruction that was more highly structured and guided by an external source (a teacher or a worksheet).

A great deal of research supports the hypothesis that weaker students need significant guidance, but the present study provides a concrete demonstration of how much guidance is necessary. Proponents of learner-centered and constructivist approaches to education place the responsibility for learning on the learners themselves. These perspectives imply that students should actively work to construct their own knowledge in the classroom (e.g., Wright, 2011).

Though the educational literature does not always intersect with theoretical work from cognitive psychology, these approaches include activities that are presumably designed to increase germane cognitive load (e.g., problem-based learning). The current findings suggest that when considering how to best support low-performing students, educators need to be particularly aware of the "curse of expertise" and the cognitive load imposed by student-directed learning activities, because these students may need more guidance than educators typically assume.

Selecting relevant information and directing attention is an important skill for students to practice; an exam (much less the real world) is not scaffolded to facilitate problem solving. The present study suggests that low-performing students in particular need a great deal of support before they are able to exercise this skill successfully. Scaffolding aids initial learning for learners of various ages, and may be adapted to fit various stages of proficiency during the learning process (Vygotsky, 1978). Scaffolding may take many forms, including guiding analogical comparison. For example, a series of studies by Kurtz et al. (2001) showed that explicitly directing participants to jointly interpret two examples of a scientific concept, and to list correspondences between the examples, led to improved performance on an immediate test. Although the authors did not frame the materials in terms of cognitive load, the learning condition that produced the greatest performance gains imposed significant germane cognitive load, while offering a highly structured worksheet to guide the process so as to avoid overloading the learner. Other studies introducing varying amounts of structure to facilitate comparison show similar results: instruction that increases germane cognitive load through analogical comparison leads to better outcomes (e.g., Gentner & Loewenstein, & Thompson, 2003; Richland & McDonough, 2010). However, many of these experiments were performed on high-achieving undergraduate populations and within limited time frames. It is unclear how well these manipulations would generalize to lower-performing populations, and it is unclear to what extent germane cognitive load may need to be moderated by additional scaffolding.

The need for additional structure for novice learners has been documented by various researchers, and several instructional techniques explicitly instantiate this idea. For example, concreteness fading recognizes that students need a great deal of practice with concrete, visualizable materials before they are able to work successfully with abstract ideas (Fyfe, McNeil, Son, & Goldstone, 2014). These strategies progressively strip down concrete representations of concepts to their abstract essence. While concreteness fading may not be directly applicable to the interventions at issue in the current work, the process of gradually removing guidance and cues to direct attention to structure may be adapted to fit the analogical interventions tested in this study. The goal is ultimately the same: to build up students' abilities to direct their attention independently.

The current study suggests that caution is warranted in extending results of instructional efficacy to lower-performing students. The present findings may underestimate the extent to which instruction should be modified for lower-performing students, because low-performing students at UCLA are still high-performing on a nationwide scale. For example, the 25th percentile composite SAT score for students at UCLA, which is 1280, is still well above the national average of 1059 (College Board, 2020).

Curriculum changes due to the Covid-19 pandemic offered a unique opportunity to investigate the proposed interventions in an online setting. During discussion sections, student groups frequently worked on the worksheets while in individual Zoom breakout rooms. Unlike in a face-to-face classroom environment, the instructor cannot supervise all groups simultaneously, and must check in with each group in a sequential fashion if they choose to do so at all. As a result, it was more difficult for instructors to intervene to supplement instructions on the worksheets or offer real-time feedback as students worked in groups. More structure may lead to better outcomes in an environment in which immediate clarification and guidance is less available. It is unclear whether similar results would be obtained in a face-to-face setting in which students can freely interact with the instructor, and future research should investigate this question.

The present study can also be conceptualized as a case study of a successful researcher-practitioner relationship. Instructional techniques that are effective over the course of a 60-minute laboratory session using undergraduate participants with questionable motivation may not generalize to complex classroom contexts. Demonstrating efficacy of these techniques in naturalistic classroom environments is of utmost importance, and the optimal method to disseminate findings to educators is an open question. There is generally a trade-off between real-world applicability and fidelity of the empirical recommendations: classrooms are, in brief, messy open systems (Bronfenbrenner, 1976), and instructors are often unable to implement materials exactly as designed by researchers. For example, after finding compelling evidence that comparison aids middle school-age students comprehend algebra across several smaller-scale studies (Rittle-Johnson & Star, 2011), a large randomized-controlled trial that recruited 141 teachers in the US found no benefit, largely because the intervention was not delivered as intended (Star et al., 2015).

The present study offers a possible model for collaborating with practitioners to test an intervention initially based on laboratory studies. Specifically, the three-stage process of material development also represents a comparison of researcher-written materials to materials designed by a researcher and then adapted by an instructor to fit the specific classroom environment. A similar number of activities were delivered to students across conditions, but the teacher-centered worksheets led to better outcomes. In the context of studying a researcher-practitioner relationship, this finding suggests that learners may benefit more from instruction that is adapted to the classroom setting by an experienced instructor, even at the expense of some of the empirically-supported interventions. Future work should investigate this possibility more thoroughly and compare efficacy of researcher-designed to instructor-designed materials across multiple classrooms.

The present study also suffered from some limitations. For example, the dependent measure of interest, exam performance, may be a particularly noisy measure of learning and memory because the exams were open-note and open-book. Despite the likely dilution of the dependent measure, differences were still observed. The magnitude of the difference may be different in exams delivered in more traditional circumstances (i.e., closed book and proctored).

Similarly, another limitation is that the "dose" of the intervention was relatively small. Although many students do not follow this model, one credit-hour corresponds to two hours of work outside of class (US Department of Education 34 CFR 600.2). For LS 30A, a five credit-hour course, this corresponds to 15 hours of work each week. Students spent 40 minutes of total class time each week on the worksheets, which was only 15% of the live time and only 4% of the total time students are supposed to spend on coursework each week. While it would behoove students to use the resources they were presented with and to attempt to use the strategies they were taught while they were in class, it is not safe to assume that all students did so. Further, students often report cramming lots of information in a few study sessions before an exam (Hartwig & Dunlosky, 2012). In short, it was not possible (or perhaps even desirable) to control student behavior outside the classroom, which may have washed out any differences attributable to the relatively modest intervention. The teacher-centered worksheets generally contained more structure and might have been easier for students to use outside of the classroom. Overall, the fact that effects were observed despite all these considerations is a positive sign for the efficacy of teacher-centered instruction, and future work should attempt to increase the strength of the signal in order to examine whether or not the effects prove robust.

Another limitation involves order effects. This limitation was unavoidable given the number of LS 30A sections offered during the data collection period, but measures were taken to reduce the impact of order effects where possible. First, the topics on the Midterm and Final exam are relatively independent, so any impact of particularly effective instruction at the start of the course is less likely to carry over to the latter half of the course. Second, to provide a more pure measure, the dependent measure was computed using only Final Exam questions that assessed topics taught after the Midterm Exam. Future work might address order effects by recruiting a larger number of instructors.

It is also possible that certain topics are particularly amenable to a particular type of instruction. Unfortunately, the data from the current project do not allow for a granular topic-level analysis, but future work should address this question more thoroughly. In matching instructional interventions to topics, collaboration between researchers and educators is paramount. In the present work, recruiting a domain expert to create the teacher-centered worksheet addressed this issue somewhat. The expert was more sparing in her use of the analogical learning principles and removed them for topics that weren't a good fit.

In sum, the present findings demonstrate that analogy is an effective addition to STEM classrooms if it is used conscientiously. Low-performing students especially seem to benefit from more explicit guidance from both teachers and worksheets. In addition, the study suggests that domain experts should be involved in both design and delivery of instructional interventions. One key strength of an analogical approach is the flexibility of the principles. Once an instructor understands the underlying principles, they should have the freedom to apply them wherever they are relevant.

Chapter 5: General Discussion and Conclusions

The aim of the current project was to investigate the efficacy of an analogical approach to STEM education in both an online lab and a naturalistic classroom setting. Across three studies, the efficacy of the approach was mixed. In two online lab studies using undergraduate participants, instruction designed using an analogical approach fared similarly to control instruction on immediate tests. In a classroom setting, use of an analogical approach was most successful when the instructional materials included considerable scaffolding. The largest benefit was observed for low-performing students.

Whether due to prior preparation, cognitive resources, or a mix of other factors, high-ability college students are already at a significant advantage. Prior learning better prepares these students for future learning (e.g., McNamara & Kinstch, 1996) and these students can learn from a variety of instructional styles (Nesbit & Adesope, 2006). Students that enter college with less preparation or fewer cognitive resources require more guidance and more conscientious instruction.

The present findings show that an analogical approach to STEM education is likewise vulnerable to these differential effects on different populations of students: high-ability students were able to learn from all instructional conditions, while lower-ability students benefitted most from a scaffolded analogical intervention. Principles from analogical reasoning literature have the potential to increase conceptual understanding (e.g., Gray & Holyoak, 2021; Vendetti, Matlen, Richland, & Bunge, 2015) but the extent to which these principles improve educational outcomes for diverse populations remains unclear.

The project also places educational analogical interventions in the context of cognitive load theory (Sweller, 2011). While analogical comparison may be a method to increase germane cognitive load, educators must take care that students are not overloaded. Future work should extend these findings to identify the proper level of support required for analogical interventions. Further, this line of work may be combined with the principles behind concreteness fading instructional interventions (Fyfe, McNeil, Son, & Goldstone, 2014): novices may begin analogical comparisons with guidance that gradually fades away as learners gain expertise.

An analogical approach to STEM education shows promise, but educators must employ it conscientiously. Classrooms and the students in them are not uniform. Educators know their classrooms well and should adapt the principles of analogical instruction to their classrooms as they see fit. In the current project, the principles proved fairly easy for an instructor to understand and adapt. Granting agency to educators in how they choose to implement the approach stands to increase the dissemination of research findings into classrooms, which will increase its reach to students and potentially improve educational outcomes.

1. Use well-understood source analogs to capitalize on prior knowledge. Explain correspondences fully.
2. Highlight shared causal structure among examples of a structurally-defined category using visuospatial, gestural, and verbal supports.
3. Identify and explain correspondences between various representations.
4. Use analogical comparison to increase germane load and modify presentation style to facilitate comparison and reduce extraneous cognitive load when appropriate.
5. Once students have some proficiency with the material, encourage generation of inferences.

Table 1. Summary of principles for analogical approach to teaching

	Instructional condition			
Self-report strategy use on posttest		*Geometric*	*Symbolic*	*JUMP Math*
	No	29	11	20
	Sometimes	45	36	61
	Yes	38	56	37

Table 2. Relationship between strategy use and instructional condition in Study 2

	Class	Minimum	25th percentile	Median	75th percentile	Maximum
Midterm	Lecture 1 (TC)	57.5	80.0	90.0	92.5	100.0
	Lecture 2 (SC)	52.5	80.0	92.5	95.0	100.0
Final	Lecture 1 (SC)	41.3	72.6	84.4	91.2	98.7
	Lecture 2 (TC)	42.6	78.8	88.1	93.5	98.7

Table 3. Overall exam performance in the classroom study

*Note. Scores are reported as percentages.

	Class	Minimum	25th percentile	Median	75th percentile	Maximum
Overall performance	Lecture 1	-31.10	-0.01	6.05	12.94	32.98
	Lecture 2	-15.79	-1.59	3.45	6.68	28.49
Nonoverlapping topics	Lecture 1	-31.80	-2.39	6.53	11.67	35.76
	Lecture 2	-23.16	-2.53	2.00	8.97	31.25

Table 4. Change scores for each class from the classroom study.

*Note. Positive change scores indicate better performance on the Midterm than the Final.

$$c = \begin{cases} \dfrac{post - pre}{100 - pre} & post > pre \\ drop & post = pre = 100 \ or \ 0 \\ 0 & post = pre \\ \dfrac{post - pre}{pre} & post < pre \end{cases}$$

Equation 1. Normalized change score calculation from Marx & Cummings (2007)

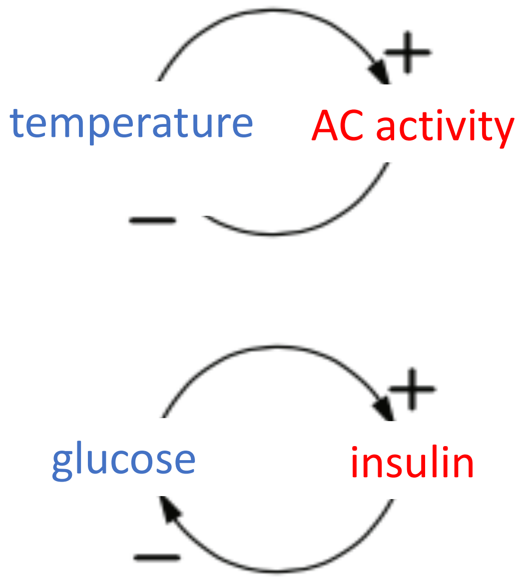


Figure 1.1. Analogous examples of a negative feedback loop
A diagram showing analogous examples of negative feedback loops. They are diagrammed to visuospatially highlight shared structure.

1. In the absence of predators, hares have a constant per-capita growth rate *0.1*.
2. The rate at which a single lynx catches hares is proportional to the hare population size with proportionality constant *0.02.*
3. The lynx birth rate is proportional to the amount of food caught by the population with proportionality constant *0.05.*
4. Lynx have a constant per capita death rate, *0.1.*

$$H' = 0.1 * H - 0.02 * LH$$

$$L' = 0.05 * .02 * LH - 0.1 * L$$

Figure 1.2. Example of model writing problem in LS 30A

Example of model writing problem from LS 30A taught using an analogical approach. H' represents how the population of hares changes and L' represents how the population of lynx changes. Color coding is used to highlight correspondences between the verbal/semantic and symbolic representation.

Figure 1.3. Stable and unstable equilibria
A diagram of stable (left) and unstable (right) equilibria used in LS 30A. The diagram relates equilibrium points in a differential equation to the behavior of a ball at the bottom of a cup (left) or on top of a hill (right). This figure appears in the Modeling Life textbook as Figure 3.3.

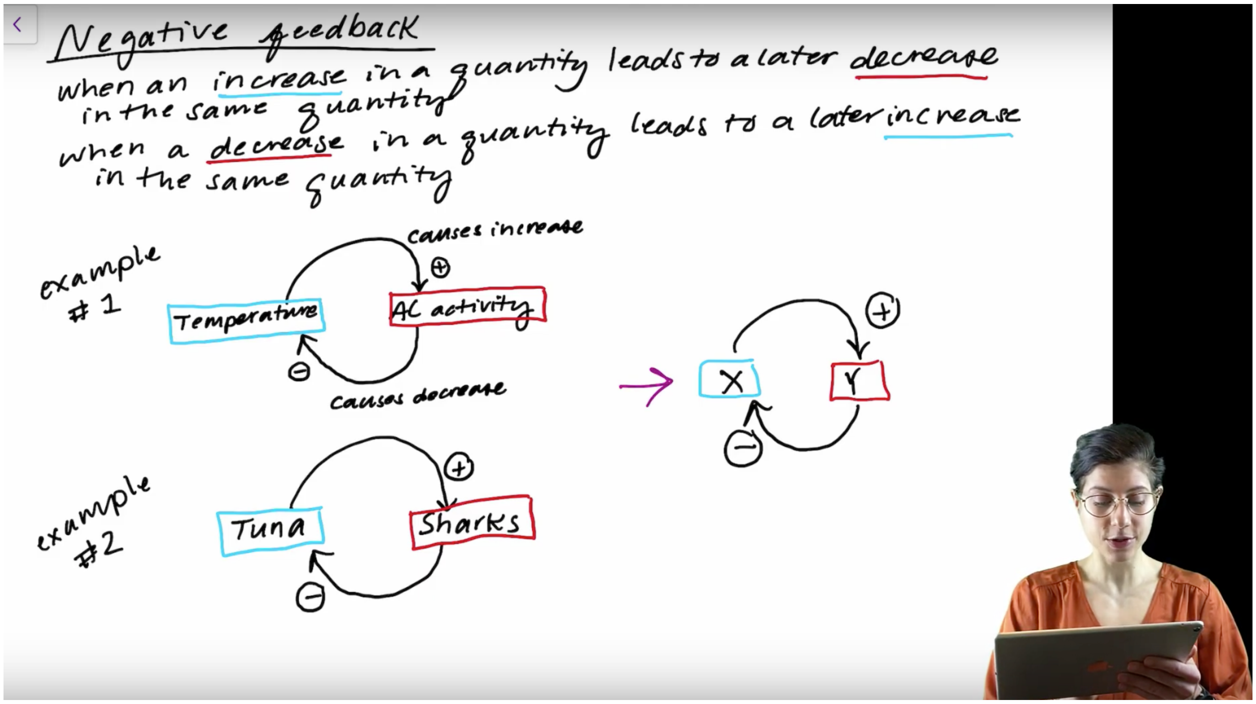


Figure 2.1. Instructional video from Study 1
Screenshot of an instructional video from Study 1 showing the instructor superimposed over a tablet screen. The image shows the analogical-feedback loop condition.

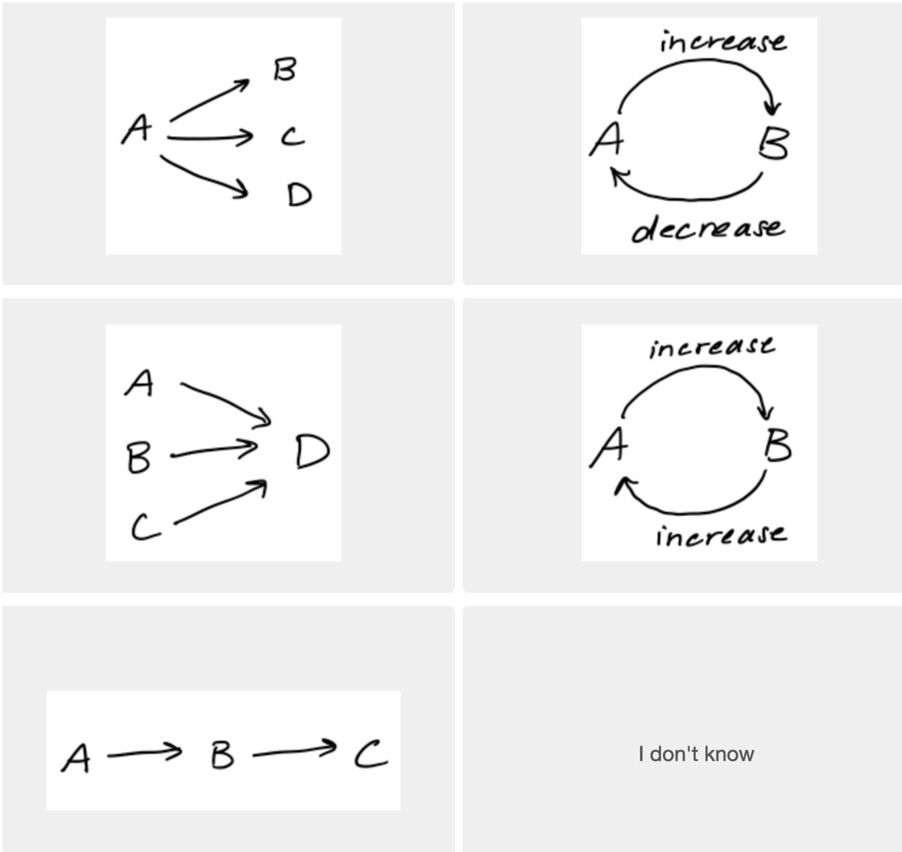


Figure 2.2. Causal relational diagrams used in Study 1 assessments
An image of the causal-relational diagrams used to assess understanding of feedback loops in Study 1. Participants clicked the diagram that best fit the description of each scenario.

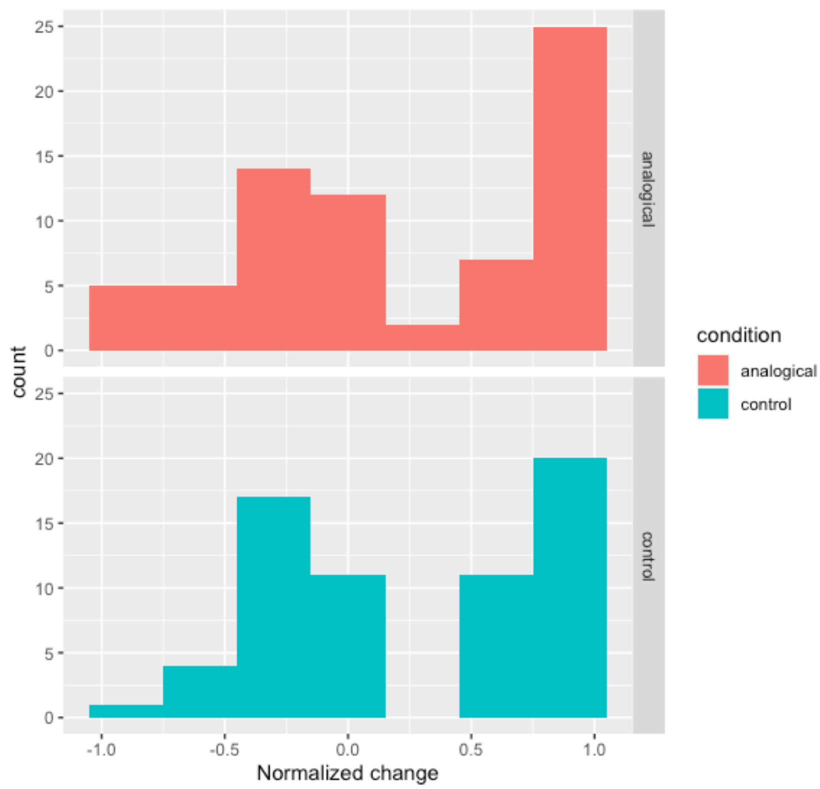


Figure 2.3. Normalized change scores for the feedback topic in Study 1
Normalized change scores broken down by condition for the feedback topic in Study 1.

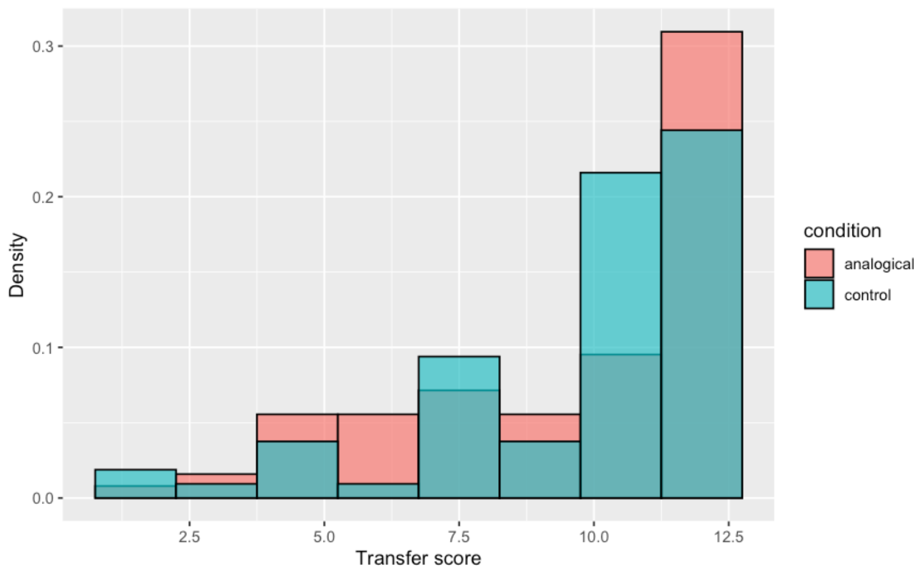


Figure 2.4. Transfer scores for the feedback topic in Study 1
Transfer performance broken down by condition for the feedback topic in Study 1. The maximum score was 12 points.

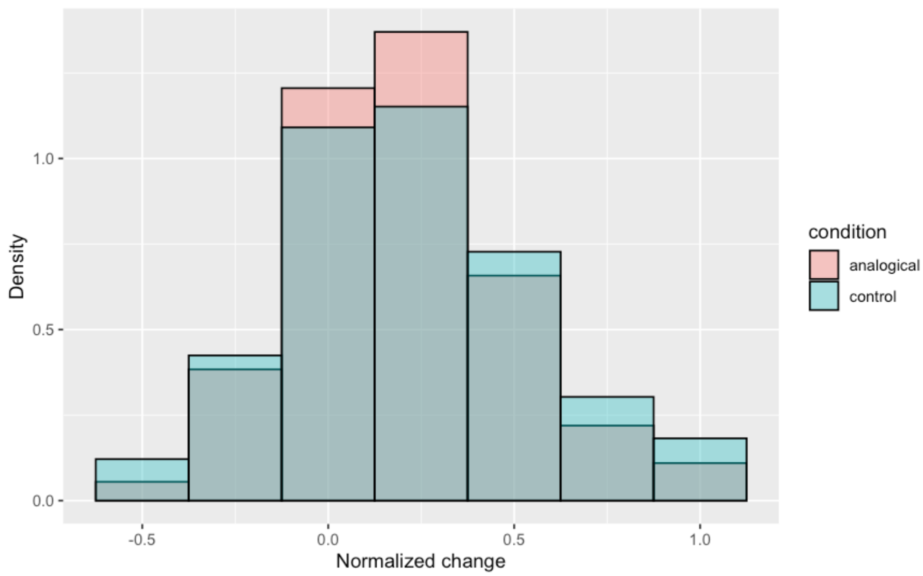


Figure 2.5. Normalized change scores for the functions topic in Study 1
Normalized change scores broken down by condition for the functions topic in Study 1.

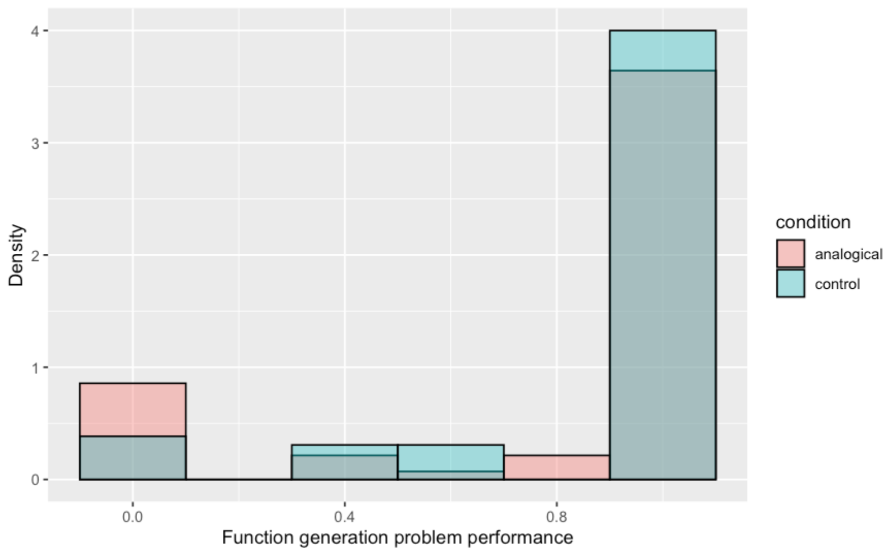
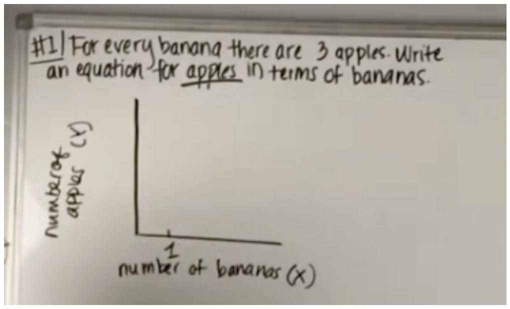
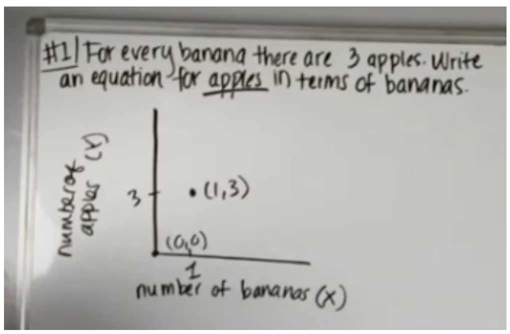


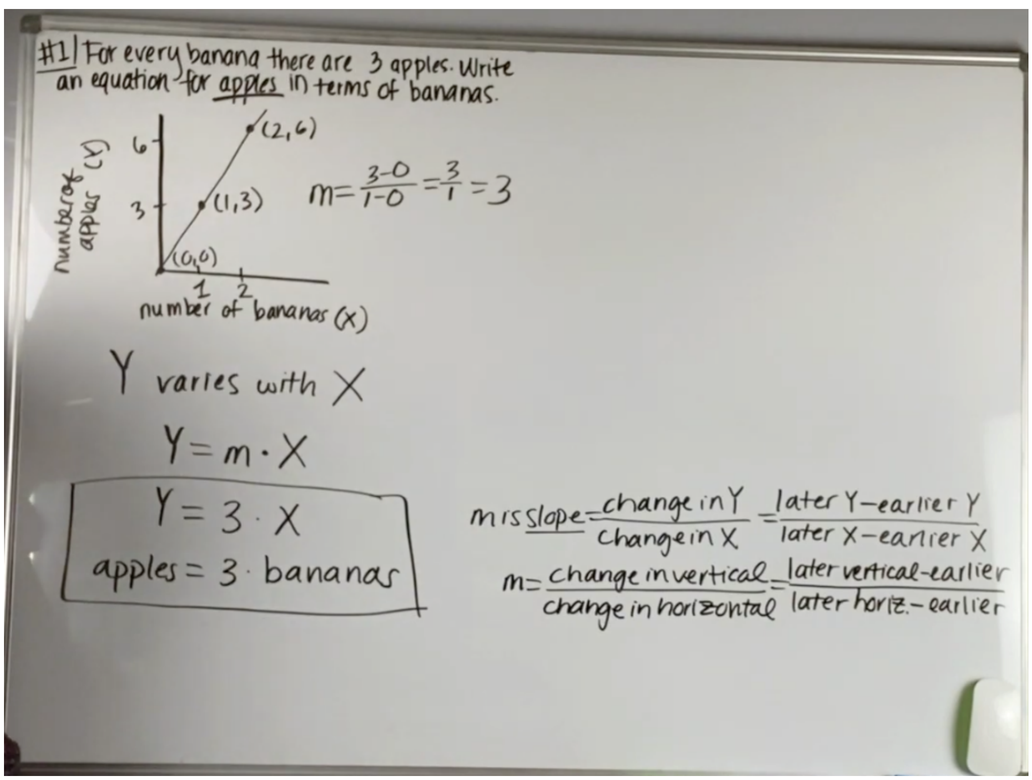
Figure 2.6. Transfer scores for the functions topic in Study 1
Function generation scores broken down by condition for the functions topic in Study 1. Scores are shown as proportion correct.



a.



b.



c.

Figure 3.1. Sample board work for the geometric condition in Study 2
Sample board work from the geometric instructional condition in Study 2.

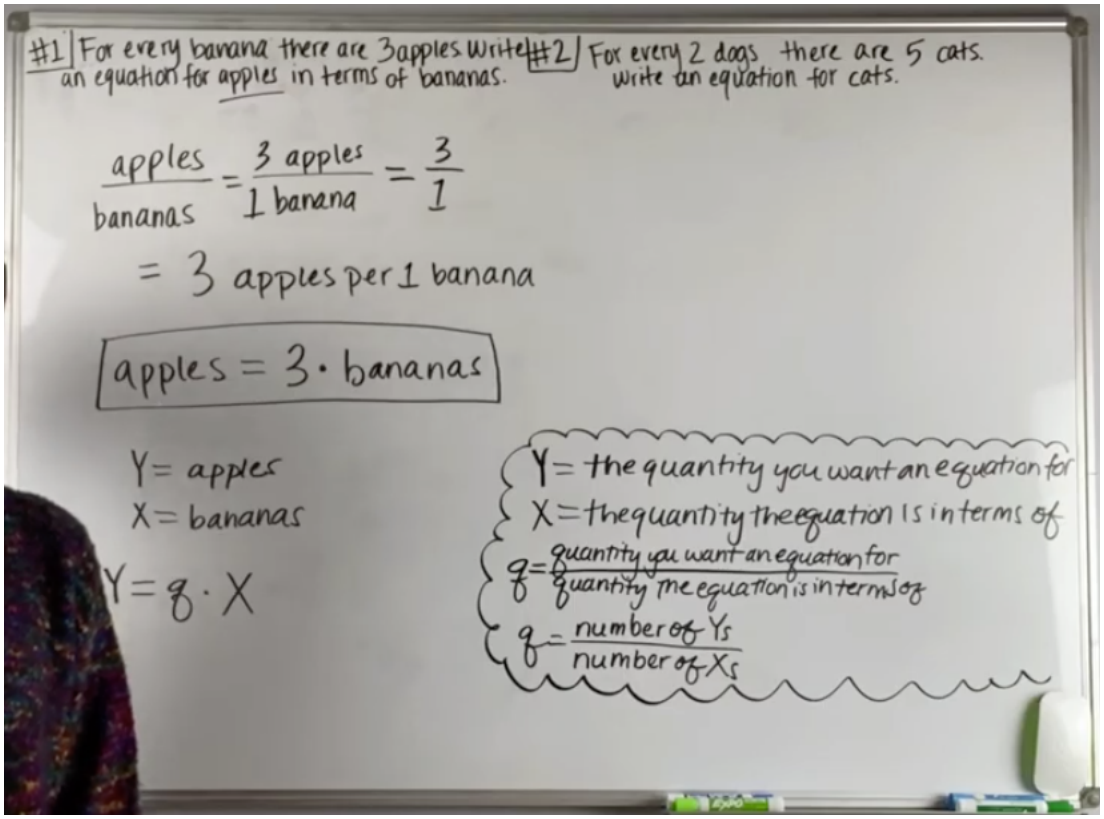
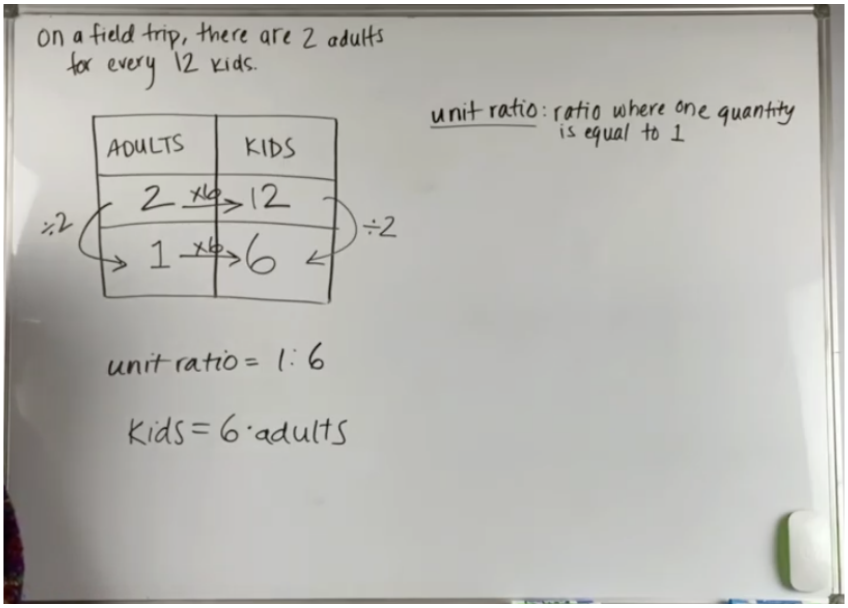
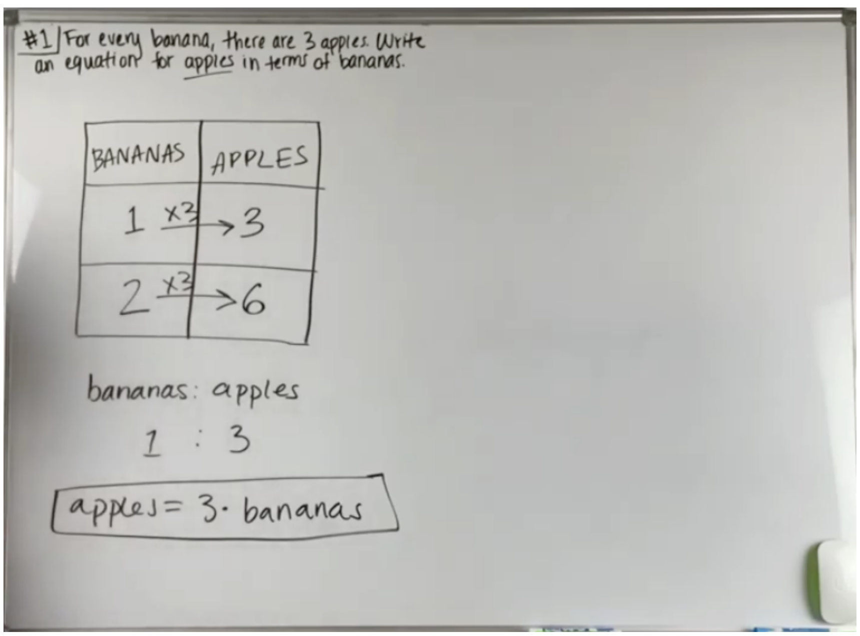


Figure 3.2. Sample board work for the symbolic control condition in Study 2
Sample board work from the symbolic control instructional condition in Study 2.



a.



b.

Figure 3.3. Sample board work for the JUMP Math condition in Study 2
Sample board work from the JUMP Math instructional condition in Study 2.

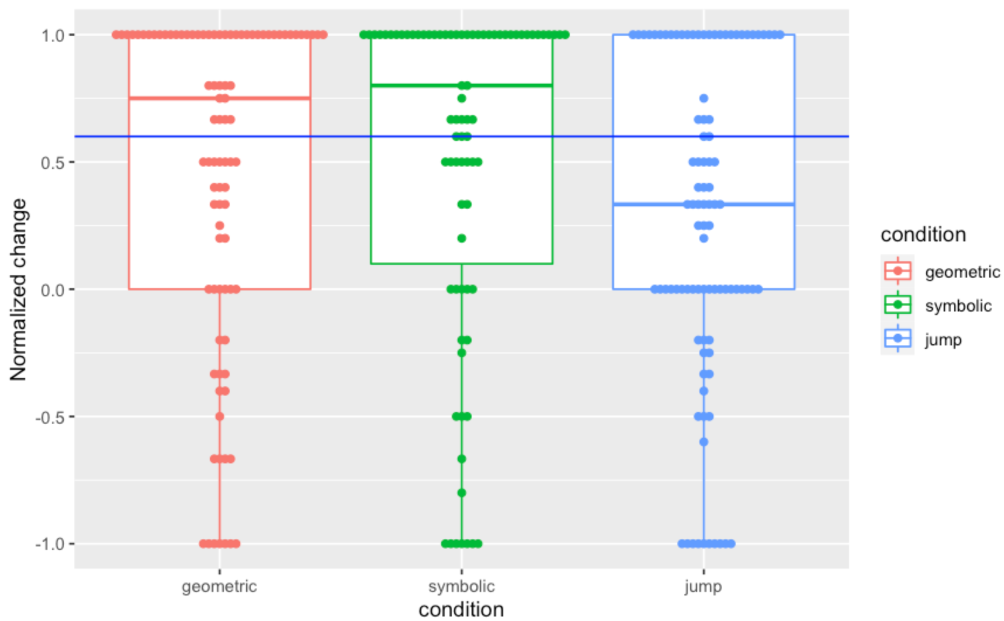


Figure 3.4. Normalized change scores for equation construction in Study 2
Normalized change scores for equation construction problems in Study 2 broken down by condition. The solid blue line indicates the grand median computed from all conditions.

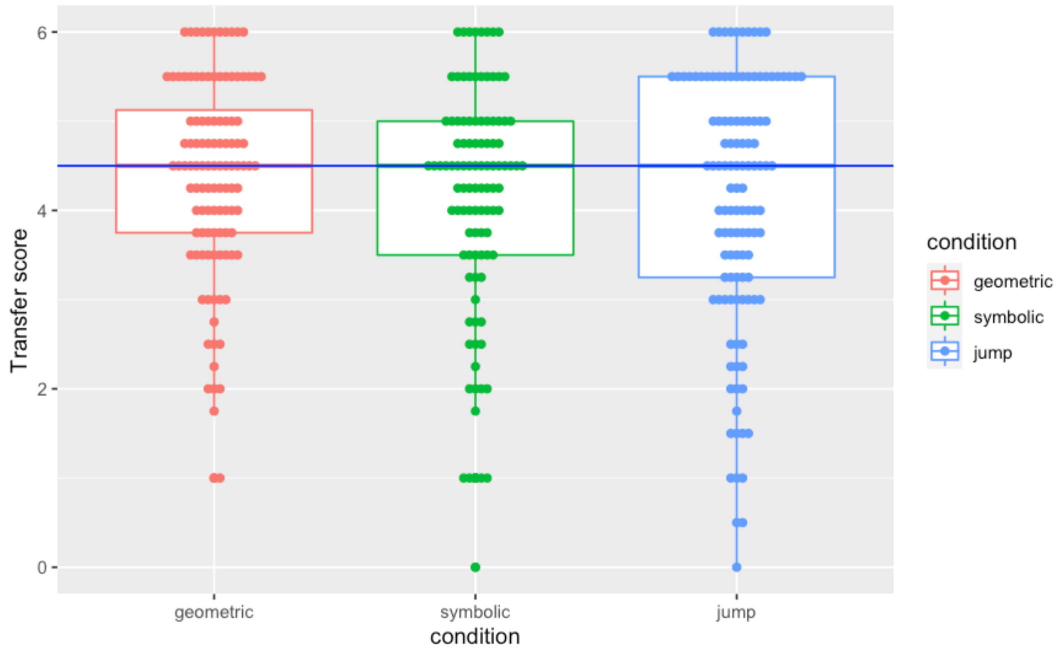


Figure 3.5. Transfer scores for equation construction in Study 2
Transfer for equation construction problems in Study 2 broken down by condition. The maximum score was 6. The solid blue line indicates the grand median computed from all conditions.

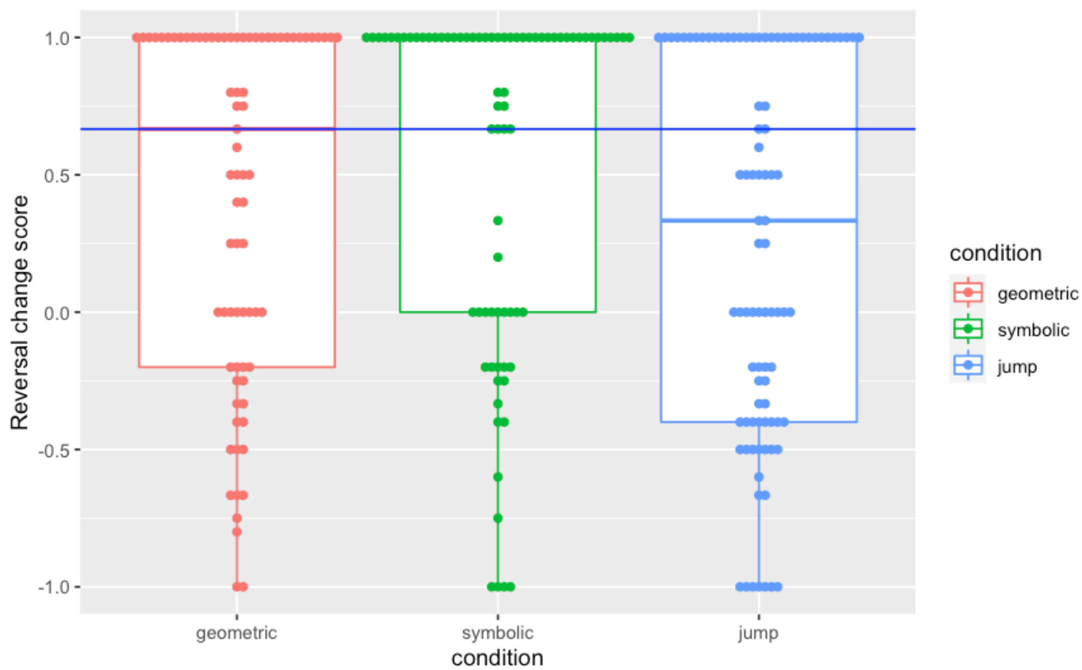


Figure 3.6. Reversal change scores for equation construction in Study 2
Reversal change scores for Study 2 broken down by condition. Solid blue line indicates grand median computed across all conditions.

Learning Team Activity: Compare/contrast models

Logistic equation $X' = bX - cX^2 = bX\left(1 - \frac{X}{k}\right)$ Where $c = \frac{b}{k}$	Ants/Aphids $N' = 0.03P - 0.02N$ $P' = 0.12P - 0.001P - 0.7 * \frac{1}{N} * P$
Eucalyptus/Koalas $E' = 0.01E - 0.02E - 0.05EK$ $K' = 0.15EK - 0.03K$	

1. Find the outflow terms in each model. Mark these in red.
    a. What is similar about the outflow terms?
    b. What is different about them?
    c. Which of the similarities you wrote down *makes something an outflow term?*
2. Find the inflow terms in each model. Mark these in blue.
    a. What is similar about the inflow terms?
    b. What is different about them?
    c. Which of the similarities you wrote down *makes something an inflow term?*
3. Find the parameters in each model. Mark these in green.
    a. What is similar about the parameters?
    b. What is different about them?
    c. Which of the similarities you wrote down *makes something a parameter?*
4. Find the terms that show state variables interacting with one another. Mark these in yellow.
    a. What is similar about the interaction terms?
    b. What is different about them?
    c. Which of the similarities you wrote down *makes something an interaction term?*

Figure 4.1. Compare/contrast modeling activity used in Classroom Study
Example activity from the SC worksheet showing a color coding scheme that highlights correspondences between different examples of the same concept (terms in a differential equation). In the TC worksheet, students completed the color coding activity in Learning Teams. Instead of generating similarities and differences, students were asked to describe how they identified each term.

	Euler's Method	Riemann Sum
What's the purpose of the process?		
What information do you have at the start of the process?		
What is the result of each process?		
What happens to Δt from row to row?		
What happens to t from row to row?		
What's the relationship between rows?		
What happens to the approximation as $\Delta t \to 0$?		

Figure 4.2. Compare/contrast table activity used in Classroom Study

Example of table used to organize compare/contrast exercises in the classroom study. In the SC worksheets, students worked through the table row-by-row, comparing concepts back and forth. In the TC worksheets, students were guided through the table column-by-column, completing each concept individually before any comparison was attempted.

Learning Team Activity: Explain Terms

My team's assignment: ____ (+ all teams will do the last row)

3. Here are the terms that should appear in your differential equation model for ants and aphids. For each term, (a) decide which assumption the term connects to (b) describe why it makes sense for the term to have the mathematical form that it does (c, cXY, etc.). (See the first row for an example.)

Term	(a)	(b)
$0.03NP$ *(N')*	Assumption (1)	This term is positive because it corresponds to an increase in ants. There is multiplication in this term because the increase of ants by birth depends on how many aphids are around to produce honeydew and how many ants are around to reproduce.
$-\ 0.02N$ *(N')*		
$0.12P$ *(P')*		
$0.001P^2$ *(P')*		
$-\ 0.7 * \frac{P}{N}$ *(P')*		

Figure 4.3. Connecting representations activity used in Classroom Study
Example of generating explanations to connect a mathematical representation to its verbal counterpart used in the SC worksheet. In the TC worksheets, the response in the (b) column was provided for students with some key phrases left blank. Students filled in the blanks.

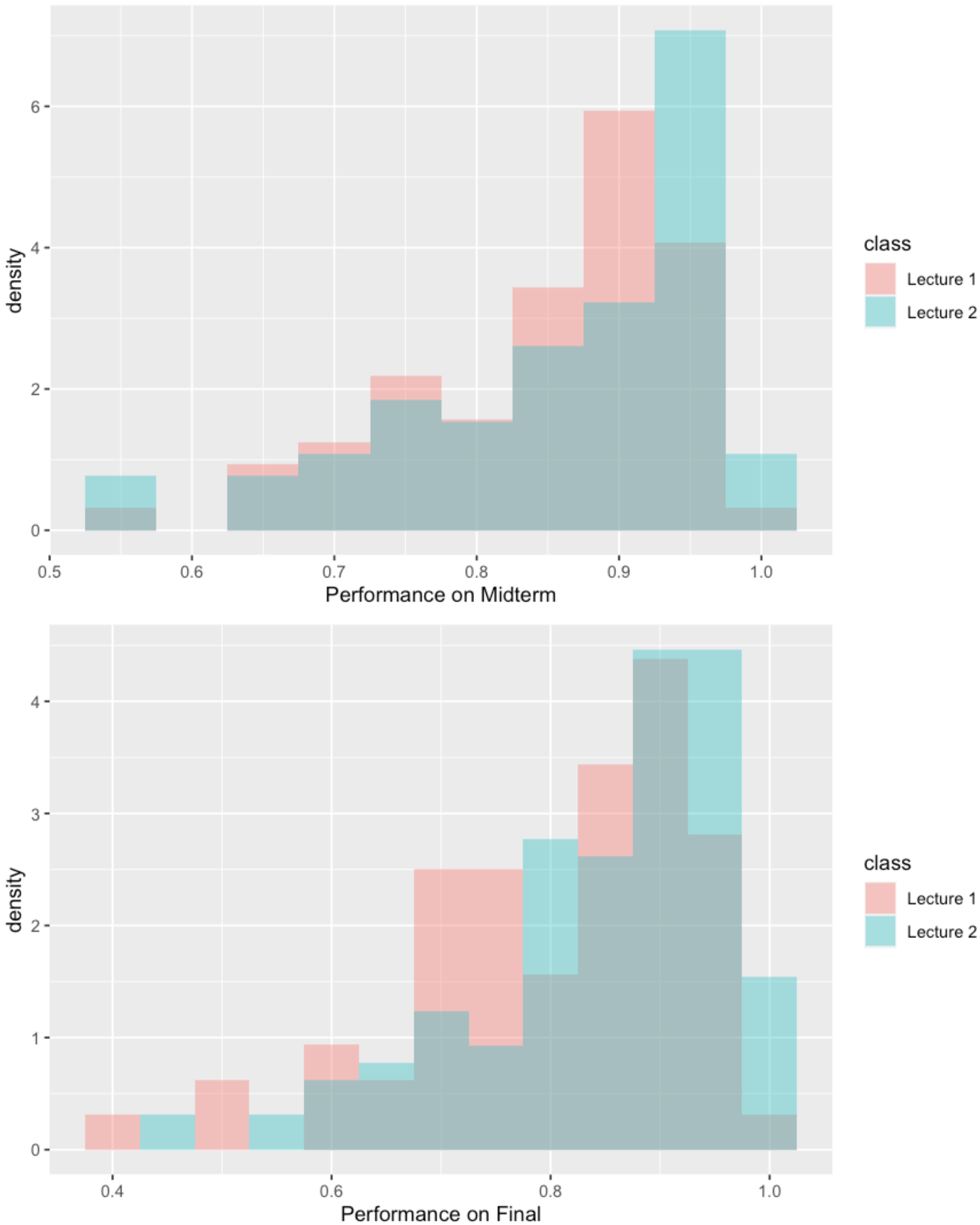


Figure 4.4. Overall Midterm and Final Exam performance in Classroom Study
Student performance on the Midterm and Final Exams (including all topics). Histograms are normalized due to unequal class size. Lecture 1 received TC instruction and Lecture 2 received SC instruction for the Midterm Exam. Lecture 1 received SC instruction and Lecture 2 received TC instruction for the Final Exam. Exam performance is shown as proportion correct.

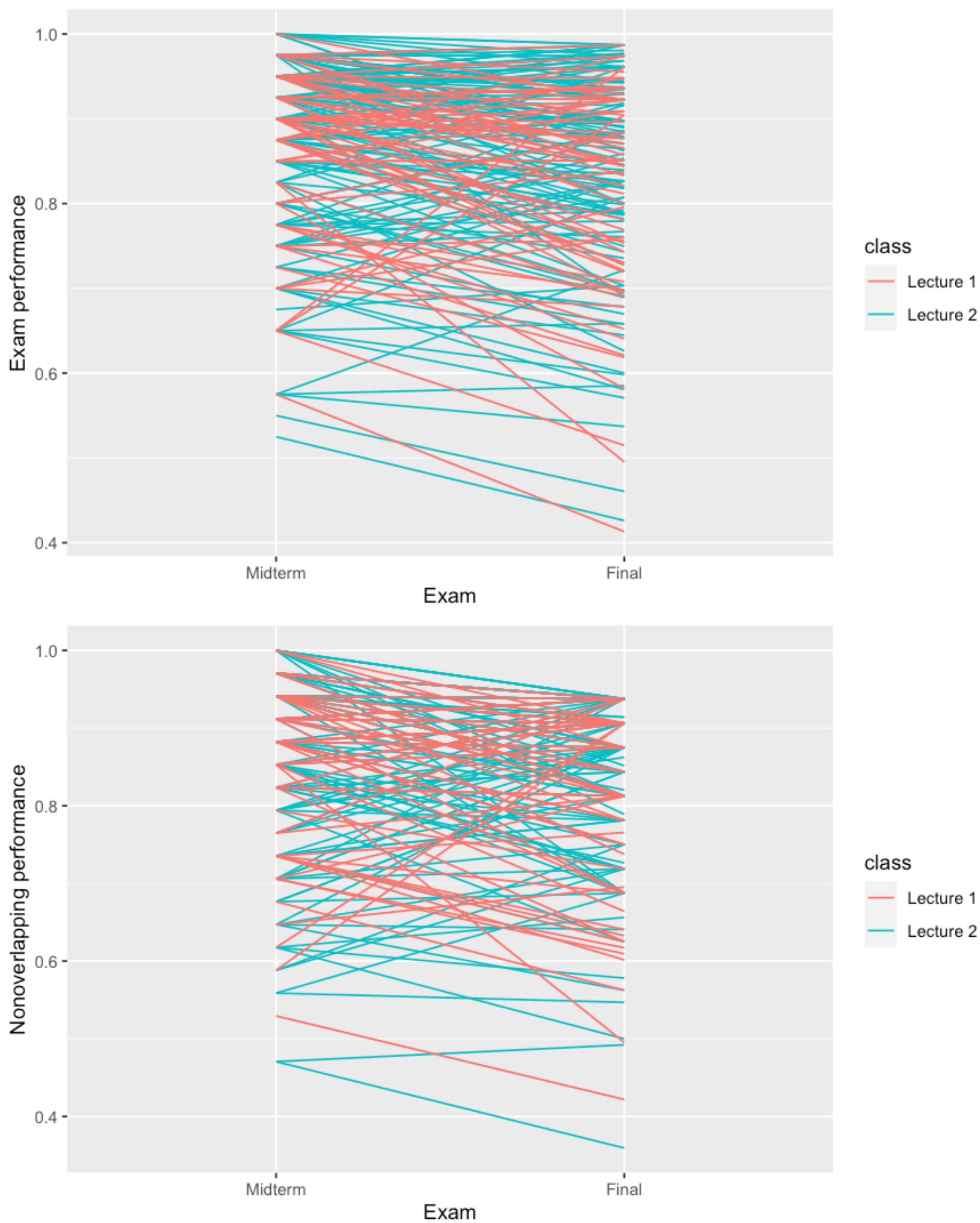


Figure 4.5. Changes in exam performance in Classroom Study
Change in performance from the Midterm to the Final Exam (both total performance and performance on non-overlapping topics unique to each instructional condition) broken down by class. Exam performance is shown as proportion correct.

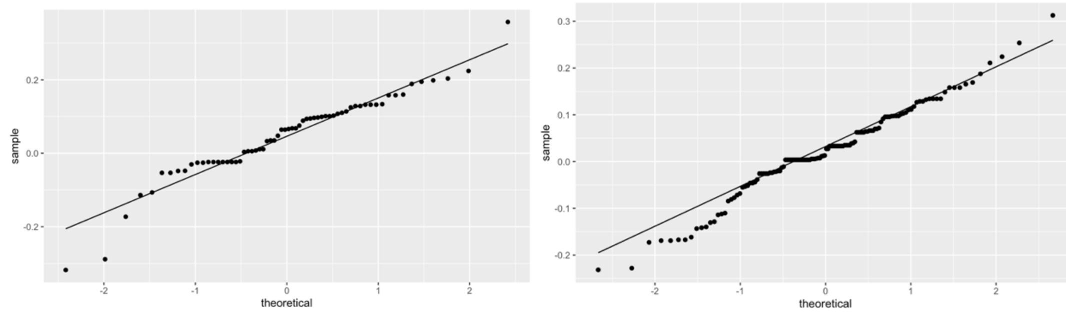


Figure 4.6. Q-Q plots of change scores in Classroom Study

QQ plots for change scores in non-overlapping topics. The left panel shows Lecture 1 and the right panel shows Lecture 2. Change scores are not normally distributed, so parametric analyses are not appropriate.

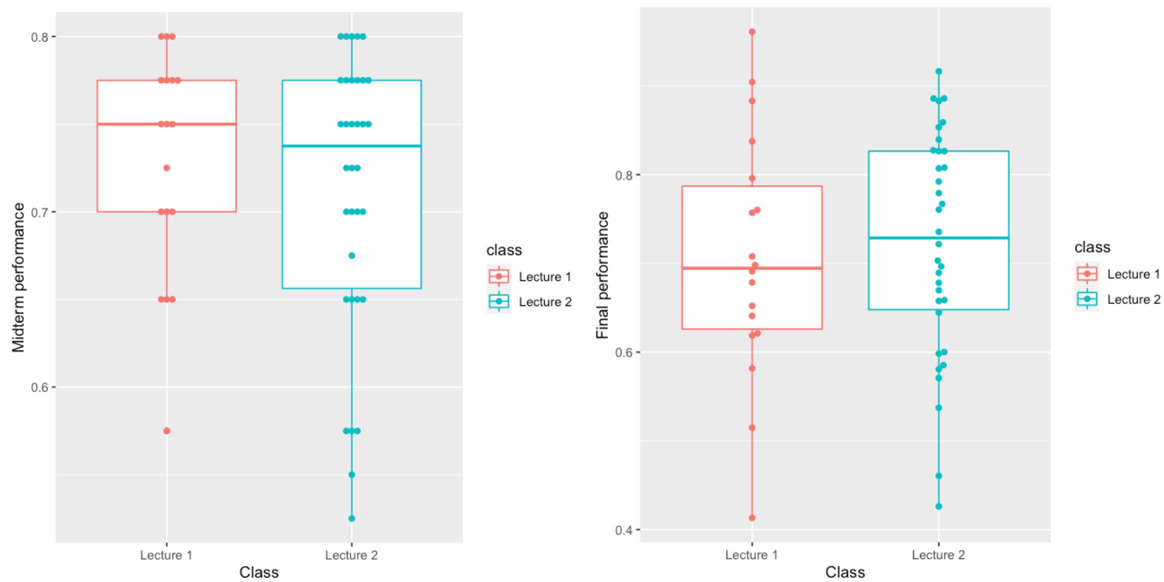


Figure 4.7. Exam performance for low-performing students in Classroom Students
Total score on each exam for students that performed at or below the 25th percentile on the Midterm exam.

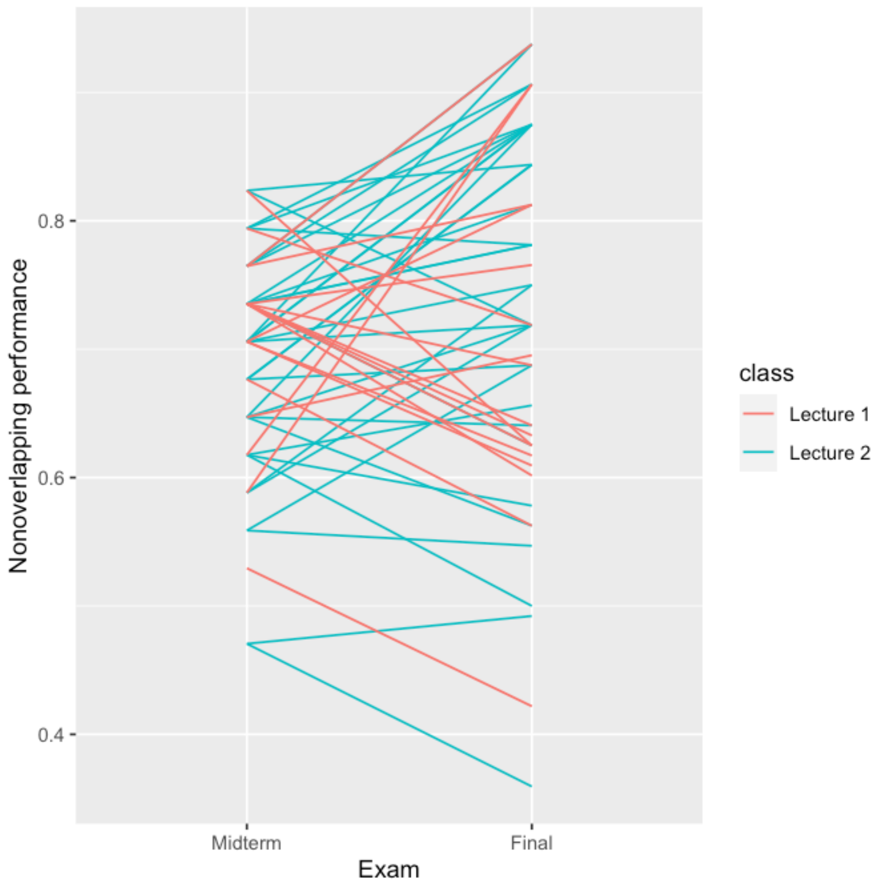


Figure 4.8. Changes in exam performance for low-performing students in Classroom Study
Changes in Midterm and Final exam performance in non-overlapping topics for students that performed at or below the 25th percentile on the Midterm.

References

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

Baker, W. P., & Lawson, A. E. (2001). Complex instructional analogies and theoretical concept acquisition in college genetics. *Science Education*, *85*(6), 665–683.

Bassok, M., Chase, V. M., & Martin, S. A. (1998). Adding Apples and Oranges: Alignment of Semantic and Formal Knowledge. *Cognitive Psychology*, *35*, 99–134.

Bean, T. W., Searles, D., Singer, H., & Cowen, S. (1990). Learning concepts from biology text through pictorial analogies and an analogical study guide. *The Journal of Educational Research*, *83*(4), 233–237.

Begolli, K. N., & Richland, L. E. (2016). Teaching mathematics by comparison: Analog visibility as a double-edged sword. *Journal of Educational Psychology*, *108*(2), 194.

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, *35*(2), 201–210. https://doi.org/10.3758/BF03193441

Biwer, F., de Bruin, A. B. H., Schreurs, S., & oude Egbrink, M. G. A. (2020). Future Steps in Teaching Desirably Difficult Learning Strategies: Reflections from the Study Smart Program. *Journal of Applied Research in Memory and Cognition*, *9*(4), 439–446. https://doi.org/10.1016/j.jarmac.2020.07.006

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology*, *64*(1), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Boaler, J. (2015). *What's Math Got to Do with It?: How Teachers and Parents Can Transform Mathematics Learning and Inspire Success*. Penguin.

Bronfenbrenner, U. (1976). *The Experimental Ecology of Education*. 12.

Boshuizen, H. P., & Schmidt, H. G. (2008). The development of clinical reasoning expertise. In J. Higgs, M. A. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (3rd ed., pp. 113–121). Philadelphia, PA: Elsevier.

Braasch, J. L., & Goldman, S. R. (2010). The role of prior knowledge in learning from analogies in science texts. *Discourse Processes*, *47*(6), 447–479.

Calmettes, G., Drummond, G. B., & Vowler, S. L. (2012). Making do with what we have: Use your bootstraps. *British Journal of Pharmacology*, *167*(2), 233–237. https://doi.org/10.1111/j.1476-5381.2012.02101.x

Campos, I. S., Almeida, L. S., Ferreira, A. I., Martinez, L. F., & Ramalho, G. (2013). Cognitive processes and math performance: A study with children at third grade of basic education. *European Journal of Psychology of Education*, *28*(2), 421–436.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1147.

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: When the "have-nots" gain but the "haves" gain even more . *The American Psychologist*, *60*(2), 149–160. https://doi.org/10.1037/0003-066X.60.2.149

Chen, X. (2005). *First-Generation Students in Postsecondary Education: A Look at Their College Transcripts*. 103.

Cheng, P. C.-H. (1999). Unlocking conceptual learning in mathematics and science with effective representational systems. *Computers & Education*, *33*(2–3), 109–130.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*(3), 257–273.

Christianson, K., Mestre, J. P., & Luke, S. G. (2012). Practice Makes (Nearly) Perfect: Solving 'Students-and-Professors'-Type Algebra Word Problems. *Applied Cognitive Psychology*, *26*(5), 810–822. https://doi.org/10.1002/acp.2863

Clement, C. A., & Yanowitz, K. L. (2003). Using an analogy to model causal mechanisms in a complex text. *Instructional Science*, *31*(3), 195–225.

Cohen, E., & Kanim, S. E. (2005). Factors influencing the algebra "reversal error." *American Journal of Physics*, *73*(11), 1072–1078.

College Board. (2020, February 28). What is the Average SAT Score? Retrieved November 12, 2021, from College Board Blog website: https://blog.collegeboard.org/what-is-the-average-sat-score

Devlin, K. J. (2012). *Introduction to mathematical thinking*. Palo Alto, Calif: Devlin.

DeWolf, M., Bassok, M., & Holyoak, K. J. (2015). Conceptual structure and the procedural affordances of rational numbers: Relational reasoning with fractions and decimals. *Journal of Experimental Psychology: General*, *144*(1), 127–150. https://doi.org/10.1037/xge0000034

Duit, R. (1991). On the role of analogies and metaphors in learning science. *Science Education*, *75*(6), 649–672. https://doi.org/10.1002/sce.3730750606

Durkin, K., Rittle-Johnson, B., Star, J. R., & Loehr, A. (2021). Comparing and discussing multiple strategies: An approach to improving algebra instruction. *The Journal of Experimental Education*, 1–19.

Fries, L., Son, J. Y., Givvin, K. B., & Stigler, J. W. (2021). Practicing connections: A framework to guide instructional design for developing understanding in complex domains. *Educational Psychology Review*, *33*(2), 739–762.

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review*, *26*(1), 9–25.

Garfinkel, A., Shevtsov, J., & Guo, Y. (2017). *Modeling life*. New York, NY: Springer Berlin Heidelberg.

Gentner, D., Holyoak, K. J., Holyoak, K. J., & Kokinov, B. N. (2001). *The Analogical Mind: Perspectives from Cognitive Science*. MIT Press.

Gentner, D., & Kurtz, K. J. (2005). Relational categories. In W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin.* (pp. 151–175). Washington: American Psychological Association. https://doi.org/10.1037/11156-009

Gentner, D., Levine, S. C., Ping, R., Isaia, A., Dhillon, S., Bradley, C., & Honke, G. (2016). Rapid learning in a children's museum via analogical comparison. *Cognitive Science*, *40*(1), 224–240.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393–408. https://doi.org/10.1037/0022-0663.95.2.393

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*(1), 45.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*(3), 306–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38.

Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, *31*(1), 1–40.

Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition*, *137*, 137–153.

Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin*, *142*(7), 729.

Gray, M. E., & Holyoak, K. J. (2020). Individual differences in relational reasoning. *Memory & Cognition*, *48*(1), 96–110. https://doi.org/10.3758/s13421-019-00964-y

Gray, M. E., & Holyoak, K. J. (2021). Teaching by Analogy: From Theory to Practice. *Mind, Brain, and Education*. https://doi.org/10.1111/mbe.12288

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74. https://doi.org/10.1119/1.18809

Harrison, A. G., & Coll, R. K. (2008). *Using Analogies in Middle and Secondary Science Classrooms: The FAR Guide – An Interesting Way to Teach With Analogies*. Corwin Press.

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, *19*(1), 126–134. https://doi.org/10.3758/s13423-011-0181-y

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, *28*(5), 524–549.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*(3), 427.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *110*(2), 220.

Jaeger, A. J., Taylor, A. R., & Wiley, J. (2016). When, and for whom, analogies help: The role of spatial skills and interleaved presentation. *Journal of Educational Psychology*, *108*(8), 1121.

Jaeger, A. J., & Wiley, J. (2015). Reading an analogy can cause the illusion of comprehension. *Discourse Processes*, *52*(5–6), 376–405.

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T. F., & Sageman, B. (2013). Finding faults: Analogical comparison supports spatial concept learning in geoscience. *Cognitive Processing*, *14*(2), 175–187.

Jonassen, D. H. (2003). Designing Research-Based Instruction for Story Problems. *Educational Psychology Review*, *15*(3), 267–296. JSTOR. Retrieved from JSTOR.

Jones, B., & Kenward, M. G. (1989). *Design and analysis of cross-over trials*. Chapman and Hall/CRC.

Kalyuga, S. (2007). Expertise Reversal Effect and Its Implications for Learner-Tailored Instruction. *Educational Psychology Review*, *19*(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3

Kimball, D. R., & Holyoak, K. J. (2000). Transfer and expertise. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 109–122). New York: Oxford University Press.

Klein, P. D., Piacente-Cimini, S., & Williams, L. A. (2007). The role of writing in learning from analogies. *Learning and Instruction*, *17*(6), 595–611.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, *67*(6), 2797–2822.

Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., … Knowlton, B. J. (2008). Distraction during relational reasoning: The role of prefrontal cortex in interference control. *Neuropsychologia*, *46*(7), 2020–2032.

Kubricht, J. R., Lu, H., & Holyoak, K. J. (2017). Individual differences in spontaneous analogical transfer. *Memory & Cognition*, *45*(4), 576–588. https://doi.org/10.3758/s13421-016-0687-7

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

Larkin, J. H., & Simon, H. A. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science*, *11*(1), 65–100. https://doi.org/10.1016/S0364-0213(87)80026-5

Lee, H. S., DeWolf, M., Bassok, M., & Holyoak, K. J. (2016). Conceptual and procedural distinctions between fractions and decimals: A cross-national comparison. *Cognition*, *147*, 57–69. https://doi.org/10.1016/j.cognition.2015.11.005

MacGregor, M., & Stacey, K. (1993). Cognitive models underlying students' formulation of simple linear equations. *Journal for Research in Mathematics Education*, *24*(3), 217–232.

Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics*, *69*(S1), S12–S23. https://doi.org/10.1119/1.1371296

Martin, S. A., & Bassok, M. (2005). Effects of semantic cues on mathematical modeling: Evidence from word-problem solving and equation construction tasks. *Memory & Cognition*, *33*(3), 471–478. https://doi.org/10.3758/BF03193064

Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics*, *75*(1), 87–91. https://doi.org/10.1119/1.2372468

Matlen, B., Vosniadou, S., Jee, B., & Ptouchkina, M. (2011). Enhancing the comprehension of science text through visual analogies. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*(3), 247–288.

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors, and feedback. *Psychonomic Bulletin & Review*, *14*(2), 225–229. https://doi.org/10.3758/BF03194056

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, *76*(3), 413–448.

Nistal, A. A., Van Dooren, W., Clarebout, G., Elen, J., & Verschaffel, L. (2009). Conceptualising, investigating and stimulating representational flexibility in mathematical problem solving and learning: A critical review. *ZDM*, *41*(5), 627–636.

Nokes-Malach, T. J., VanLehn, K., 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.

OECD. (2018). *Education at a Glance 2018: OECD Indicators*. OECD. https://doi.org/10.1787/eag-2018-en

Orgill, M., & Bodner, G. M. (2006). An analysis of the effectiveness of analogy use in college-level biochemistry textbooks. *Journal of Research in Science Teaching*, *43*(10), 1040–1060.

O'Shea, A., Breen, S., & Jaworski, B. (2016). The development of a function concept inventory. *International Journal of Research in Undergraduate Mathematics Education*, *2*(3), 279–296. https://doi.org/10.1007/s40753-016-0030-5

Paas, F. G., & Van Merriënboer, J. J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, *6*(4), 351–371.

Rapp, M., Bassok, M., DeWolf, M., & Holyoak, K. J. (2015). Modeling discrete and continuous entities with fractions and decimals. *Journal of Experimental Psychology: Applied*, *21*(1), 47–56. https://doi.org/10.1037/xap0000036

Reed, S. K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(1), 124.

Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, *31*(2), 115–121. https://doi.org/10.1207/s15326985ep3102_3

Research Supporting JUMP Math | JUMP Math. (n.d.). Retrieved February 23, 2020, from https://jumpmath.org/jump/us/node/56#two-year

Richland, L. E., & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology*, *35*(1), 28–43.

Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the Conceptual Structure of Mathematics. *Educational Psychologist*, *47*(3), 189–203. https://doi.org/10.1080/00461520.2012.667065

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science*, *316*(5828), 1128.

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.

Rittle-Johnson, B., & Star, J. R. (2011). The power of comparison in learning and instruction: Learning outcomes supported by different types of comparisons. In *Psychology of Learning and Motivation* (Vol. 55, pp. 199–225). Elsevier.

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.

Roth, K., & Garnier, H. (2006). What science teaching looks like: An international perspective. *Educational Leadership*, *64*(4), 16.

Rottman, B. M., Gentner, D., & Goldwater, M. B. (2012). Causal Systems Categories: Differences in Novice and Expert Categorization of Causal Phenomena. *Cognitive Science*, *36*(5), 919–932. https://doi.org/10.1111/j.1551-6709.2012.01253.x

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565.

Simon, H. A., & Hayes, J. R. (1976). The understanding process: Problem isomorphs. *Cognitive Psychology*, *8*(2), 165–190. https://doi.org/10.1016/0010-0285(76)90022-0

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*(2), 176–199.

Solomon, T., Dupuis, A., O'Hara, A., Hockenberry, M.-N., Lam, J., Goco, G., … Tannock, R. (2019). A cluster-randomized controlled trial of the effectiveness of the JUMP Math program of math instruction for improving elementary math achievement. *PLOS ONE*, *14*(10), e0223049. https://doi.org/10.1371/journal.pone.0223049

Stains, M., & Talanquer, V. (2008). Classification of chemical reactions: Stages of expertise. *Journal of Research in Science Teaching*, *45*(7), 771–793. https://doi.org/10.1002/tea.20221

Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015). Learning from comparison in algebra. *Examining Innovations—Navigating the Dynamic Complexities of School-Based Intervention Research*, *40*, 41–54. https://doi.org/10.1016/j.cedpsych.2014.05.005

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*(2), 455–488.

Stigler, J. W., Givvin, K. B., & Thompson, B. J. (2010). What community college developmental mathematics students understand about mathematics. *MathAMATYC Educator*, *1*(3), 4–16.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. *The Cambridge Handbook of Multimedia Learning*, *3*(2), 19–30.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138.

Sweller, J. (2011). Cognitive load theory. In J. P. Mester & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Cambridge, MA: Elsevier.

Teacher Resource for Grade 8—New US Edition | JUMP Math. (n.d.). Retrieved February 19, 2020, from https://jumpmath.org/jump/us/tg_usa_8

Tekkumru-Kisa, M., Kisa, Z., & Hiester, H. (2020). Intellectual Work Required of Students in Science Classrooms: Students' Opportunities to Learn Science. *Research in Science Education*. https://doi.org/10.1007/s11165-020-09924-y

Tohill, J. M., & Holyoak, K. J. (2000). The impact of anxiety on analogical reasoning. *Thinking & Reasoning*, *6*(1), 27–40.

Touw, K. W. J., Vogelaar, B., Verdel, R., Bakker, M., & Resing, W. (2017). Children's progress in solving figural analogies: Are outcomes of dynamic testing helpful for teachers? *Educational and Child Psychology*, *34*(1), 21–38.

Treagust, D. F., Harrison, A. G., & Venville, G. J. (1998). Teaching Science Effectively With Analogies: An Approach for Preservice and Inservice Teacher Education. *Journal of Science Teacher Education*, *9*(2), 85–101. https://doi.org/10.1023/A:1009423030880

Tyumeneva, Y. A., Larina, G., Alexandrova, E., DeWolf, M., Bassok, M., & Holyoak, K. J. (2018). Semantic alignment across whole-number arithmetic and rational numbers: Evidence from a Russian perspective. *Thinking & Reasoning*, *24*(2), 198–220. https://doi.org/10.1080/13546783.2017.1374307

Vendetti, M. S., Matlen, B. J., Richland, L. E., & Bunge, S. A. (2015). Analogical reasoning in the classroom: Insights from cognitive science. *Mind, Brain, and Education*, *9*(2), 100–106.

Vendetti, M. S., Wu, A., & Holyoak, K. J. (2014). Far-Out Thinking: Generating Solutions to Distant Analogies Promotes Relational Thinking. *Psychological Science*, *25*(4), 928–933. https://doi.org/10.1177/0956797613518079

Venville, G. J., Bryer, L., & Treagust, D. F. (1994). Training students in the use of analogies to enhance understanding in science. *Australian Science Teachers Journal*, *40*(2), 60–66.

Viskontas, I. V., Morrison, R. G., Holyoak, K. J., Hummel, J. E., & Knowlton, B. J. (2004). Relational integration, inhibition, and analogical reasoning in older adults. *Psychology and Aging*, *19*(4), 581.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition*, *28*(7), 1205–1212.

Weaver III, C. A., & Kintsch, W. (1992). Enhancing students' comprehension of the conceptual structure of algebra word problems. *Journal of Educational Psychology*, *84*(4), 419.

Weinstein, R. S. (1983). Student Perceptions of Schooling. *The Elementary School Journal*, *83*(4), 287–312. JSTOR. Retrieved from JSTOR.

Wright, G. B. (2011). Student-centered learning in higher education. *International Journal of Teaching and Learning in Higher Education*, *23*(1), 92–97.

Yanowitz, K. L. (2001). Using analogies to improve elementary school students' inferential reasoning about scientific concepts. *School Science and Mathematics*, *101*(3), 133–142.

Zook, K. B., & Maier, J. M. (1994). Systematic analysis of variables that contribute to the formation of analogical misconceptions. *Journal of Educational Psychology*, *86*(4), 589.