

UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

The Role of Executive Functions for Overcoming Misconceptions through Structure-Mapping in Mathematics Classrooms

Permalink

<https://escholarship.org/uc/item/2bj05317>

Author

Begolli, Kreshnik

Publication Date

2015

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

The Role of Executive Functions and Visual Supports for Overcoming Misconceptions through
Structure-Mapping in Mathematics Classrooms

DISSERTATION

Submitted in partial satisfaction of the requirements

For the degree of

DOCTOR OF PHILOSOPHY

In Education

by

Kreshnik Begolli

Dissertation Committee:
Assistant Professor, Susanne Jaeggi, Chair
Associate Professor, Lindsey Richland
Associate Professor, Rossella Santagata

2015

DEDICATION

In remembrance of my father, Nasi Begolli (1932 – 2007).

Nasi Begolli embodied a treasure of virtues with remarkable poise, yet thought by his friends to own a single vice – *“he was too kind for his own good.”*

My “babi” (father in Albanian) was a soft-spoken man of collected thoughts who lived by essential principles. A living reminder of equilibrium, a stark contrast from me, he taught me there can be “too much of a good thing.” His selflessness was thus, deliberate. Babi’s attitude empowered me with a legacy of appreciating a life devoted to bettering the lives of others. As a teacher, governmental administrator, executive, and father, he has positively shaped countless lives. I hope that through this and future research, I will continue his legacy of deliberate selflessness.

Table of Contents

DEDICATION	ii
ACKNOWLEDGEMENTS	v
INCLUDED PAPERS.....	vii
LIST OF FIGURES	viii
LIST OF TABLES	ix
CURRICULUM VITAE.....	x
ABSTRACT OF THE DISSERTATION	xiii
The Role of Executive Functions and Visual Supports in Instructional Analogy	1
Research Overview.....	4
Chapter 1: Bridging Cognitive Experimentation and Real Classrooms: A Video	
Methodology for Experimental Research in Education	8
The Importance of Capturing the Environment in Stimuli	12
From Video-Based Observations to Video-Based Experimentation	15
A Video Methodology for Bridging Cognitive Research with Teaching Practices	17
Video-Editing for Creating Experimental Conditions that Test Efficient Classroom	
Techniques: Case Study.....	20
Discussion.....	26
Summary	31
Chapter 2: Teaching Mathematics by Comparison: Analog Visibility as A Double-Edged	
Sword (Begolli & Richland, 2015).....	33
Experiment 1.....	38
Method.....	38
Participants	38
Materials	38
Interactive Instructional Lesson.....	39
Mathematics Assessment.....	42
Design & Procedure.....	46
Results	46
Baseline Data.....	46
Condition Effects.....	47
Discussion.....	51
Chapter 3: The Role of Executive Functions for Structure-Mapping in Mathematics	
(Begolli et al., 2015).....	55

The Relationship between Analogy, Conceptual Change, Working Memory, and Inhibitory Control.....	57
EF and Instructional Analogy.....	60
Experiment 2.....	61
Method.....	61
Participants	61
Design & Procedure.....	61
Results	67
Baseline Analyses.....	67
Relationship between EF and Mathematics.....	69
Discussion.....	74
Chapter 4: General Discussion of Theoretical and Practical Implications	78
Implications for Theory	78
Implications for Practice	81
Limitations and Future Directions	84
General Summary	87
References	89
APPENDIX A	113
APPENDIX B	118

ACKNOWLEDGEMENTS

I cannot begin to imagine my academic experience without the continued mentoring, understanding, and support I received from Lindsey Richland. Lindsey has the mind of a genius and the heart of a saint. I have had the incredible fortune to have a mentor surpassing anything I would have attempted to imagine. Despite the distance, time zone differences, incredibly busy schedules, Lindsey has always managed to make time to advise me on every single academic decision I have made in the past five years. Lindsey is my role model for the researcher, mentor, and person, I wish to be. I would not be here without her.

I have a profound appreciation for the support that Susanne Jaeggi has given me. She has patiently and consistently committed to helping me succeed in my projects, invaluable reminding me to take careful, yet powerful steps in research. I owe the success of the work presented in my last chapter to Susanne.

A sincere thank you to all of the people that have served on my committee and provided me with priceless feedback: George Farkas, Rossella Santagata, Bradley Hughes, and Tiffany Willoughby-Herard. Special acknowledgement goes to George Farkas and Rossella Santagata for supporting me throughout my academic transitions. Thanks to Leora Fellus for listening and always managing to find a productive, painless solution.

Part of the work presented here was made possible by the National Science Foundation grant #0954222 and the Merage Fellowship for the American Dream. The views expressed here are mine, not theirs. This thesis would not have been possible without the hard and tedious work of my undergraduate collaborators: James Gamboa, Carmen Chan, Brooke Herd, Hannah Sayonno, Illiana Yopez, Hugo Yopez, Hao Ngo, “Q” Vong, Susane Nelson, Alice Kim, Grace Lin, and Allison Robertson. Also, the lab managers: Carey DeMichelis, Nina Simms, and Ellen

Sheehan. Programming help from Martin Buschkuehl and Raminder Goraya. And video-editing help from Allison O’Daniel. My deep appreciation goes to the fourteen teachers from 5 different school districts, the principals, students, and parents whose children were involved in these projects.

Next, I would like to thank my friends and colleagues who have read and listened to many of my drafts and research ideas. I want to particularly thank Erzen Hoda and his family for being part of every professional and personal milestone of mine.

A heartfelt thanks goes to my mother, Alltëne Begolli, a lifelong teacher ingraining in me an intense appreciation for education, and for proudly carrying all of us adorably in her thoughts. My two brothers, Bekim and Faton Begolli, for supporting me throughout this long journey – I would have given up long ago without your moral and financial support. I particularly appreciate the many sacrifices that my family has made to see me through this journey – I would not be alive to witness this without your tremendous help, continuous advice, and critical visits. Thank you for always picking up the phone and for always making me feel safe throughout my hardest moments.

My greatest inspiration comes from my daughter Jerina Solum-Begolli – she fuels my determination and passion for everything I do in life, especially the work that has gone into this dissertation. Another source of energy comes from my wife Visare Shala whom I met midway through the dissertation. I admire the dedication, passion, tirelessness, and the extraordinary sacrifices you made to start a new life with us. Thank you for the greatest gift in the world – our son, Rron Shala-Begolli, and your endless encouragement that has painted bright colors to my life’s canvas. Visare and Jerina have patiently endured through numerous lectures, teacher visits, and many hours of my work. My “guurlls” never cease to put a smile on my face.

INCLUDED PAPERS

Paper 1.

Begolli, K. N. & Richland, L. E. (submitted). Bridging Cognitive Experimentation and Real Classrooms: A Video Methodology.

Paper 2.

Begolli, K. N., & Richland, L. E. (2015). Teaching Mathematics by Comparison : Analog Visibility as a Double-Edged Sword. *Journal of Educational Psychology*, 107 (3).

Paper 3.

Begolli, K. N., Richland, L. E., & Jaeggi, S. M. (2015). The Role of Executive Functions for Structure-Mapping in Mathematics, 1, 1–6. In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warfaumont & J. Yoshimi (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Paper 4.

Begolli, K. N., & Richland, L. E. (Begolli & Richland, 2013) Visual Support for Instructional Analogy: Context Matters. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1863 - 1868). Austin, TX: Cognitive Science Society.

LIST OF FIGURES

<i>Figure 1.</i> A process overview of using video to create experimental manipulations.	18
<i>Figure 2.</i> Still images illustrating the experimental conditions created through video editing. From left to right, the first picture shows only the teacher while obstructing the writing on the whiteboard (Not Visible condition), the middle picture shows only the most recent problem (Part Visible condition), and the picture that shows the whole board (All Visible condition).....	20
<i>Figure 3.</i> A layout of the cameras and microphones in the classroom used for recording the lesson.....	24
<i>Figure 4.</i> Estimated marginal means across both posttests on procedural knowledge, procedural flexibility, conceptual knowledge, and use of the misconception by condition . Error bars are standard errors.....	49
<i>Figure 5.</i> In the picture on the left the cat (both the chaser and the chased) corresponds relationally to the boy in the picture on the right (both the chaser and being chased). A common error young children make is to be distracted by the perceptual cues matching the cat in the left with the cat in the right.....	57
<i>Figure 6.</i> The hearts & flowers task. Congruent trials are shown on the left and incongruent trials are shown on the right (Wright and Diamond, 2014).....	65
<i>Figure 7.</i> The Stop-Signal “game” instruction screen. The task of the students is to press the corresponding key quickly enough to “send” the fish home, but to not press the key if the manta ray fish appears. The manta ray appeared at random on 40% of the trials. Adapted from Bissett & Logan (2012).....	66
<i>Figure 8.</i> Mean scores for Pretest (PT), Immediate (IT) and Delayed (DT) tests by WM (left) and IC (right) score.....	73

LIST OF TABLES

Table 1. <i>The Differences Between Laboratory Experiments and Design-Based Research, adopted from Collins (1999)</i>	11
Table 2. <i>Problem solved and discussed during videotaped lesson</i>	40
Table 3. <i>Three student solutions compared during the videotaped instructional analogy lesson</i>	41
Table 4. <i>Solution Strategies Produced Problems, by Condition</i>	47
Table 5. <i>Student Scores for Experiment 1, by Condition</i>	48
Table 6. <i>Analyses of Covariance Results on Learning Outcomes</i>	49
Table 7. <i>Pairwise Comparisons, by Condition with Pretest & Teacher as Covariates for both Posttests</i>	50
Table 8. <i>Summary of Mann-Whitney U Pairwise Comparisons of the Common Misconception</i> ..	50
Table 9. <i>Inter-item alpha values at each test for each construct</i>	62
Table 10. <i>Illustration of the Forward Digit Span. The Backward Digit Span works in the same fashion, but the numbers need to repeated in backward order.</i>	64
Table 11. <i>Factor Loadings and Descriptives of Cognitive Tasks</i>	68
Table 12. <i>Regression models conducted in analyses. A separate regression was conducted for each mathematical construct</i>	69
Table 13. <i>Mean scores per construct, SD in parenthesis</i>	70
Table 14. <i>Beta values from regression models described in Table 1. Pretest beta vales were always significant, $p < .05$; not shown here.</i>	71
Table 15. <i>Effect sizes for each predictor from Model 1, on each mathematics outcome, shown as partial eta-squared</i>	72

CURRICULUM VITAE

KRESHNIK BEGOLLI

kbegolli@uci.edu - 310.266.3266
6314 Adobe Circle South, Irvine, CA, 92617

EDUCATION

- 2015 **PhD University of California, Irvine**
Education
Emphasis in Learning, Cognition, & Development
Thesis: Role of Executive Functions and Visual Supports for Overcoming
Misconceptions through Structure-Mapping in Mathematics Classrooms
- 2013 **MA University of California, Irvine**
Education
- 2010 **BS University of California, Los Angeles**
Cognitive Science, Specialization in Computing (major)
Linguistics (minor)
Psychology Departmental Honors
Thesis: Perceptual Learning of Word Roots
Advisor: Dr. Phillip Kellman
College Honors

AWARDS AND HONORS

- 2013 – 2014 **Community for Advancing Discovery Research in Education – CADRE Fellow**
CADRE is the resource network for NSF's discovery research K-12 program.
One of ten awarded nation-wide.
- 2008 – 2010 **The Merage Foundation for the American Dream Fellowship**
One of twelve awarded nation-wide with submissions from 22 Universities
- 2008 – 2010 **Berkman Scholar**, UCLA, Two-Year Full Tuition Scholarship
The only Los Angeles City College student and one of two UCLA-wide
- 2009 **Dean's Prize Award**, Westwind/Aleph Conference, UCLA
- 2009 **Psychology Research Opportunity Program**, University of California, Los Angeles
- 2008 **Ralph Bunche Scholar**, Los Angeles City College
- 2008 **Luise Snyder Johnson Fitzgerald Scholarship**, Los Angeles City College
-

PUBLICATIONS

Begolli, K. N., Richland, L. E. (2015). Teaching Mathematics by Comparison: Analog Visibility as A Double-Edged Sword. *Journal of Educational Psychology*.

Richland, L.E., Frausel, R. R., **Begolli, K. N.** (forthcoming, 2015). Cognitive Development, *The SAGE Encyclopedia of Theory in Psychology, Harold (Hal) L. Miller*.

Richland, L. E., **Begolli, K. N.**, Naslund-Hadley, E. (forthcoming, 2015). Current Theories of Children’s Mathematical Development. Chapter in “The Inter-American Development Bank and Technology in Education: How to Promote Effective Teaching?” The Inter-American Development Bank.

In preparation
(draft
available)

Begolli, K.N., Richland, L. E. (Submitted). Bridging Cognitive Experimentation and Real Classrooms: A Video Methodology.

Begolli, K. N., Richland, L.E., Jaeggi, S. (In preparation for submission). The Role of Executive Functions for Structure-Mapping in Mathematics.

McDonough, I., **Begolli, K. N.**, Richland, L. E., Kornell, N. (In preparation for resubmission) Limiting Feedback Increases Study Time and Metacognitive Accuracy.

In preparation
(data collected)

Begolli, K. N., Richland, L.E., Jaeggi, S. (data collection). Teaching the Wrong Solution First or Last: Children’s Mathematics Outcomes Depend on Individual Differences in Executive Functions.

*supervised
undergraduate
co-author

Herd, B.*, Sayonno, H.*, **Begolli, K. N.**, Jaeggi, S., Richland, L.E. (In preparation for submission). Stereotype Threat: Its Effects On Female Learning from Instructional Analogies.

Herd, B.*, Sayonno, H.*, **Begolli, K. N.**, Jaeggi, S., Richland, L.E. (data collection). Stereotype Threat, Executive Functions, and Teaching Scaffolds: The Effects, Mechanisms, and Aids for Learning from Instructional Analogies.

PEER-REVIEWED CONFERENCE PRESENTATIONS

Papers &
Symposia

Begolli, K. N., Richland, L.E., Jaeggi, S. (2015, July). The Role of Executive Functions for Structure-Mapping in Mathematics. Paper accepted at the Cognitive Science Society Conference, Pasadena, USA.

Richland, L. E., **Begolli, K. N.**, & Frausel, R. (2015, April) "Relational Thinking in Elementary Mathematics" Symposium accepted at the Twenty Seventh Annual meeting of the American Psychological Society, Chicago, Illinois, USA

Posters

Richland, L. E., **Begolli, K. N.**, & Frausel, R. R. (2014, May). Using the Internet as a Laboratory for Research in the Real World, Symposium Presented at the Twenty Sixth Annual meeting of the American Psychological Society, San Francisco, California, USA.

Begolli, K. N. & Richland, L. E. (2013, August) Visual Support for Instructional Analogy: Context Matters. Paper presented at the Cognitive Science Society Conference, Berlin, Germany.

Begolli, K. N. & Richland, L. E. (2013, April). Out of Sight, Out of Mind: Unloading Working Memory to Facilitate Mathematical Comparisons. Paper presented at the biennial meeting of the Society for Research in Child Development, Seattle, Washington, USA.

Begolli, K. N. & Richland, L. E. (2013, June). Out of Sight, Out of Mind: Unloading Working Memory to Facilitate Mathematical Comparisons. Paper presented at the Fourty Third meeting of the Jean-Piaget Society, Chicago, Illinois, USA.

Begolli, K. N., Richland, L. E., Jaeggi, S. (2015, March). Individual Differences in Inhibitory Control and Misconceptions in Instructional Analogy. Poster presented at the biennial meeting of the Society for Research in Child Development, Philadelphia, Pennsylvania, USA.

Begolli, K. N. (2014, July). Teacher's Use of Gesture and Stereotype Threat in Classrooms. Poster presented at the meeting of the International Society for Gesture Studies, San Diego, California, USA.

Richland, L. E., **Begolli, K. N.**, & Frausel, R. R. (2014, May). Drawing Connections Under Pressure in the Mathematics Classroom: Linking Gestures can Help, Poster Presented at the Twenty Sixth Annual meeting of the American Psychological Society, San Francisco, California, USA

Begolli, K. N. (2013, August). Using Video to Bring Analogy Research to Classroom Settings. Poster presented at the International Conference on Analogy, Dijon, France.

Begolli, K. N. & Richland, L. E. (2012, May). Out of Sight, Out of Mind: Unloading Working Memory to Facilitate Comparisons. Poster presented at the Twenty Fourth Annual Convention of the Association for Psychological Science, Chicago, Illinois, USA.

Begolli, K., Mettler, E., Kellman, P. (2009) Learning by Comparison: Extending Perceptual Learning to Vocabulary Learning. Paper presented at the Stanford Psychology Undergraduate Research Conference. (Paper also presented at UCLA's, and Whittier's Psychology Undergraduate Research Conferences 2009).

ABSTRACT OF THE DISSERTATION

The Role of Executive Functions and Visual Supports for Overcoming Misconceptions through
Structure-Mapping in Mathematics Classrooms

By

Kreshnik Nasi Begolli

Doctor of Philosophy in Education

University of California, Irvine, 2015

Assistant Professor, Susanne Jaeggi, Chair

Comparing different student solutions to a single instructional problem is a key recommended pedagogical tool in mathematics and for inducing conceptual change. However, these instructional comparisons can overload children's cognitive resources, and may be more demanding for some children than others. In my dissertation, I examine the relations between individual differences in children's cognitive resources and their thinking and learning from instructional comparisons. Individual differences in children's resource capacities provide a useful lens for understanding the contributions of cognitive mechanisms in learning from comparisons. Manipulating the kinds of instructional representations and sequence of presentations provide further insights on how to best manage students' limited processing resources. The dissertation has three main chapters. The first discusses a novel methodology for deriving causal relationships through stimuli in which the complexity of everyday classrooms is held constant. The second describes Experiment 1, which tested the role of visual representations when learning from comparing three solution strategies (one incorrect and two correct). In this experiment, fifth graders watched a mathematics video-lesson on ratio edited to either: a) present solutions only orally; b) present only the most recent solution visually; and c)

present all solutions visually. Visual representations served as a double-edged sword, where simultaneous presentations led to durable, conceptual understanding, whereas presenting solutions sequentially as in condition b) led to the poorest learning outcomes possibly reflecting limitations in cognitive resources. This result is further examined in the third chapter, which reports on Experiment 2. This experiment replicates only condition b) of Experiment 1, where solutions strategies were presented sequentially, but with a larger sample of fifth graders who were also administered four cognitive measures. The results of Experiment 2 reflect a positive relationship between children's learning and their cognitive control capacities. Finally, the theoretical and practice relevant implications of these results and future directions are discussed.

The Role of Executive Functions and Visual Supports in Instructional Analogy

Comparing different student solutions to a single instructional problem is a key recommended pedagogical tool in mathematics (Common Core State Standards Initiative, 2010; National Mathematics Advisory Panel, 2008; National Research Council, 2001) and for inducing conceptual change (Brown & Clement, 1989; Dietrich, 2000; Vosniadou, Vamvakoussi, Skopeliti, & Vosniadou, 2008). However, the cognitive underpinnings of successfully completing this task are complex. Students must perform relational structure-mapping: represent the multiple solutions as systems of relationships, align and map these systems to each other, and draw inferences based on the alignments (and misalignments) for successful schema formation (see Gentner, 1983; Gick & Holyoak, 1983; Richland, Zur, & Holyoak, 2007). Schemas can be thought of as abstract mental funnels that categorize relations between thoughts (e.g. the schema of “taking away” in mathematics). Structure-mapping is posited to underlie the processes of analogical reasoning where one familiar source representation (e.g. Besa lost an apple), set of representations (e.g. multiple people lose apples), or a schema (e.g. “having less of something” in general) is mapped to a less familiar target representation (e.g. $x - 1$), set of representations (e.g. $nx - my$) or a schema (e.g. “subtracting” in general; Gentner, 1983; Richland, Stigler, & Holyoak, 2012).

Orchestrating classroom lessons in which learners successfully accomplish relational structure mapping is not straightforward, particularly because opportunities for learning through structure mapping often fail in laboratory contexts (e.g., Gick & Holyoak, 1983; Ross, 1989). Specifically, reasoners regularly fail to notice the utility of aligning and mapping two or more available relational structures, often leading to misconceptions (Spiro, Feltovich, Coulson, & Anderson, 1989; Zook & Di Vesta, 1991).

The low success rate with which participants notice and use relational structure mapping, or analogy, within laboratory studies to solve problems may in part reflect limitations in executive functions (EFs; see Waltz, Lad, Grewal, & Holyoak, 2000). Executive functions are a set of related general-purpose processes, which share *commonalities*, but also have *diverse* functions, for controlling thought and behavior (Miyake & Friedman, 2012). In structure-mapping, executive functions are required to relationally represent systems of objects (e.g. steps to solution strategies) to re-represent these objects as systems of relations so that their structures can align and map together, to identify meaningful similarities and differences, and to derive conceptual/schematic inferences that better inform future problem solving (Morrison et al., 2004).

The aim of this dissertation is to examine the relationship between distinct processes in EFs and specific teaching strategies that support learning mathematics through instructional opportunities for structure mapping. This proposal draws from basic studies in cognitive science literature of analogical reasoning which have given rise to the discovery of key external factors (e.g. keeping source and target visible simultaneously) and cognitive mechanisms (i.e. executive functions) implicated in increasing/decreasing the likelihood of successful schema formation from structure mapping (Gick & Holyoak, 1983; Morrison, Dumas, & Richland, 2006). While these findings have important implications for classrooms, these literatures have avoided experimentation in in these contexts in order to maintain high experimental control, because naturalistic settings, such as classrooms, are ripe with variability. But, this has also limited the generalizability of their findings for instructional practice.

As a result, more research is needed to discover key instructional practices that help teachers orchestrate lessons around opportunities for structure mapping and how cognitive

mechanisms operate in *in vivo* classroom settings using educationally relevant materials (Richland, Linn, & Bjork, 2007). Recent studies (Richland & McDonough, 2010; Richland & Hansen, 2013; Rittle-Johnson & Star, 2009) have alluded that the effectiveness of teaching strategies may be explained in part because they provide EF offloads for the reasoner, freeing up necessary processing resources for structure mapping. In support of these interpretations, empirical data suggest that populations with limited EF capacity perform worse on analogy tasks (e.g. studies with children, Richland, Morrison, & Holyoak, 2006; TBI adolescents, Krawczyk et al., 2010; older adults, Viskontas, Morrison, Holyoak, Hummel, & Knowlton, 2004; and brain damaged adults, Morrison et al., 2004; for a computational model see Morrison, Dumas, & Richland, 2011).

To gain a better understanding about the EF processes that underlie reasoning from an analogy based lesson, in this dissertation I will draw from literature and methodology that attempts to isolate the contribution of EF on academic performance by investigating individual differences in children's EF. Specifically, this literature examined differential advantages/disadvantages within children's particular EF components, such as inhibitory control (IC) and/or working memory (WM), and their relationship to advantages/disadvantages in academic performance (Diamond, 2013; Gathercole, Pickering, Knight, & Stegmann, 2004). One purpose of the current work is to extend such studies and examine the role of IC and WM on *learning* from a single classroom lesson, in contrast to performance on academic achievement tests.

Within this realm, students' prior knowledge plays a critical role in the learning process, which can have detrimental effects when their conceptual intuitions conflict with normative mathematical concepts (Brown & Clement, 1989). This increases the complexity of identifying

relevant instructional practices for teaching through analogies, because teachers also need to account for students' common misconceptions when providing students with opportunities for structure mapping. To gain a better grounding on how students' misconceptions play a role in the structure mapping process, the current work also draws from a second literature, the body of educational research exploring how students undergo conceptual change (Clement & Brown, 2008; Vosniadou et al., 2008). This literature posits that people must first elicit and explore their intuitions about concepts in order to change them. Thus, the stimuli used in the studies incorporate a common misconception to challenge students' incorrect intuitions.

Research Overview

Two strands of research run throughout this dissertation. One strand seeks to gain specificity about teaching strategies that facilitate students' schema formation through alignments (or accounts for their misalignments) between structural relations in mathematics. The other strand intends to discover the underlying cognitive mechanisms of learning from analogies and misconceptions in classroom settings. In order to investigate both of these strands of research, I will present a research methodology that has high implications for cognitive scientists and education researchers to derive causal relationships about teaching practices and cognitive mechanisms with high internal and external validity.

Classrooms are vibrant, complex environments in which the high level of unexpected variability makes experimental control often impossible (Brown, 1992). The overarching commitment to controlled manipulation of experimental contexts within psychological research has led much cognitive scientific study of learning behavior to be conducted in controlled laboratory settings. While in some ways this model leads to the production of data that can be

easily interpreted (x behavior derived from y manipulation), the meaningfulness of these results for educational practice has been less clear. Theoretically, this research epistemology has also meant that the search for universal cognitive processes of learning can best be accomplished through the design and examination of cognition within atypical, impoverished environments (see Shweder, 2012). Thus, the assumption that cognitive mechanisms underlying classroom learning are not moderated by environmental factors is unexplored. In this dissertation I will not investigate that question but rather in my studies I hope to reduce the assumption by situating the stimuli creation in the naturalistic classroom context itself.

In the first chapter, a video methodology is proposed as a new model for experimental designs that aim to bridge education and cognition, enabling experimental control while testing in real classroom contexts. Video recordings of live lessons can be systematically video-edited to create multiple versions of a single lesson to be used as instructional stimuli, in which only one aspect of the lesson is varied while otherwise maintaining constancy of contextual factors that have been shown to impact learning, such as instructional discourse, gestures, student participation, emotional valence, or background noise. These matched versions of the videotaped lesson can then be used as stimuli for teaching new samples of students. When delivered on individual computers, students can be randomized by condition within classrooms to meet a gold standard of random assignment. This technology provides opportunities for cognitive scientists and education researchers to extend their research and derive causal relationships about teaching practices through stimuli that approximate a real lesson. This approach is utilized in the studies presented in this manuscript.

In the second chapter, I will describe Experiment 1 which draws from an important mode for developing flexible mathematical thinking and a high leverage recommendation in education:

comparing multiple solutions to a single problem, Yet, instructionally leading this activity is challenging (Stein, Engle, Smith & Hughes, 2008). I test one decision teachers must make: whether to have students describe solutions orally, to have students show their solutions one after another, or to maintain all solutions visible throughout a discussion. Sixth grade students were presented with one of three versions of a videotaped mathematics lesson on ratio. The video was manipulated to create three versions with constant audio but in which the compared solutions were a) presented only orally, b) visible sequentially in the order they were described, or 3) all solutions were visible after being described throughout the discussion. Pretest, posttest, and delayed posttest measures assessed 11-12 year old children's learning, revealing that making all solutions visible throughout the discussion led to the highest gains in all measured aspects of mathematics knowledge: procedural knowledge, reduced use of misconceptions, procedural flexibility, and conceptual/ schematic knowledge. Interestingly, showing students visual representations of solutions sequentially led to the lowest gains and highest rates of misconceptions, suggesting that visual representations may be powerful but must be visible simultaneously in order to facilitate full understanding of the comparisons. Structure-mapping failures may reflect a cognitive overload of children's executive functions (EF).

In the third chapter, I describe Experiment 2. In Experiment 2, I examine the role of individual differences in EF resources for learning from an everyday mathematics video-lesson placing a particular strain on children's cognitive resources: comparing three analogs presented sequentially. Specifically, I examined the separate contributions of working memory (WM) and inhibitory control (IC) on successful schema-formation. Overall, WM and IC explained distinct variance for predicting improvements in procedural knowledge, procedural flexibility, and conceptual knowledge after a 1-week delay. WM & IC are less predictive at immediate post-test,

suggesting that these functions are not simply correlated with mathematics skill, but may be particularly important in the process of structure-mapping for durable schema-formation. These results inform the literature on both analogy and mathematics *performance* implicating EFs as key for successful structure-mapping, and extend them to an ecologically valid *learning* context.

The three studies will be described in more detail below. I begin with Study 1, where the motivation behind the video methodology used in Studies 2 and 3 of the dissertation will be explained, and then the methodology itself articulated explicitly.

Chapter 1: Bridging Cognitive Experimentation and Real Classrooms: A Video Methodology for Experimental Research in Education

Bridging experimental psychological studies with classroom needs, interests, and contextual dynamics is challenging. In the past decade, there has been a surge of laudable experimental work in cognition that has shifted from the laboratory to the classroom (Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012; Klahr & Li, 2005; Koedinger, Aleven, Roll, & Baker, 2009; Rittle-Johnson & Star, 2009; Roediger & Pyc, 2012; Schwartz & Bransford, 2005; for review see Mayer, 2008; Richland, Linn, et al., 2007; Pashler et al., 2007), but controlled examinations of instructional manipulations executed by a teacher are less frequent (for critiques of existing methodologies, see: Design-Based Research Collective, 2003; McCandliss, Kalchman, & Bryant, 2003 and often misunderstood (Klahr, 2013)). While teachers continue to play the most important role in everyday classroom instruction, many psychological experiments seek to derive abstractions by manipulating software or paper executed strategies, often minimizing the role of the teacher in the process (Daniel, 2012), due to concerns about control and reliability. The aim of conducting carefully controlled work in dynamic classrooms in ways that both inform teaching and the underpinning cognitive mechanisms poses difficult methodological challenges in terms of preserving high internal and external validity.

I describe a methodology for using videotaped classroom instruction as experimental stimuli, which has the potential to make great strides toward this aim. Edited videotapes are used as a basis for experiments in which new students learn from the recorded, and thus well-controlled, instruction. Video provides an efficient, reliable, and relatively inexpensive technological tool for creating experimental stimuli that approximate the real world. Video

stimuli can serve to answer both applied and theoretical questions, since it provides a mode for standardizing classroom-relevant instruction across conditions.

Further, the methodology can provide both high internal and external validity. When conducting research to both inform theory and practice, there is often a trade-off made between external validity and internal reliability. I posit that video technologies provide an opportunity to create controlled stimuli embedded within the contextual variability that is integral to real classroom lessons, making this tradeoff less severe. Many of these hard-to-control features of the classroom context have been shown to impact learning, such as teacher and student gesture, (e.g., Alibali, Flevares, and Goldin-Meadow, 1997), or affect (Pardos et al., 2013). Thus, making experimentally-derived causal claims about classroom learning principles without incorporating these multi-faceted aspects of classroom contexts in the stimuli may be failing to account for the complexity of the interrelationships between cognitive principles. This excluded variability may also in part explain regular failures for laboratory-based principles to generalize with large effect sizes to classrooms (Dunlosky and Rawson, 2012; Donovan, 2013). These challenges in part echo the concerns of many cognitive scientists whose ultimate goal is to understand the workings of the brain in the real world, but who are unable to account for the interrelated, highly complex brain activation that occurs in response to dynamic, real-world contexts (e.g. for discussions on the topic in memory see Kvavilashvili and Ellis, 2004; in attention see Kingstone et al., 2003).

An alternative model of studying learning in classroom contexts is to use iterative designs to formulate, test and refine optimal curriculum and instructional stimuli. These projects embrace the complexity of the classroom context and are much better able to describe the complex interplay between an instructional manipulation and student thinking. At the same time,

this work is not designed to maintain internal reliability, and thus causal claims about the efficacy of instructional choices are not part of the theory building enterprise.

I describe a mode through which video may be used to bridge these endeavors, such that it enables the researcher to design experimental materials that are well controlled but that also include more of the ecological valid variability that is part of everyday classroom learning contexts. The goal is to more fully meet Brown's (1992) vision to "traverse between the real world and the laboratory" and understand the causal elements arising from qualitative designs (e.g. teaching strategy "x" causes change in student learning "y"; Design-Based Research Collective, 2003; McCandliss, Kalchman, and Bryant, 2003).

In sum, the approaches of experimental psychologists and education scientists have been generally limited in either external or internal validity, respectively, which echoes the concerns of many cognitive scientists attempting to understand the workings of the brain in the real world (e.g. for discussions on the topic in memory see Kvavilashvili & Ellis, 2004; in attention see Kingstone, Smilek, Ristic, Friesen, & Eastwood, 2003). Overall, researchers have been generally sitting on two opposite sides of the same coin, as outlined in Table 1. In this chapter I revisit the opportunities that video-based stimuli afford for bridging the methodological gap between these two sides to derive cause-and-effect relationships, which examine the workings of the human brain in classrooms and have direct implications for teaching practices.

Table 1
The Differences Between Laboratory Experiments and Design-Based Research, adopted from Collins (1999).

Traditional Laboratory Experiments	Design-Based Experiments
Learners focus on stimuli without interferences or disturbances with controlled materials (stimuli) presented in a standardized manner	Design-experiments are carried out in “messy” classroom settings, characteristics learning in the real world
One or two variables of interest	Multiple dependent variables
Controlling the research setting	Characterizing the situation
Theory based manipulations of design	Iterative revisions of design based on student responses
Participants are isolated	Participants are in a social setting
Testing hypothesis	Characterizing a profile of the design
Experimenter only	Co-participations in research design and analysis

This chapter will describe a model for creating video-based stimuli for instructional experiments. I begin by explaining the rationale for studying cognition in the classroom context and provide a brief overview on the usage of video to capture realistic settings. Afterwards, I walk the reader through the process of the methodology and present a case study (Begolli & Richland, 2015) to provide the technical details involved in video-recording and video-editing and conclude the chapter with the limitations of video-stimuli. Finally, the benefits of the video methodology as an alternative are discussed with respect to current research approaches to suggest how this methodology could overcome the limitations in either internal or external validity to inform teaching practice and cognitive theories.

The Importance of Capturing the Environment in Stimuli

A substantial number of findings arising from laboratory contexts have revealed cognitive processes that have potential to generalize to real world contexts. Nevertheless, researchers are usually very careful to limit the generalizability of their findings within the context being studied and for good reason (see Duncan & Owen, 2000 for review on brain imaging studies tying stability of cognitive processes with context). Ecological psychology studies have established that reasoning processes are grounded in the modality, action, and perspective of the thinker and the environment (Carragher, Schliemann, & Carragher, 1988; Lave, 1988; Saxe, 1988; Suchman, 1987), where the most minor change within a laboratory context could compromise the replicability of an effect (e.g. Atchley & Kramer, 2001; Berry & Klein, 1993; Bindemann, Mike Burton, & Langton, 2008). Considering the few published works on failed replications, this represents as Kingstone, Smilek, & Eastwood, (2008) put it “a small tip of a very large iceberg of failed replications ... that are never published.” However, a recent study by Klein et al., (2014) which attempted to replicate 13 psychological effects across 36 independent samples in either lab or online contexts is informative. While only 10 out of 13 effects were replicated, they found that the *effect* itself is a better predictor than the population sample or the setting (laboratory or online). This work suggests a great promise for the generalizability of laboratory findings to online administration, across different populations, though it is administered on WEIRD populations (western, educated, industrialized, rich and democratic; Henrich, Heine, & Norenzayan, 2010), with 28 out of 36 samples consisting of mostly native English speakers. This and future phases of the Klein et al., (2014) and the Open Science Collaboration (2012) projects are important steps at understanding whether sources of variability exist between populations and the setting for reproducing psychological effects. Yet,

in contrast to the current study, the projects examine variability regarding *who* is presented with stimulus and *where* the stimulus is present, but it does not attempt to transform the stimulus itself from static, text-based (e.g. written vignettes used in Klein et al., 2014) into stimuli that better approximate a realistic event (see Monin et al., 2014 for a similar critique about lack of stimulus variability & sampling). Others have argued unrealistic and/or unreliable dependent measures, analyses tools, and properties of data lead to the misinterpretation of results (Rotello, Heit, & Dubé, 2014). Thus, I applaud Klein et al.'s approach, and stress that researchers need not only be aware of *who* is presented with stimuli and *where* the stimuli is presented, but critically whether the stimuli themselves generalize by modeling a realistic event. These concerns continue to challenge the epistemic value about the generalizability of laboratory findings to realistic environments, such as classrooms.

In this vein, there is a growing number of theoretical accounts emphasizing that cognition emerges as a result of the interaction between environmental factors and the thinker (e.g. Gibson, 1979) which have been supported in the fields of embodied (e.g. Barsalou, 1999), ecological (e.g. Reed, 1996), distributed (e.g. Hutchins, 2000; Clark & Chalmers, 1998), situated (Greeno, 1998), and socio-cultural cognition (e.g. Bender & Beller, 2013; Correa-Chávez & Rogoff, 2005). Inherent in these ideas is that the environment-thinker interactions entail events occurring both in the thinker and the environment. As such, there is mounting evidence that cognitive processes in basic perception (e.g. Gibson, 1979), attention (e.g. Simons & Chabris, 1999), and memory (Barnier, 2012; Kvavilashvili & Ellis, 2004) operate differently in response to an event (in this case a classroom lesson) than in response to stationary stimuli.

One common difference between laboratory and real world contexts for cognition is the static versus active nature of the stimuli used to prompt a studied cognitive change. Laboratory

as well as controlled classroom experiments frequently use static images or written text to provide a highly controlled stimulus prompt in order to more carefully examine the cognitive responses to variations in the stimulus. At the same time, this raises several problems. First, in a laboratory, one may find that processing for static materials is quite different and not generalizable to the dynamic, more complex processing that unfolds when learning stimuli are active lesson materials in a context such as a classroom. Secondly, even when static materials are presented in a controlled way, providing them in a dynamic setting such as a classroom means that the experimenter then cannot control what the reasoner does in response to these materials, and how they talk or act to engage with them.

In contrast, video-based stimuli afford a better approximation of real world events and in turn a higher likelihood of generalization of educationally relevant findings to teaching practices in classroom contexts. Using videotaped classroom lessons as stimuli means that the materials are dynamic, involving complex linguistic and interactional information in addition to the cognitive prompts, they carry normative classroom cultural information, and they carry dynamic social and perceptually rich information so they are more likely to stimulate more of the complex cognitive work engaged when students are learning in an everyday classroom. Thus while they can be used to manipulate specific cognitive factors, as will be described below, they can also more closely approximate classroom learning so require less of a leap for generalization. Additionally, they can be administered in one-on-one computer delivery, meaning that research subjects' participation can be highly controlled.

For many reasons, therefore, videotaped stimuli provide a rich methodology for bridging classroom studies and controlled experimentation. The idea of using video to capture the real world is not novel in behavioral experiments (Gibson, 1947) and education research, thus before

presenting the methodological approach, I present a brief review of video usage in observational and experimental studies.

From Video-Based Observations to Video-Based Experimentation

Video-based experimental stimuli have had widespread use for examining human and animal behavior, most pervasively in studies of visual perception where video has been deemed a favorable as stimulus for studying visual processing of motion, shape, texture, size, and brightness (Gibson, 1947; Oliveira et al., 2000; Webb, Knott, & MacAskill, 2010). Other scientists have used it to investigate, for example, social behavior (Soble, Spanierman, & Liao, 2011), gestures (Alibali et al., 1997; Cook, Duffy, & Fenn, 2013; Valenzeno, Alibali, & Klatzky, 2003), health and clinical training (Gaba, 2004; Heath, Luff, & Sanchez Svensson, 2007; Lyden et al., 1994), aggression (Kilduff, Hopp, Cook, Crewther, Manning, 2013), animal behavior (D'Eath, 1998), and autism (Marcus & Wilder, 2009).

In education, however, video has been mostly employed as a data collection tool for the purposes of observing teacher or student behaviors (e.g. TIMSS, 1999). In fact, most published guides on the use of video for research in education revolve around methods for how to conduct careful observations in classrooms, such as selecting appropriate video segments, conducting video analysis, and developing descriptive video coding schemes (Derry, 2007; Derry et al., 2010; Goldman, 2007; Santagata, Gallimore, & Stigler, 2005). Such observational videos, for example, have been used to create video clubs where teachers reflect on their teaching (van Es & Sherin, 2008) to investigate reflections between expert and novice teachers (Rich & Hannafin, 2008), to assess teacher knowledge (Kersting, 2008) and efficacy (Hill, Ball, Blunk, Goffney, &

Rowan, 2007), and to examine classroom behavior in children with attention-deficit hyperactivity disorder (Lauth, Heubeck, & Mackowiak, 2006).

Observational studies have revealed important questions about many of the details of interactions within a classroom, which may causally affect student learning such as teacher and student gestures (e.g. Alibali, Flevares, and Goldin-Meadow, 1997; Richland, Zur, & Holyoak, 2007). The role of teacher and student gestures on learning is one such example. Many gesture researchers have developed experimental designs that attempt to make causal claims about the correlational and qualitative data arising from observational studies (Alibali et al., 1997; Cook et al., 2013; Sueyoshi & Hardison, 2005; Valenzeno et al., 2003).

In such studies, the stimuli are recorded at two different time points, but share the same audio (Cook, Duffy, and Fenn, 2013) or multiple cameras record the same event from different perspectives either capturing gestures or not (Sueyoshi and Hardison, 2005). These approaches exploit the main advantage of video – capturing realistic details of teacher-student interactions (e.g. gestures, affect, etc.) while keeping at least parts of the event in the stimulus constant across participants, and randomly assigning participants to different video versions. However, this approach has largely been overlooked as a method for testing the effect of a particular teaching practice on student learning within a classroom setting. This could be due to the traditional difficulties of conducting research in schools and embedding research questions within a full classroom lesson (for exceptions see Richland and Hansen, 2013; Begolli & Richland, 2015).

To promote the discovery of learning principles that optimize teaching practices, a methodology for creating video-based stimuli that could contribute to the experimental study of scientific hypotheses with high internal and external validity is presented next.

A Video Methodology for Bridging Cognitive Research with Teaching Practices

Video maximizes external validity because it is a better approximation of real world settings, such as a classroom, than paper or computer-based stimuli. Video-lessons maintain high internal validity because: a) video remains unchanged, b) video-recordings of a single lesson can be systematically edited to create two versions of a stimulus (e.g. students either see or do not see a teacher's gestures), c) two or more versions of a video-lesson can be randomized within two student populations. Video-lessons maximize external validity because they approximate a true classroom experience. The proposed method for employing video-lessons to maintain experimental control is embedded in a methodological approach illustrated in Figure 1. The methodological process begins with a hypothesis, which may come from video or field observations of classroom practice, shared professional knowledge of teaching practices, basic research, or a combination of these. For instance, a specific hypothesis could be: should the teacher keep multiple solutions to a single problem visible on the board throughout the lesson? Afterwards, the researcher collaborates with a teacher to co-design a lesson that becomes the basis for the stimulus of each experimental condition (e.g. the teacher conducts the lesson while leaving all solutions visible on the board). In this manner, theoretical and/or practical questions are embedded in a teacher-guided lesson.

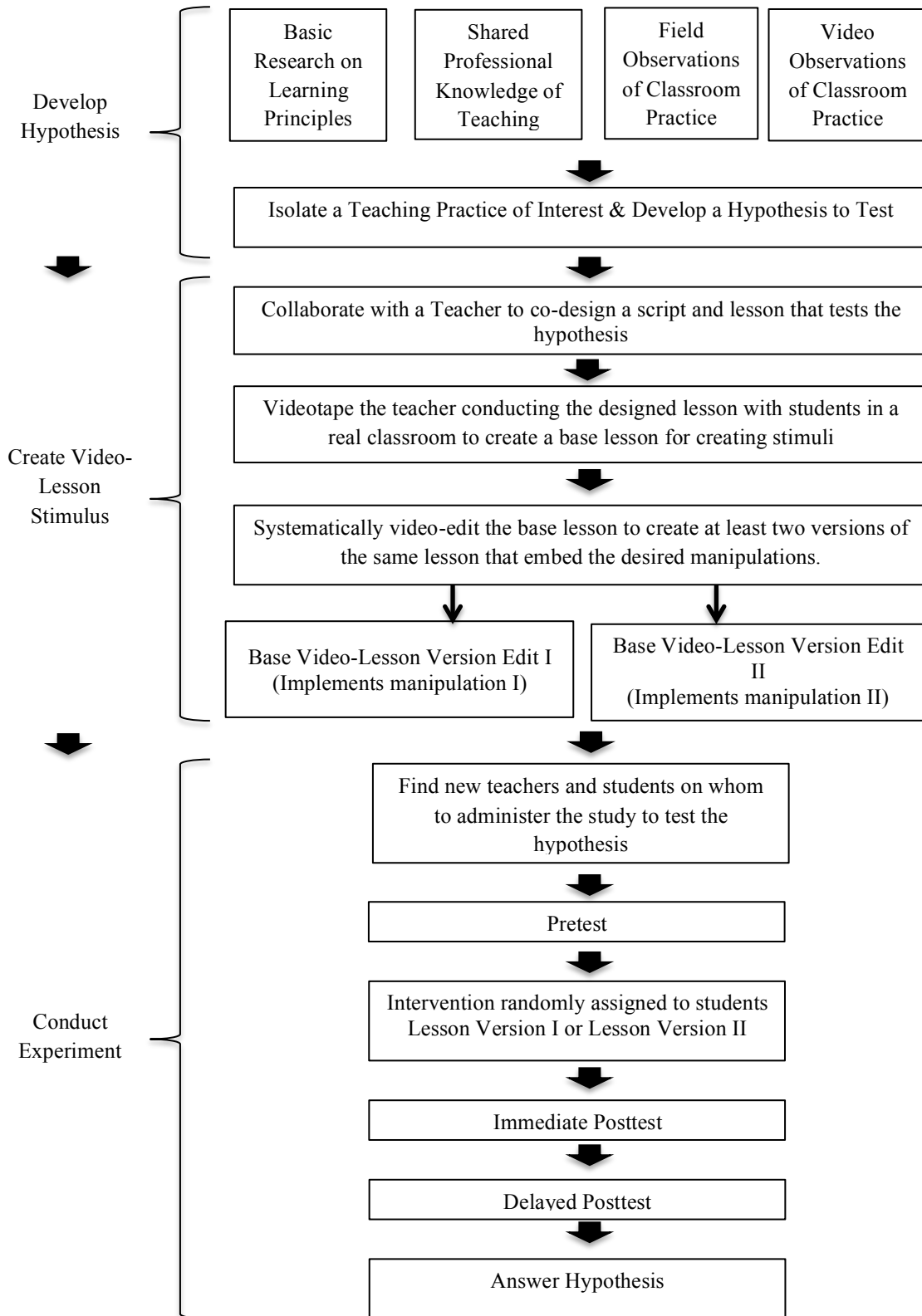


Figure 1. A process overview of using video to create experimental manipulations.

Based on the lesson script, the teacher conducts the designed lesson in her classroom while the researchers videotape it, capturing the complexity of a real classroom including spontaneous student-teacher interactions. To avoid confusion the classroom will be called the stimulus-creation classroom because it is only used for creating a base video-lesson. This video-lesson maximizes the representation of a natural teaching and learning environment, and maximizes external validity of the stimuli. The recording itself is done in a manner that allows for later video-edits to create stimuli that will become the basis of experimental manipulations.

Next, the recorded video-lesson is systematically video-edited to create two or more *identical* instructional-videos with a single systematic difference – the manipulation of interest. The result is two or more versions (e.g. Version I - solutions are visible on the board and Version II - solutions are not visible on the board) of the same instructional-video that approximate a real classroom experience. This phase completes the stimulus creation of the video-lesson.

In a different school, new teachers are recruited and the new teachers' students are administered a pretest. These are intervention-classrooms. Using schools' computer labs, or small netbooks, each student or students with partners, can be randomized within classrooms to either watch video-lesson Version I (e.g. solutions visible) or Version II (e.g. solutions not visible), while remaining within the natural classroom environment. This is followed by an immediate posttest then a delayed posttest of students' learning outcomes. Finally, the outcomes of each condition are analyzed to answer the hypothesis of interest. This methodology is a direct extension of work from Richland and colleagues (Richland & McDonough, 2010; Richland & Hansen, 2013) who have utilized video-lessons recorded in the laboratory to test the presence or absence of a combination of teaching strategies on student learning. The proposed methodology builds on this work by recording real classrooms to capture teacher-student interactions, instead

of only the teacher or experimenter; and tests a single teaching strategy, instead of a combination of strategies.

Video-Editing for Creating Experimental Conditions that Test Efficient Classroom

Techniques: Case Study

A recent study by Begolli & Richland (2015), discussed in chapter 2, illustrates the potential of video editing for examining questions that lead to causal inferences about specific teaching practices. They examined the benefit of making student responses visible during a math lesson. This chapter will illustrate how to create different experimental conditions from a single lesson through systematic video-editing techniques to vary only one teaching practice.

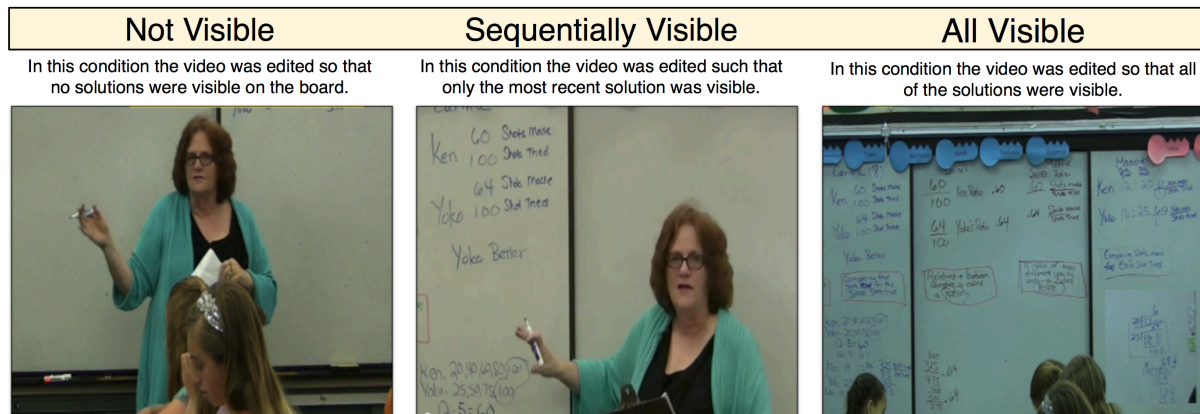


Figure 2. Still images illustrating the experimental conditions created through video editing. From left to right, the first picture shows only the teacher while obstructing the writing on the whiteboard (Not Visible condition), the middle picture shows only the most recent problem (Part Visible condition), and the picture that shows the whole board (All Visible condition).

The Begolli & Richland (2015) experiment stems from studies that have shown that making comparisons is highly taxing on processing demands of working memory and executive function (Richland, Morrison & Holyoak, 2006). In support of unloading working memory to

facilitate comparative thinking, Richland & McDonough (2010) and Richland & Hansen (2013) suggest that scaffolding comparisons during instruction leads to higher learning rates. This study focuses on one of those scaffolds – making solutions visible during comparisons, where Begolli & Richland (2015) examined the following hypothesis:

- Does putting problem solutions on the board enhance students' ability to draw connections when compared with hearing them verbally?
- Does making multiple solution strategies visible on the board enhance students' ability to draw connections when compared to only having the most recent solution visible?

The answers to these questions will provide teachers with a specific practice relevant teaching strategy, whether making student solutions visible supports students' ability to draw connections from instructional comparisons.

Example-study design. The video-instruction was a real classroom lesson and administered *in vivo* with fifth grade students, during regular class time. Students solved word problems invoking rate and ratio concepts, guided by a computerized video lesson, edited in three different ways, and followed a standard experimental procedure (pretest, intervention and immediate posttest, and 1-week delayed posttest). Students, randomized within each classroom, either watched an instructional version video edited so that no solutions were visible on the board – Not Visible (NV), a version where the most recent solution was visible – Sequentially Visible (SV), or a version of the video that showed all solutions on the board throughout the lesson - All Visible (AV). Begolli & Richland (2015) found that students who watched a video where all solutions were visible on the board outperformed both groups in conceptual knowledge. Surprisingly students who saw solutions presented one-at-a-time performed worse

than NV-students, who saw only the teacher and students, but not the board, in procedural knowledge. The authors suggest that SV-students may have confirmed their misconception when it was presented by the teacher and failed to learn from the comparison because the misconception was no longer visible. This may be related to working memory (WM) resources necessary to learn new concepts through instructional comparisons further discussed in Chapter 2. These results provide further questions to research such as the role of individual differences in WM and potential interactions with teacher scaffolds. This study shows how video-lessons can be utilized to answer questions with implications for advancing theory and practice.

Next, I provide insight in the creation of the video-lesson and how each manipulation was created to explore some possibilities of video-editing for future researchers interested in employing this tool to answer empirical questions.

The use of Zooming, Cropping, and Different Camera Angles. The most important aspect of creating a video-lesson stimulus is planning ahead. The lesson script should be well thought out to embedded the hypothesis of interest and cooperation from the teacher is critical to ensure natural delivery of the lesson. The teacher and researcher should coordinate practice trials before recording the lesson. Also, careful consideration needs to be given to acquire recording equipment and to place the equipment appropriately in the classroom. The hypothesis informs lesson design, and camera and microphone placement such that the recorded footage can be utilized to create the manipulations of interest during postproduction.

To examine the hypothesis in Begolli & Richland (2015) and create three versions of the same lesson, the plan in camera placement was: 1) to place one camera that would capture the whole board (board-cam), 2) a camera that would capture the teacher and only the most recent

solution visible (teacher-cam), and 3) a camera that captured the teacher and students but not the board (student-cam). Omnidirectional microphones were planted around the room and the teacher to capture all of student speech and ambient sounds (including distractions such as a pen dropping on a table) and one lapel microphone on the teacher (Figure 2). Each microphone was connected via a splitter jack to a camera audio input or audio recorder to avoid the use of an audio mixer. The objective of the researcher while editing the video footage was to maintain all information contained in the lesson constant across conditions, while manipulating whether the writing on the board will be visible or not. This burdens the video-editor with the challenge of trying to create the manipulations while maintaining a coherent video-lesson. For instance, a simple, but inadequate way of creating a Not Visible condition (in the NV condition students do not see anything written on the board) from the same footage could be to simply show a camera always focused on the students sitting at their desks. But this would create a confound, because it would also test whether seeing or not seeing the teacher influenced student outcomes. Thus, other cameras that capture the teacher were used and the writing on the board was obstructed by zooming in on footage from the teacher-cam, thus, cropping out the writing on the board from the visual canvas. Another method for obscuring the writing on the board was to select a perspective from the student-cam focused on the profile of the teacher, which due to its angle did not capture the board (Figure 2).

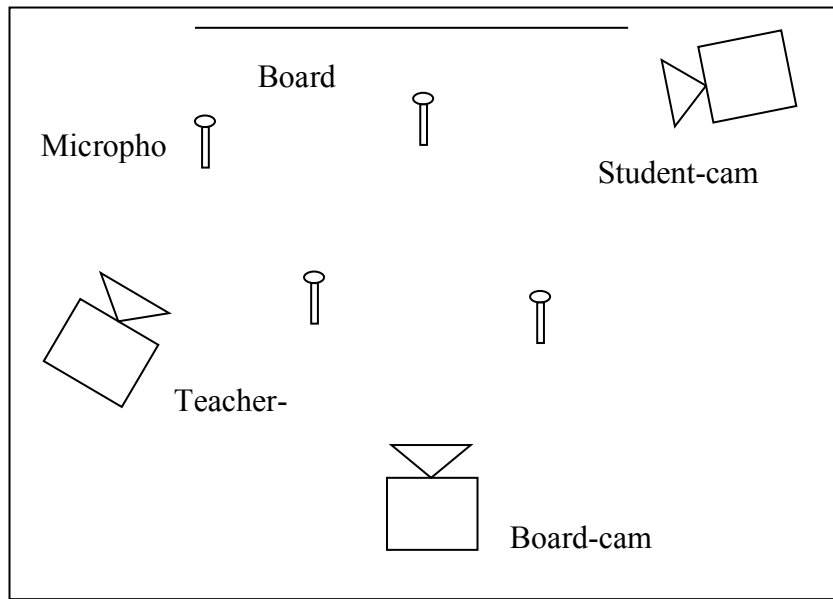


Figure 3. A layout of the cameras and microphones in the classroom used for recording the lesson.

After editing was finished, three versions of the same lesson with equal length were created (see Figure 3). Typical lessons generally last between 40 – 60 minutes, which has been shown to provide difficulties for students to sustain attention whether in real classrooms or in flipped video-lectures (Bunce, Flens, & Neiles, 2010; Risko, Buchanan, Medimorec, & Kingstone, 2013; Wilson & Korn, 2007). While this is not a necessary component of the methodology, to overcome attention difficulties, the lesson-script includes prompts, which require short answers throughout the lesson. Recent work has shown that similar memory tests reduce mind wandering and improve retention (Szpunar, Khan, & Schacter, 2013). In order to make the lesson interactive, each version was strategically divided into 9 independent clips ranging from 1-min to 5-min. The endpoints of each clip were chosen based on when the teacher asked questions to the class. These clips were embedded in a computer program that, at the end of each clip, prompted students with questions that were asked by the teacher in the videotaped

classroom. Students either wrote their answers on paper, or selected multiple choice questions that the computer program collected as assessment data.

Video editing techniques enabled the creation of controlled stimuli that approximate a *live* lesson and learning in a dynamic, realistic context, which can be randomized within two student populations. While video-based stimuli can overcome many limitations of decontextualized stimuli, they suffer from drawbacks of their own.

Limitations of Video-Stimuli. A major limitation lies in the fact that video is an *approximation* of a classroom experience, not a true classroom experience. At the same time, it offers a more realistic medium than text-based or computerized materials. Video can convey emotion, body language and other non-verbal cues, which can be filtered by the student based on their own individual differences. Some insight about the advantages of video over written text comes from research in medical education where ethical and logistic restraints restrict the use of real and simulated patients. Balslev, De Grave, Muijtjens, & Scherpbier, (2005) compared video as an alternative to written text examples and reported that students showed improvements in theory building and theory exploration after watching a video. Despite video not being a real classroom experience, it is a better approximation of the real world compared to text-based stimuli commonly used by experimental psychologists.

Another consideration is that there are limitations in the range of questions that can be answered by systematically video-editing a single lesson or set of lessons. For instance, it may be difficult to edit the same lesson to introduce new or different information into each version. In a possible extension to Begolli & Richland (2015), it would be challenging to create one version of

a lesson in which a teacher compares solutions to a single problem and a second version in which a teacher compares solutions to multiple problems.

A final limitation for this methodology is that impromptu actions and instructional errors may arise during key instructional moments, which make it a challenge when recording a live lesson unfolding. This could obscure the goal of the lesson due to reduced learning for all students and/or confound the research question if the error is related to the manipulation. While errors are natural, this is a challenge to this methodology, because reshooting the lesson with the same teacher and students would be unnatural in the classroom leading to an unnatural video-lesson, and practically, would disrupt the teacher's curriculum. One technique for overcoming this is by doing short reshoots of the problematic parts of the lesson with the same teacher in the same classroom, and inserting them in the video-lesson during postproduction. Despite these limitations, there are vast opportunities for answering questions within these methodological constraints. Many of these questions could stem from surveying current experimental work in classroom settings.

Discussion

Decades of research in cognitive science on learning behavior have not made serious inroads into our educational discussions, and less so into teaching practices (Dunlosky & Rawson, 2012). This may be in part due to the fact that most of this research has been conducted in laboratory settings using stimuli not relevant to classroom instruction (Richland, Linn, Bjork, 2009). Recently, there has been a surge of classroom experiments of learning processes drawing from laboratory work of cognitive and educational psychologists which have suggested methods for improving student learning (e.g., see Roediger and Pyc, 2012; Richland, Stigler, and

Holyoak, 2012; Booth, Lange, Koedinger, & Newton, 2013; Carpenter et al., 2012; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Fyfe, Decaro, & Rittle-Johnson, 2014; Glogger-Frey, Fleischer, Grüny, Kappich, & Renkl, 2015; Klahr & Li, 2005; Metcalfe, Kornell, & Son, 2007; Pashler et al., 2007; Schwartz, Chase, Oppezzo, & Chin, 2011)..

This body of work has been able to make precise claims about various cognitive strategies, succinctly summarized by Koedinger, Booth, & Klahr (2013). The overarching commitment to maintain experimental control, however, has driven many studies bridging laboratory and classroom settings to utilize decontextualized stimuli, where the manipulation is embedded in paper and computer-based materials. This may contribute to a lack of specificity when making recommendations about teaching practices (e.g., see Daniel, 2012; Kornell, Rabelo, & Klein, 2012; Mayer, 2012). Designing manipulations that maintain experimental control but involve teacher-led instruction is often difficult to achieve (Brown, 1992). Recommendations stemming from laboratory work in cognitive and educational psychology (e.g. immediate versus delayed feedback, use of concrete versus abstract materials), often are in contradiction, and their effectiveness varies across contents and contexts (Koedinger, Booth, Klahr, 2013). This issue is further aggravated for teachers looking to adopt these recommendations, because most experiments derive abstractions through manipulations and/or stimuli that rarely mimic an actual teacher's behavior (for exceptions see Richland and McDonough, 2010; Cook, Duffy, and Fenn, 2013). These factors may contribute to the reasons why the scientific community and disciplinary panels have little evidence that directly translates into teaching practices, thus providing teachers with broad recommendations that lack specificity (e.g., see Daniel, 2012; Kornell, Rabelo, Klein, 2012; Mayer, 2012). While promising work in emerging fields is attempting to tame the complexity of instructional recommendations (Society

for Research on Educational Effectiveness; International Education Data Mining Society), the proposed methodology focuses on the latter issue, that of discovering findings that better translate to *teaching practices* by using video-based stimuli.

The video methodology and video editing techniques stemming from gesture research and the studies presented in this paper provide a model that can be adopted by many researchers interested in making causal claims regarding teaching practices. This model also provides opportunities for design-based studies to create a bi-directional exchange between the classroom and the laboratory.

Design-experiments provide rich descriptions of the classrooms and have provided valuable insights on the mechanisms of a classroom that lead to successful teaching. A major concern for design-based researchers is that most stimuli used by laboratory researchers de-emphasize the role of the teacher, as well as the students, and interactions between them (Collins, Joseph, & Bielaczyc, 2009; Collins, 1999). However, video-based stimuli of classrooms can overcome this, and the design-based research community has been encouraging scientists to increase their efforts to examine questions that arise in classrooms using controlled experiments (Design-Based Research Collective, 2003; McCandliss, Kalchman, and Bryant, 2013).

In sum, recommendations from design-research largely lack specificity due to multiple factors being studied simultaneously. On the other hand, paper and computer-based materials used by experimental psychologists often result in recommendations that are difficult to translate into teaching practice. Teachers are left with the challenge of how to incorporate these principles in their teaching routines or adopt them for the complex cognitive tasks they may expect their students to engage in. Video is presented as a tool to help researchers with the specific techniques that can partly account for students' reasoning challenges arising from interactions in

classroom settings. Thus, next I exemplify how phenomena that may be familiar to our readers stemming from either controlled studies bridging the laboratory and classroom or qualitative designs and teacher observations could be transformed into experiments that test teaching practices on student learning as outlined in Figure 1.

Testing Teaching Practices based on Cognitive and Education Theories. There are many ways that the video-lesson methodology could be used. To illustrate the breadth of the potential within this methodology, one might consider a theory-based question such as: how do invention activities (DeCaro & Rittle-Johnson, 2012; Schwartz et al., 2011) guided by a teacher affect student outcomes when presented before versus after direct instruction? This question could be answered by recording a lesson that begins with (a) invention activities that are followed by (b) direct instruction and practice. In video post-production, researchers could create two versions of the lesson in which the order of these two activities (a) and (b) is counterbalanced. The video-lesson could be split into two (or more) clips, one clip contains (a) the invention activity and the other clip (b) direct instruction with practice. This would enable the researcher to create two versions of the lesson which maintain constancy, apart from manipulating the sequence of instructional activities (invention vs. direct instruction). To make the link seamless, I recommend to include a question and answer period in between the two clips. This would provide a more rigorous test of the order effects than having teachers teach different students in the two counterbalanced versions, or different teachers teach the two versions. Each of those has the potential for confounding the order effects with contextual features of the teacher-student interactions.

Another example derives from (Stein, Engle, Smith, & Hughes, 2008) recommendation for how to sequence multiple solutions to a single problem. While Stein et al. (2008) propose that teachers should sequence solutions from simpler to more complex, they admit that more work needs to be done to compare the effects of different sequencing methods on student learning. For example, prior work stemming from observations of Japanese lessons from the TIMSS 1999 study (Stigler & Hiebert, 1999) show that teachers in higher achieving countries present students a solution that represents a common misconception before moving on to the correct solution (Shimizu, 2003). But whether common misconceptions should be presented first or last, remains to be tested empirically. A video-based lesson could be split into independent clips (similarly as in the example above) and the order of problem solutions could be manipulated through video-editing. One version orders the video clips so the common misconception is presented first in the sequence of multiple problem solutions, whereas in another version the common misconception is presented last.

Video-lessons could also include minor variations of key moments within a lesson, but maintain the rest of the lesson unchanged. The literature is vast with recommendations that stem from observational data or from educational psychology experiments that positively affect student learning, which are not tested using rigorously controlled teacher-guided lessons (for review see Pashler et al., 2007; Mayer, 2008; Roediger and Pyc, 2012; Richland, Linn, and Bjork, 2007). Richland, Zur, and Holyoak (2007) have identified scaffolds correlating with students' success in relational thinking of mathematics concepts that could be tested using this method. For example, utilizing spatial cues to highlight ideas between two representations (e.g. drawing a scale underneath an equal sign of an equation), mental imagery (e.g. "picture a scale when balancing an equation"), gestures, and spatial alignment of mathematical representations

(i.e. organization of instructional problems on the board). There is also significant literature that examines the effects of external factors on performance, which could be applied to examine learning. For instance, language complexity of mathematics tests (Abedi, 2008) and stereotype threat (Beilock, Rydell, & McConnell, 2007) have been shown to affect student performance, but are not tested within the context of learning in a classroom.

Summary

Thus in summary, video-editing provides a rich opportunity for bridging cognitive and educational research through experimental designs. Unlike other methodologies that test learning principles but may leave teachers with the challenge of knowing how or which recommendations to incorporate into a real lesson, video can be used as a tool to give specificity to teaching strategies. I presented how videotaped-lessons can be scripted to incorporate learning principles, modified to create manipulations that test hypothesis in learning behavior, used as stimulus for a randomized experiment. In particular, I describe tests of whether making problem solutions visible embedded in a videotaped-teacher guided lesson affected student learning, which showed how zooming and different camera angles can be used to examine whether students' cognitive load is reduced by writing problem solutions on the board. These stimuli that approximate a real lesson maximize external validity and can be used to isolate cause and effect on students' outcomes, depending on the intervention of interest. Many teachers use effective instructional techniques embedded in their lessons and this tool provides us with ways that can test which of these techniques are most efficient when the rest of the lesson is held constant. The examples provided in the discussion of the manuscript were aimed to illustrate the opportunities

of transforming and developing research in the context of testing educational strategies that have particular relevance for teachers looking to support student learning.

We hope that future researchers will begin considering how to embed learning principles in real lessons, and in particular, will adopt video and video editing tools to create manipulations that test the efficiency of learning principles in a contextualized way that is close to everyday classroom practice. In this way, ideally they will make gains in providing teachers with successful, specific, and relevant instructional techniques.

Chapter 2: Teaching Mathematics by Comparison: Analog Visibility as A Double-Edged Sword (Begolli & Richland, 2015)

Comparing different student solutions to a single instructional problem is a key recommended pedagogical tool in mathematics leading to deep, generalizable learning (see (Common Core State Standards Initiative, 2010; National Mathematics Advisory Panel, 2008; National Research Council, 2001); however, the cognitive underpinnings of successfully completing this task are complex. In order to understand that $2 + 2 + 2$ conveys the same relationships as 2×3 , for example, students must perform what has been theoretically described as structure-mapping: represent the multiple solutions as systems of mathematical relationships, align and map these systems to each other, and draw inferences based on the alignments (and misalignments) for successful schema formation (see Gentner, 1983; Gick & Holyoak, 1983). Structure-mapping is posited to underlie the processes of analogical reasoning where one *source* representation (e.g. $3 - 1 = 2$), is mapped to a *target* representation (e.g. $x - 1 = 2$), (Gentner, 1983).

Orchestrating classroom lessons in which learners successfully accomplish such structure mapping is not straightforward for many reasons. First, classroom discussions often involve comparisons between a misconception and a valid solution strategy, which may be particularly effortful in regards to structure mapping and schema formation, because misconceptions often derive from deeply or long held beliefs that may be difficult to overcome (Chi, 2013; Chinn & Brewer, 1993; Vosniadou, 2013). Secondly, reasoners often fail to notice the relevance or importance of doing structure mapping unless given very clear and explicit support cues to do so (Alfieri, Nokes-Malach, & Schunn, 2013; Gentner, Loewenstein, & Thompson, 2003; Gick & Holyoak, 1980, 1983; Schwartz & Bransford, 1998). Third, reasoners may intend to perform

structure-mapping but the process breaks down because their working memory or cognitive control processing resources are overwhelmed: (Cho, Holyoak, & Cannon, 2007; English & Halford, 1995; Paas, Renkl, & Sweller, 2003; Richland et al., 2006; Waltz et al., 2000). Working memory is required to represent the relationships operating within systems of objects as well as the higher order relationships between a familiar representation (*source analog*) and less familiar representation (*target analog*). In this case, to mentally consider the relationships between two solution strategies, one must hold in mind the steps to each solution strategy being compared, must re-organize and re-represent these systems of relations so that their structures can align and map together, identify meaningful similarities and differences, and derive conceptual/schematic inferences from this structure-mapping exercise to better inform future problem solving (Morrison et al., 2004). Lastly, reasoners' prior knowledge plays an additional role. Those without adequate knowledge of the key relationships within the source and target representations are either unlikely to be able to notice structure mapping (Fyfe, Rittle-Johnson, & DeCaro, 2012; Gentner & Rattermann, 1991; Goswami, 2001), or this process will impose higher processing load than it would for those with more domain expertise (Novick & Holyoak, 1991).

These challenges mean that the instructional supports are very likely essential to whether students notice and successfully execute structure-mapping between multiple solution strategies. The current study tests a classroom-relevant mode for providing such support - providing visual representations of the source and target analogs. The study manipulation assesses whether 1) making source and target analogs visual (versus oral) increases the likelihood that participants will notice and successfully benefit from structure mapping opportunities, and 2) whether learning is enhanced if the visual representations of all compared solutions are visible simultaneously during structure-mapping. The former should increase the salience of the

relational structure of each representation, while the latter should reduce the working memory load and cognitive control resources necessary for participants to engage in structure-mapping and inference processes.

Understanding the relationships between visual representations and learners' structure-mapping provides insights into both a key pedagogical practice and improving theory on structure-mapping and analogy more broadly. Teachers tend to find it difficult to lead students into making connections between problem solutions, and one productive way to support them is to provide guidelines for such discussions (e.g., see Stein, Engle, Smith, & Hughes, 2008).

The study methodology is designed to lead to generalizable guidelines for the use of visual representations during classroom discussions comparing multiple solutions to a single problem. Observational data suggest that U.S. teachers do not regularly provide visual representations to support multiple compared solution strategies, and when they do, they are less likely than teachers in higher achieving countries to leave the multiple representations visible simultaneously (Richland, Zur & Holyoak, 2007). The literature on the role of making representations visible suggests that presenting source and target analogs simultaneously versus sequentially leads to better learning (Gentner, Loewenstein, & Thompson, 2003; Rittle-Johnson and Star, 2009), but these studies did not examine comparisons between an incorrect and a correct strategy. On the other hand, learning from incorrect and correct strategies was better than learning from correct strategies only (Durkin and Rittle-Johnson, 2012; Booth et al., 2013), but these studies have not investigated the role of visual supports. Thus this study may provide first evidence toward a guideline for teaching instructional comparisons with visual representations, particularly in the context of comparing a misconception and a correct student solution.

To maximize the relevance of the findings for teaching practices, I test alternative uses of visual representations within a mathematics lesson on proportional reasoning – a topic central to curriculum standards. Stimuli and data collection are conducted in everyday classrooms. The proportional reasoning lesson is situated in the context of a problem asking students to find the best free-throw shooter in a basketball game. In this lesson, students are guided to perform structure-mapping between three commonly used solution strategies: a) subtract between two units (e.g. subtract shots made from shots tried, which is incorrect and a common misconception), b) find the least common multiple between two ratios (e.g. proportionally equalize shots made to compare the shots tried), and c) divide two units to find a success rate (e.g. divide shots made by shots tried).

In addition, the work provides insights into theory on structure-mapping and analogy. I examine a specific case of schema formation from structure-mapping: identifying *misalignments* between two representations, in this case “subtraction” (a common misconception) and “proportions” (e.g. rate or ratio). To benefit from this structure-mapping exercise, students have to identify elements that are not aligned between the two relational structures. Namely, the difference between comparing a single unit (e.g. shots missed) and a relationship between two units (e.g. shots made and shots tried). Schema formation about proportional reasoning would derive from understanding the higher order differences between these two ways of attempting to solve the proportion problem. In contrast, structure-mapping failures may lead to the adoption of an inappropriate source (single unit comparison), or at best the target (relational comparison), but neither of which would be schema formation. In fact, either of these could hinder structure-mapping, lead to misconceptions, and/or reduce transfer when solving later problems. I expect these findings to provide a more nuanced view on the possible implications of visual

representations in terms of supporting or straining WM resources necessary for successful structure-mapping and its influence on students' mathematical knowledge.

We examine these research questions using an experiment that employs methods and measurements designed with the aim to optimize both ecological validity and experimental rigor.

I utilize stimuli that approximate a true classroom experience – a single mathematics video-lesson recorded in a real classroom – then randomly assign students within each classroom to watch one of three versions of the lesson (see Figure 2). The recording is video-edited to support or strain WM resources through variations in the visibility of representations. I use four carefully designed pre-post-and delayed posttest measures to assess the impact of these manipulations on: 1) procedural understanding - students' ability to reproduce taught procedures; 2) procedural flexibility - participants' ability to understand multiple solutions and to deploy the optimal strategy; 3) conceptual understanding – understanding the concepts underpinning rate and ratio; and 4) use of misconceptions. These measures enable us to not only assess which use of visual representations is most effective for promoting learning, but they also let us better understand the processes by which children have been learning in each of the three conditions. Memory and retention of the instruction would be reflected in procedural understanding measures, while schema formation would be better reflected in the procedural flexibility and conceptual understanding measures. I theorize that working memory is the mechanism underpinning differences between these conditions on learning, since more working memory is required to hold visual representations in mind when reasoning about information that is not currently visible.

Thus, findings from this experiment will yield both theoretical insight into the role of visual representations for complex structure mapping, retention, and schema formation, and provides practice relevant implications for everyday mathematics teachers.

Experiment 1

Method

Participants

Eighty-eight participants were drawn from a suburban public school with a diverse population (33% Asian, 25% Hispanic or Latino, 25% White, 6% Filipino, 5% Black or African-American, 6% Other race/ethnicity; 38% disadvantaged SES; 37% English Language Learners). Data from students who missed the intervention were omitted since their scores were not affected by the manipulation. Students who missed the pretest were also excluded from the analyses. They were excluded rather than having their data imputed (Peugh & Enders, 2004) due to concerns that solving pretest problems may have changed the learning context for those who took it due to a testing effect (Bjork, 1988; Carrier & Pashler, 1992; McDaniel, Roediger, & McDermott, 2007; Richland, Kornell, & Kao, 2009; Roediger & Karpicke, 2006a, 2006b; for a review see Richland, Bjork, Linn, 2007). The final analyses included 76 students (32 girls) who completed all three tests (pretest, immediate posttest, and retention test) with ages ranging between 11-12 years old.

Materials

Materials for the intervention consisted of a worksheet, a netbook, and a pre-recorded video-lesson embedded in an interactive computer program. Figure 1 provides a visual of the process for developing the lesson and administering it as stimulus to students in different schools.

Interactive Instructional Lesson

Proportional Reasoning. There is a large literature researching student thinking about ratio that has contributed to evidence, which can predict students' responses to proportion problems. (e.g., Hart, 1984; Hunting, 1983; Kaput & West, 1994; Karplus, Pulos, & Stage, 1983; Lamon, 1993a, 1993b; Lo & Watanabe, 1997; Shimizu, 2003). The National Research Council (2001) identified proportional reasoning as requiring refined knowledge of mathematics and as the pinnacle of elementary arithmetic critical for algebraic and more sophisticated mathematics. Ratio was chosen for this study for two reasons: (a) it is part of the common core standards for 6th grade because it is essential for subsequent learning of algebra and (b) previous research has shown that ratio problems are cognitively taxing, leading to more diverse systematic student responses, useful for understanding mathematical thinking.

Lesson content and teacher collaboration. An approximately 40-minute lesson was developed by the authors in collaboration with a nationally board certified public school teacher (see Appendix A for a sample of the transcript from the lesson). First, a lesson script was written based on a previously published lesson model (Shimizu, 2003) on which the teacher and the authors performed practice trials without students. During practice, the transcript was modified to feel more natural with the teacher's instructional style. The script provided specific details on how to present the main instructional problem; identify key student responses, present them on the board in a predetermined sequence, organize student responses on the board; the type of gestures to use, and so on. The teacher then taught the scripted lesson to her students in her regular classroom. Students were not given instructions and were expected to act spontaneously as they normally would during class hour. This teacher and students were not participants in this

study. They only partook during the recording of the lesson, which was used as stimulus for other students.

The lessons began with the teacher asking her students to solve the problem below (Table 2).

Table 2

Ken and Yoko were shooting free-throws in a basketball game. The results of their shooting are shown in the table below. Who is the better free-throw shooter?

Problem solved and discussed during videotaped lesson.

	Shots-Made	Shots-Tried
Ken	12	20
Yoko	16	25

This was a novel problem and students were not given hints or instruction on how to solve it. The teacher's only instruction was to "solve any way you know how," and that "the class can learn from all the answers." If students objected because they did not know how to solve this, the teacher encouraged them to use any strategy they liked.

During the time when the students solved the problem, the teacher circled the room to identifying three students that used the three strategies listed in Table 3.

After a 5-min period, those three students were called to the board to share their solutions with the class. The sequence of strategies was presented based on this published lesson model (Shimizu, 2003). First, a student verbalized the incorrect solution C while the teacher wrote it on the board. Next, solutions B and then A were presented in the same manner with different students describing their strategies, followed by a short discussion on what the student was thinking when using the strategy. After all solutions were presented the teacher orchestrated a

discussion comparing the different solution methods to achieve specific goals. The teacher's goals were: (a) to challenge students' common misconception (strategy C – subtraction) by asking students whether strategy C is reasonable if the numbers changed (Ken makes 0 out of 4 free throws and Yoko makes 5 out of 10), (b) introduce the concept of proportional reasoning (strategy B) by leading students to notice that proportions can be compared by making one number (i.e. the denominator) constant in each ratio and then comparing the other number (i.e. the numerator) to determine which ratio is larger, (c) to challenge students to notice that the least common multiple strategy can become more difficult for larger and prime numbers, (d) to notice that using division (solution A) is the most efficient strategy, since it does not change much in difficulty, regardless if the numbers increase. These points led the teacher to introduce the concept of ratio through pre-designed comparisons while the class responded spontaneously to her prompts.

Table 3
Three student solutions compared during the videotaped instructional analogy lesson.

A	Student finds the number of goals made if each player shoots only 1 free throw. Ken: $12 \text{ goals} \div 20 \text{ shots} = 0.6$, and Yoko: $16 \text{ goals} \div 25 \text{ shots} = 0.64$. <i>Answer: Yoko, because she gets more goals for the same number of free throws (.64 > .60).</i>	Most efficient generalizable strategy
B	Student compares the number of goals if each player shoots the same number of free throws. Using 100 as the last common multiple, we get Ken: $60/100$ and Yoko: $64/100$. <i>Answer: Yoko, because she would get more goals if they each shot 100 times ($64/100 > 60/100$).</i>	Finding Least Common Multiple: Drawback, difficult when larger numbers
C	Student compares the players by finding the difference between the number of free throw shots and the number of goals. Ken: $20 \text{ shots} - 12 \text{ goals} = 8 \text{ misses}$, and Yoko: $25 \text{ shots} - 16 \text{ goals} = 9 \text{ misses}$. <i>Answer: Ken, because he missed fewer times than Yoko ($8 < 9$).</i>	Misconception (incorrect): subtract values and compare differences without considering the ratio.

The crux of this manipulation came from applying video-editing techniques to the recording to create three different versions of the same lesson. FINAL CUT PRO's (FCP) 7.0.3 academic version's various editing features were used such as zooming, cropping, or different camera perspectives of the screen canvas to either: (a) hide the board to create a version of the lesson for the Not Visible condition; (b) show only the chapter of the board most recently discussed, but hide other areas of the board to create a version of the lesson for the Sequentially Visible condition; or (c) show the whole board throughout the version of the lesson in the All Visible condition. Thus, the same content was verbalized in all three lesson versions, but with systematic differences in visual cues.

Each version of the lesson was strategically divided into 9 clips with an approximate range from 1-min to 8-min. The endpoints of each clip were chosen based on when the teacher asked questions to the class. Each version of the video-lesson was made interactive by embedding clips of the video in a computer program written specifically for this study. At the end of each clip, the program prompted students with questions that were asked by the teacher in the videotaped classroom. Students in all conditions either wrote their answers on a packet provided by the experimenters, or selected multiple choice questions that the computer program collected as assessment data. This methodological approach of stimuli creation provided a rigorous level of experimental control of a highly dynamic context – an everyday classroom. Further, it allowed for randomization within each classroom.

Mathematics Assessment

The assessment was designed to assess schema formation and generalization. Mathematically, the assessment included four constructs, procedural knowledge, procedural

flexibility, and conceptual knowledge (Rittle-Johnson & Schneider, 2013). The first three constructs were conceptually derived from Rittle-Johnson & Star (2007, 2009), and adapted to the core concepts and procedures underlying ratio problems. Items used for assessing each knowledge type are included in Appendix B. The items on the pretest and posttests were identical, but the pretest contained 5 additional procedural knowledge problems used to assess students prerequisite knowledge of basic procedures (e.g. division by decimals, finding the least common denominator). Detailed scoring on all of the items can be found in Appendix B. Scores for each construct were averaged to yield an overall mean for that particular construct. Student produced strategies were coded for use of: (1) division, (2) least common multiple, (3) subtraction, (4) ratio, (5) other valid strategy (e.g. cross multiplication), (6) other invalid strategy (e.g. addition). For the analyses, these codes were collapsed for each problem into correct or incorrect. Open-ended questions (e.g. what is the definition of ratio?) were given a binary score of either correct or incorrect. Interrater agreement for 3 raters was calculated on 20% of each test phase on whether the strategy was correct or incorrect ranging from 95-99% (0.88 - 0.97 Kappa). Strategy specific coding was less reliable ranging from 86% - 92% (0.79- 0.85 Kappa).

Procedural Knowledge. Seven problems measured baseline ability and growth in procedural understanding. The procedural knowledge construct was designed to test: (a) students' knowledge for producing and evaluating solutions of: familiar problems (i.e. similar in appearance to the problem used in the video lesson; n=2 produce; n=3 recognition), (b) transfer problems (i.e. students' competence to extend these solution strategies to problems with novel appearance, but similar context; n=2). Students received 1 point if they produced the correct solution method and another point if they produced a correct solution method and made an inference to reach a correct answer. Students received 1-point if they recognized the correct

strategy on multiple-choice questions. One procedural transfer problem showed no sensitivity to the intervention, so it was dropped from the analyses (average change of 5% from pretest to posttest; see Problem 4 in Appendix B for details). Cronbach's alpha on the remaining items was .88 at posttest, .91 at delayed posttest, and .86 at pretest, which are above the suggested values of .5 or .6 (Nunnally, 1967).

Procedural Flexibility. The procedural flexibility construct measured: (a) students' adaptive production of solution methods (n=3), (b) their ability to identify the most efficient strategy (n=1), and (c) students' ability to identify a novel solution method which was related to a taught strategy (n=1). For (a) students were presented with 1 problem containing 3 items. The first item asked students to produce two strategies (and correct answers) for the same problem. The second item asked students to evaluate which of the two strategies was most effective. The third item asked students to select 1 out of 4 reasons for their choice on item 2. Students could receive 2 points for the first item and 1 point on the last two items. For (b) students were presented a multiple choice problem, which required them to identify the optimal strategy from 2 valid, and 2 invalid strategies. For (c) students were presented with a multiple choice problem that probed students' competence to identify the correctness of a related but novel method of solving a problem (i.e. finding the lowest common multiple for the numerator instead of the denominator). Both (b) and (c) were scored for accuracy. Cronbach's alpha on the flexibility construct was .68 at posttest, .67 at delayed posttest, and .57 at pretest.

Conceptual Knowledge. The conceptual knowledge construct consisted of 7 items that were designed to probe students' explicit and implicit knowledge of ratio. Students' explicit knowledge was measured by asking them to write a definition for ratio, which was scored for accuracy. The other six items measured students' implicit understanding and transfer to new

contexts. One problem probed whether students could conceptually examine two sets of non-numerical quantities (i.e. pictures of lemon juice and water), adapt their just learned solution methods to this novel context, and compare the sets to decide which ratio was greater (i.e. which lemonade was more lemony?). On this problem, students were scored on whether they could produce the correct setup given objective quantities and choose the correct set (when using the correct setup). The remaining problems were scored for accuracy. One multiple-choice item that was part of a procedural knowledge problem probed whether students could conceptually evaluate the multiplicative properties of a solution procedure they had produced (see Problem 1 in Appendix B under the Procedural and Conceptual Knowledge chapters). Three conceptual questions, 2 fill-in-the-blank and 1 multiple choice, probed students' understanding of units in correspondence to ratio and rate numerical quantities. One of these unit questions was dropped due to floor effects (only 6 out of 97 students responded correctly; see Problem 5 part 3 in Appendix B for details). Lastly, one problem asked students to recognize the correct solution and setup for a novel problem type and context. However, this problem suffered from ceiling effects and was not sensitive to intervention (averages ranged between 81%-89%; average change -3% from pretest to posttest; see Problem 8 in Appendix B for details). Thus, it was omitted from the analyses. Cronbach's alpha on the remaining items was .66 at posttest, .64 at delayed posttest, and .42 on pretest. Reliability at pretest was lower due to floor effects.

Common Misconception. Misconceptions are mistakes that students make based on inferences from prior knowledge, which obstruct learning (Smith III, diSessa*, & Roschelle, 1994). Based on a published lesson (Shimizu, 2003), pilot data, and pretest data, a solution involving subtraction was expected to be the most common misconception participants would bring to the study. A sub-coding assessed how frequently students used the subtraction method.

The common misconception measure examined students' use of subtraction on near transfer procedural problems that looked like the instructed problem in the video lesson.

Design & Procedure

Students within four classrooms, not in the videotaped classroom, were randomly assigned to three experimental conditions: Not Visible ($n = 26$), Sequentially Visible ($n = 26$), or All Visible ($n = 24$). All students were administered a pretest, one week later completed the video-lesson intervention and an immediate posttest, and one week later completed a delayed posttest. Students underwent the intervention before being introduced to rate and ratio in their regular curriculum.

Results

Baseline Data

One-way ANOVAs were conducted first to establish that the randomization was successful and there were no differences between conditions on each of the above described constructs. At pretest, there were no differences between conditions on any of the outcome constructs: procedural, procedural flexibility, conceptual knowledge, and common misconception with all p -values above .05, $F_s(2, 80) = 0.69, 0.53, 0.71, \text{ and } 0.29$, respectively. At pretest, students used mostly invalid strategies when solving ratio problems, and left a significant portion of the problems blank (Table 4). The average scores at each test point by condition are summarized in Table 5.

Table 4
Solution Strategies Produced for Ratio Problems by Condition

	Blank	Division	Least Common Multiple	Ratio Setup	Subtraction	Other Valid	Other Invalid
Pretest							
All Visible	25%	4%	18%	9%	20%	7%	18%
Seq. Visible	24%	8%	7%	4%	22%	4%	30%
Not Visible	17%	10%	10%	8%	22%	2%	30%
Immediate							
All Visible	14%	48%	22%	1%	9%	3%	4%
Seq. Visible	12%	19%	15%	1%	35%	1%	17%
Not Visible	5%	39%	21%	1%	19%	1%	14%
Retention							
All Visible	16%	39%	22%	1%	13%	3%	6%
Seq. Visible	15%	25%	12%	0%	31%	5%	11%
Not Visible	12%	30%	19%	4%	21%	4%	12%

Condition Effects

Analysis plan overview. I next sought to examine the effects of condition on each of the dependent variable constructs measured. There were three primary constructs: procedural knowledge, procedural flexibility, and conceptual understanding, and one additional measure to gather deeper information on the impact of the manipulations on inappropriate retention – use of the misconception.

I conducted separate ANCOVAs for each outcome measure with both posttests as a within-subjects factor (immediate test and delayed test) and condition as a between-subjects factor. Students' pretest accuracy and their classroom (i.e. teacher) were included as covariates. In the model, the pretest measure matched the posttest measure, such that procedural knowledge pretest served as a covariate for the procedural knowledge posttests, and so on. Levene's test of variance homogeneity was used to ensure that all measures were appropriate for use of the ANCOVA statistic. These analyses yielded no significant differences in variance between groups on all measures (F -values range $.078 < F < 1.76$ and p -value range $.17 < p < .92$), apart

from the score for how often students used the misconception. The measure of misconception use was therefore analyzed using Mann-Whitney U comparisons of pretest to posttest gain scores across conditions, with a Bonferroni correction for multiple comparisons.

Table 5
Student Scores, by Condition

	Pretest		Posttest		Delayed	
	Mean	SD	Mean	SD	Mean	SD
<i>Procedural</i>						
All Visible	33%	0.29	59%	0.37	67%	0.35
Sequentially Visible	25%	0.28	38%	0.35	38%	0.35
Not Visible	27%	0.32	52%	0.34	56%	0.35
<i>Flexibility</i>						
All Visible	11%	0.14	37%	0.29	42%	0.26
Sequentially Visible	14%	0.22	17%	0.24	22%	0.29
Not Visible	10%	0.14	23%	0.25	30%	0.31
<i>Conceptual</i>						
All Visible	19%	0.19	44%	0.29	47%	0.26
Sequentially Visible	13%	0.17	30%	0.30	27%	0.28
Not Visible	17%	0.22	31%	0.25	35%	0.23
<i>Common Misconception</i>						
All Visible	20%	0.21	13%	0.20	10%	0.15
Sequentially Visible	21%	0.21	29%	0.31	32%	0.27
Not Visible	22%	0.22	21%	0.24	19%	0.24

When a main effect of condition was present on an ANCOVA analysis, least significant difference tests were used to determine whether there were differential effects of condition on posttest performance. Student performance was not expected to change between posttests because students continued to learn about ratio related concepts after the intervention and the within-subject test for time confirmed this prediction.

Main Effects of Condition. The results of each ANCOVA are summarized in Table 6. For each outcome there was a main effect of condition with moderate to high effect sizes ($.11 < \eta^2 < .15$) and sufficient power ($.77 < (1-\beta) < .90$). Pretest was a significant predictor for each construct, though misconception use at pretest was not independently predictive. There were no

expectations that time of test or classroom teacher would interact with condition and these tests support this. Pairwise comparisons between conditions on each construct are reported below (see Table 7 and Figure 4).

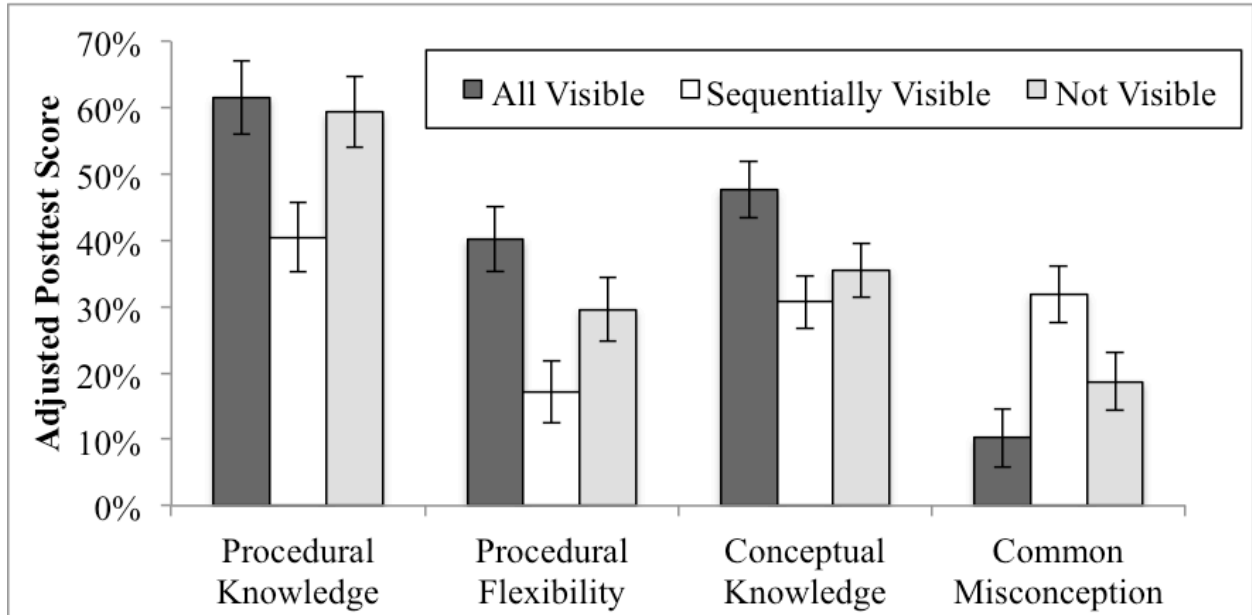


Figure 4. Estimated marginal means across both posttests on procedural knowledge, procedural flexibility, conceptual knowledge, and use of the misconception by condition for Experiment 1. Error bars are standard errors.

Table 6
Analyses of Covariance Results on Learning Outcomes

Factor	Procedural			Flexibility			Conceptual			Used Misconcept.		
	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2	<i>F</i>	<i>p</i>	η^2
Condition	4.70	.012	.13	5.62	.006	.15	4.64	.013	.12	7.39	.001	.18
Pretest	32.19	.012	.33	4.49	.038	.06	29.69	.013	.31	1.74	.192	.03
Teacher	2.53	.117	.04	0.69	.410	.01	1.07	.304	.02	0.06	.816	.00

*Condition degrees of freedom are (2, 66); all others are (1, 66).

Procedural knowledge. Students in the All Visible condition outperformed students in the Sequentially Visible condition in procedural knowledge. An unexpected finding was that

students in the Not Visible condition also outperformed students in Sequentially Visible condition (see Table 7). Not seeing the board did not affect students procedural knowledge compared to students who saw all solutions on the board simultaneously.

Procedural Flexibility. Students in the All Visible condition outperformed students in the Sequentially Visible condition, but there were no differences between any other groups (Table 7).

Conceptual Knowledge. Pairwise comparisons reveal that students in the All Visible condition scored significantly higher than students in the Not Visible condition and students in the Sequentially Visible condition (Table 7). There were no differences between students in the Sequentially Visible and Not Visible condition on conceptual knowledge (see Table 7).

Table 7
Pairwise Comparisons, by Condition with Pretest & Teacher as Covariates for both Posttests

Knowledge Type	AV vs. SV	AV vs. NV	SV vs. NV
Procedural Knowledge	.007	.778	.013
Procedural Flexibility	.001	.129	.071
Conceptual Knowledge	.005	.041	.407

These results could have been driven by the types of solution strategies that students used (Rittle-Johnson and Star, 2007; 2009), particularly subtraction (a common misconception).

Table 8
Summary of Mann-Whitney U Pairwise Comparisons of the Common Misconception

	Immediate Posttest		Delayed Posttest	
	<i>Mann Whitney U</i>	<i>p-value</i>	<i>Mann Whitney U</i>	<i>p-value</i>
AV vs. SV	208	0.034	178.5	0.007
AV vs. NV	277.5	0.49	270.5	0.41
SV vs. NV	263	0.159	266	0.179

Bonferroni adjustment renders alpha levels at $p = .008$ (.05/6).

Common Misconception. The general pattern for students' use of the common misconception displays a reverse pattern compared to the procedural knowledge performance,

providing insight into why the Sequentially Visible condition led to low accuracy rates. Mann Whitney U pairwise comparisons with a Bonferroni adjusted alpha level of $p = .008 (.05/6)$ show that students in the Sequentially Visible condition used the common misconception significantly more from pretest to delayed test compared to students in the All Visible condition, but no other comparisons were significant (Table 8).

Discussion

Overall, this study supported the hypothesis that the presence of visual representations during a discussion comparing multiple solutions to a problem can serve as a double-edged sword. The presence and timing of visual representations impacted children's learning from a mathematical classroom lesson on ratio when comparing a misconception to two correct strategies in both positive and negative ways. Having all visual representations available simultaneously led to the highest rates of learning, while having them presented sequentially led to the highest rates of misconceptions.

Specifically, the ability to see all compared representations simultaneously throughout the discussion increased the likelihood of schema formation and optimized learning when compared with seeing compared representations only sequentially. This was evidenced by greater ability to: 1) use taught procedures, 2) understand multiple accurate solution strategies and select the most efficient strategy, 3) explain and use the concepts underlying taught mathematics, and 4) minimize use of a misconception.

Strikingly, presenting mathematical solutions sequentially led to the lowest performance on these positive measures of learning, and the highest rates of misconceptions at posttests. This condition led to even lower learning rates overall than having no visual representations present

during any of the comparison episodes, though these differences were not present on all measures. The details of how these conditions differed are informative to building theory regarding the role of visual representations in comparisons and schema formation.

Having the solution strategies presented only verbally (Not Visible condition) led to performance rates that fell in between the two visual representation conditions. NV presentation did lead to some retention of taught procedures and some schema formation, but not as universally as in the All Visible condition. At the same time, these participants (Not Visible condition) were less likely than in the Sequential condition to produce the misconception, suggesting that they did not retain the instructed representations as well or uncritically as in that condition. It may be that the Not Visible condition was most effortful for students and thus some students were less successful than in the All Visible condition, but for those students who were able to perform that effort, their learning was strong.

Drawing on theory on the cognitive underpinnings of structure-mapping, I interpret the differences between these conditions based on their likely load on students' executive function resources. Structure-mapping is well established to require both the ability to hold representations in mind and manipulate the relationships to identify and map structural alignments or misalignments (e.g. Waltz et al., 2000; Morrison, et al. 2004), as well as to effortfully inhibit attention to invalid relationships (e.g., Cho, Holyoak & Cannon, 2007; Richland & Burchinal, 2013). I suggest that having all visual representations available during structure-mapping reduced the working memory load required for participants to hold the representations active in mind, so they could use those resources more directly for structure-mapping.

In contrast, I suggest that having the representations presented sequentially may have imposed the highest burden on the executive function system, requiring students to effortfully inhibit attention to the misconception representation presented first. This representation was likely salient for its visual cues as well as for its coherence with prior knowledge (hence being a common misconception). So, suppressing the impulse to retain and use this representation as it was and rather to re-represent this information through structure-mapping may have been particularly effortful and thus successful less of the time. Performing analogical reasoning would have required EF resources to revisit the misconception in light of subsequent strategies to discard its validity. However, for the SV group the misconception was no longer visible throughout the comparison, thus, making it more difficult to identify misalignments between the appropriate strategy and the misconception. Students in the NV condition did not visualize any of the solutions, which may have created “desirable difficulties” (Bjork, 1994; Schmidt & Bjork, 1992) and potentially helped students focus more on the teacher’s verbal explanations that guided students to notice subtraction as an inappropriate strategy. Thus, NV students may have attempted to memorize the other strategies discussed more than the SV students. Yet, the WM load may have still been too great for NV students to develop a durable schema through structure mapping: reconstruct each solution to notice the alignments and misalignments between each strategy.

As dual coding theory would suggest, reinforcing exemplars through visual and auditory presentations leads to greater retention. Higher retention for the details of presented representations might explain why participants in the sequential condition were most likely to retain and produce the misconception at posttest, rather than showing evidence of schema formation – which would have been expected if the students performed structure mapping. There

is significant literature suggesting that people tend to use data in the world to confirm their biases, potentially leading towards difficulties in drawing appropriate inferences from analogies (Brown, 2013; Zook, 1991). Thus in this case this confirmation bias seems to have led participants to retain the misconception as presented.

In sum, these results provide insight into the role of visual representations in schema formation. Presence of visual representations can aid structure-mapping and schema formation when representations of *all* compared solutions are visible, in particular to improve conceptual understanding. However, having visual representations presented only sequentially can actually hinder structure-mapping, leading to retention of the details of the representations rather than the overarching schema. This is particularly evident in the situation tested here, in which one representation being compared is a common misconception. Presenting analogs sequentially increased usage of the misconception on posttest when compared to having the analogs presented simultaneously, which suggests that the visibility of the analogs may play an important role in either supporting or derailing structure-mapping.

Chapter 3: The Role of Executive Functions for Structure-Mapping in Mathematics

(Begolli et al., 2015)

Misconceptions are common throughout the curriculum, and researchers focused on the potential of analogies to overcome these through conceptual change have revealed the real challenges of teaching children to reconsider their misconceptions. For example Chinn and Brewer (1993) provide evidence that many students finish high-school and University without giving up pre-Newtonian perspectives of motion (e.g. Clement, 1982).

A common instructional recommendation to help students confront misconceptions is to directly contrast them with valid relational structures, (i.e. in this case, solution strategies; Dreyfus, Jungwirth, & Eliovitch, 1990; Jonassen, 2008; Posner, Strike, Hewson, & Gertzog, 1982). As was discussed in Study 2, identifying contrasts and similarities engages complex cognitive processes that place a burden on reasoners' working memory (WM) and executive functions (EFs). Learners use WM and EFs to perform relational structure-mapping: represent systems of relationships, align and map these systems to each other, and draw inferences based on the alignments (and misalignments) for successful schema formation (see Gentner, 1983; Gick & Holyoak, 1983; Morrison et al., 2011). At the same time, engaging with misconceptions without fully encoding the higher order relation between that misconception and the correct analog may also lead to reification of these intuitions (Begolli & Richland, 2015).

While eliciting learners' prior knowledge and potential misconceptions is recommended, Experiment 1 also shows; however, that not providing sufficient supports may also lead a large number of students to use the misconception. This was particularly true for students who saw solution strategies presented one at a time. Study 3 seeks to explore how the cognitive

underpinnings of analogy may be integral to undergoing conceptual change and engaging with key misconceptions in a productive and non-interfering way. Analogical reasoning and conceptual change both require reasoning about a familiar (base) and a less familiar (target) representation, and in this respect there seems to be an overlap in the cognitive mechanisms required for performing each. Perhaps a key difference between the two is that in conceptual change the structure of the familiar (base) representation is modified or changed based on contrasts with the less familiar (target) representation. In the case of persistent misconceptions, to induce shifts in their thinking, children need to evaluate their intuitions and inhibit irrelevant elements in order to focus on elements that align with the correct representation. For instance, children need to align mammal properties of a dolphin (breathing through air), while inhibiting fish properties of dolphins (swim in water) to understand that a dolphin is not a fish. Similarly, when reasoning by analogy, children need to inhibit distractors, in order to map the correct relations between source and target. For example, in Figure 5 (taken from Richland, Morrison, & Holyoak, 2004), children need to inhibit the cat in the right figure in order to correctly map the relation of the boy as the chaser and being chased to the cat in the left figure.

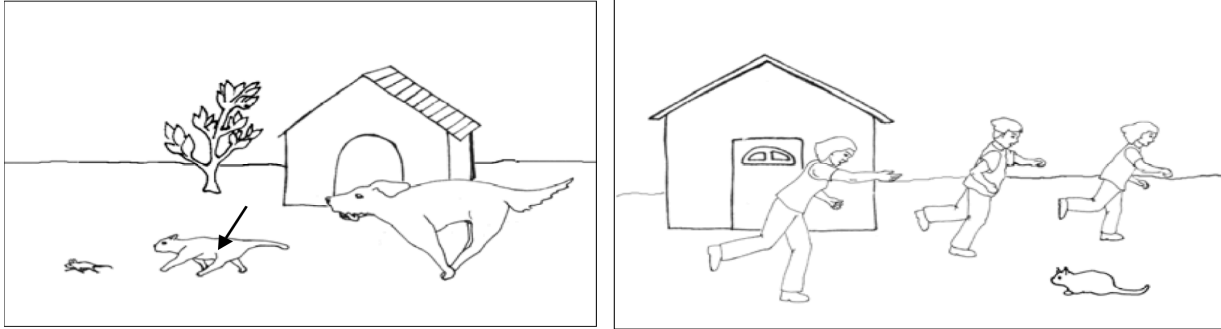


Figure 5. In the picture on the left the cat (both the chaser and the chased) corresponds relationally to the boy in the picture on the right (both the chaser and being chased). A common error young children make is to be distracted by the perceptual cues matching the cat in the left with the cat in the right.

Research in both analogical reasoning and conceptual change posit EFs as key in the success of analogy and long lasting change (Houdé et al., 2011; Richland et al., 2004; Waltz et al., 2000). Thus, I explore the hypothesis that children’s success in overcoming misconceptions through comparisons with correct analogs may vary based on limitations in children’s developing EF (see Waltz et al., 2000). Because misconceptions are often deeply embedded in intuitive beliefs, drawing a higher order relation between a misconception and a correct analog to form a valid schema is highly effortful and requires a combination of executive functions.

The Relationship between Analogy, Conceptual Change, Working Memory, and Inhibitory Control

Working memory is argued to be necessary for representing systems of objects (e.g. steps to solution strategies) and re-representing these systems of relationships in order to align and map their structures. Successful mapping and alignment requires flexible switching between

these systems of relations to attend to relevant elements within each system and inhibit irrelevant elements to identify meaningful similarities and differences, in order to derive conceptual/schematic inferences from this structure-mapping exercise and better inform future problem solving (see Morrison et al., 2011). Thus, limitations of EFs – working memory, task switching, and inhibition throughout this reasoning process could explain failures in schema formation through structure-mapping.

Current models describe working memory and inhibitory control (IC) as components that share responsibility for carrying out the three core components of executive functions: a) *Shifting* (or task switching), b) *Updating*, and c) *Inhibition* of pre-potent responses (Friedman & Miyake, 2004; Miyake et al., 2000). In this case, one can think of IC as a system that helps us in a problem solving situation to suppress or *inhibit* our initial, but less efficient (or incorrect) solution and *switch* our attention towards a perhaps less intuitive, but more adequate strategy. Whereas, WM is responsible for holding information in mind, searching for information in long-term storage, and *updating* it (Miyake et al., 2000). The exact nature of WM and IC processes are complex and hotly debated (Wright & Diamond, 2014). However, the field seems to be favoring the idea that WM & IC have distinct and common roles within executive functions. Thus, IC may have a distinct role in inhibiting pre-potent responses and irrelevant information. Inhibitory control also shares common roles with processes of *switching* and *updating* from irrelevant to relevant tasks/representations (Miyake et al., 2000). For the purposes of simplicity, the term inhibitory control (IC) will be used to refer to both distinct and common processes of suppressing responses, and the term *Inhibition* to refer to the distinct process of suppressing pre-potent responses.

To examine the relationship between IC and learning from misconceptions embedded in instructional comparisons, in Experiment 3, the following four tasks are utilized: Forward Digit Span, Backward Digit Span, Hearts & Flowers, and a Stop Signal Task. This is because they measure the core components of executive function (short-term memory, working memory, and task-switching, and inhibition) which are posited to be implicated in reasoning by analogy and conceptual change (Houdé et al., 2011; Morrison et al., 2011).

The Forward Digit Span task (repeat numbers in the same sequence as presented) is used to assess the ability to hold information in mind for a short time otherwise known as short-term memory. The Backward Digit Span (repeat numbers in reverse sequence of presentation) is used to assess participants' working memory – the ability to manipulate information in short-term memory. Thus, participants both need to keep an item in mind then manipulate the information in order to repeat it in reverse order. The hearts & flowers task is utilized to measure students' broader IC processes, such as *task switching* and *inhibition*, while the stop signal task is used as a purer measure of *inhibition*. These measures are further described in the Methods chapter.

The general findings from studies that have used WM/IC measures suggest that participants who score higher in measures of WM/IC tasks have higher IQs, score higher in scholastic achievement tests (e.g. SAT; Duncan et al., 2007) and analogical reasoning tasks (Richland and Burchinal, 2013), and undergo conceptual change at an earlier age (Houdé et al., 2011). Further, early performance on EFs, particularly inhibitory control measures predict future outcomes in a range of fields. However, there seems to be a gap in understanding the role of WM/IC in the process of learning, as distinct from performance. Most studies take WM/IC measures of participants and then examine their correlations with performance on achievement tests (e.g. SAT). In contrast, in the current study I took measures of WM/IC, gave participants a

pretest, intervened with a learning event, and then assessed their learning on two posttests, one administered immediately after the intervention, and after 1-week-delay.

Overall, there have been two common approaches for understanding the relationship between EF/IC and thinking. In the first approach, researchers have looked at correlations between individual differences in children's WM/IC abilities with common tasks of analogy or conceptual change (for analogy, see Richland and Burchinal, 2013; for conceptual change, see Houdé et al., 2011; for mathematics and IQ see Alloway & Alloway, 2010; Gathercole, Pickering, Knight, & Stegmann, 2004b; Gathercole, Alloway, Willis, & Adams, 2006). In the second approach, researchers use an experimental design where the demands of task are manipulated to increase the demands of EF/IC. The current study combines aspects of both approaches to exploit the strengths of correlational and experimental design, described next.

EF and Instructional Analogy

In Experiment 1, students were presented with a common misconception followed by two correct solution strategies and examined whether presenting analogs either simultaneously, sequentially, or only verbally would support structure-mapping in a mathematics lesson based on instructional analogy. While students' schema formation was best supported when analogs were visible throughout the structure-mapping, sequential presentations of analogs led to the lowest performance suggesting that object-level encoding of misconceptions interfered with schema formation, perhaps due to limitations of EFs (Begolli & Richland, 2015). Sequential presentation of analogs may place a greater strain on EF resources, potentially revealing EF mechanisms responsible for structure-mapping failures.

This study examines correlations between schema formation from sequential presentation of analogs (as in Begolli & Richland, 2015) and individual difference measures of EF – particularly working memory processes (WM; short-term and domain general WM) and inhibitory control processes (IC; response inhibition and task switching). Working memory is likely to facilitate the manipulation of relational systems while holding them in mind and IC is hypothesized to decrease distractional elements within these systems, enable disattention to an intuitive misconception, and aid in switching between relations to derive appropriate schemas.

This work has the potential to contribute to both the theoretical understanding of the role of EFs in successful structure-mapping within the ecologically valid context of a classroom as well as practical implications for designing technology and instruction.

Experiment 2

Method

Participants

Participants were 107 fifth graders (44 girls) drawn from a public school with largely homogenous population (78% White, 11% Hispanic or Latino, 11% other; 12% disadvantaged SES; 2% English Language Learners). Fifteen to seventeen students either missed a test or a cognitive measure or both due to absences. Ten additional participants were dropped from analyses because their pretest scores for procedural & conceptual knowledge were at ceiling (100%). The maximum number of participants at each test point and cognitive measure was included in the analyses ($n = 82-80$) with age range 10-11 years old. The *design, procedures, and measures* remained the same as in Experiment 2, apart from the changes described next.

Design & Procedure

The materials for the intervention were the same as in Experiment 1. The mathematics assessment was highly similar to Experiment 1, but only included the items that captured the most variance (see Appendix B). Three items were added and four were modified. Only one item was dropped from the flexibility measure since it showed no inter-item correlation ($r \sim 0.02 - r \sim 0.04$). The reliability of the new mathematics assessment is summarized in Table 9. The coding & scoring mimicked Experiment 1 and the results of inter-rater reliability were between 92-98% across the three primary constructs. Also, the video-lesson was not embedded in a computer program; however, it remained interactive such that screen prompts were embedded in the video. To place a greater WM/IC strain on students only the Sequentially Visible version of the video-lesson was used.

All participants followed the same procedure. Day 1: pretest and individual difference measures of WM (Forward and Backward Digit Span; Hearts & Flowers Task). Day 2: (2 days later), interactive instructional video as the intervention, followed by an immediate posttest. Day 3 (1 week later): delayed posttest and EF measures (Stop-Signal Task).

Table 9
Inter-item Cronbach's alpha values at each test for each construct

	Pretest	Immediate	Delayed
Procedural	0.72	0.84	0.87
Flexibility	0.55	0.73	0.79
Conceptual	0.76	0.78	0.84

In each class, 24 students completed the cognitive tasks first, then either took a pretest or a delayed posttest. The remainder of the students in the class (8-10 students) completed the cognitive tasks and pretest in reverse order. This was due to a lack of netbooks available for

administering the tasks. The cognitive measures used in Experiment 2 remained the same, apart from adding 12 mixed trials to the hearts and flowers task.

Working Memory and Inhibitory Control Measures. The WM and IC measures were used to determine whether individual differences in students' processing resources were related to their baseline performance and learning gains as a result of the video-lesson. I used the following computerized versions of WM and IC: Forward Digit Span, Backward Digit Span, Hearts and Flowers, and a Stop Signal Task.

Forward and Backwards Digit Span. The forward and backwards digit span are derived from the Automated Working Memory Assessment (AWMA) battery (Alloway, Gathercole, Kirkwood, & Elliott, 2009; Alloway, 2007), with superb construct validity, standardized on 1,470 children 5 to 6 years and 1,719 children ages 8 to 9 years (Alloway et al., 2009; Diamond, 2013). The Forward Digit Span (repeat numbers in the same order) is a measure of short-term memory, whereas the Backward Digit Span (repeat numbers in reverse order) is a measure of EF/WM. On both tasks, the string of numbers increased for each correct trial and numbers were shown serially such that only one number appeared on the netbook screen at a time. Students responded using a number pad that was attached to the netbook. Each student started with 3 practice trials. After practice trials, students had to respond correctly in order to increase the string of numbers. If students answer incorrectly on a specific set of strings twice, the program ends, and the maximum number of strings in the previous set is recorded (Table 10). For example, in Table 10, the recorded number would be 4. This number is used as a dependent measure on both the Forward Digit Span and the Backward Digit Span.

Table 10

Illustration of the Forward Digit Span. The Backward Digit Span works in the same fashion, but the numbers need to be repeated in backward order

		Prompt	Response	
3 digits	Trial 1	7, 9, 3	7, 9, 3	correct
	Trial 2	5, 3, 9	5, 3, 9	correct
4 digits	Trial 1	4, 1, 2, 5	4, 1, 2, 5	error
	Trial 2	3, 8, 2, 0	3, 8, 2, 0	correct
5 digits	Trial 1	4, 6, 3, 1, 9	4, 3, 1, 3	error
	Trial 2	0, 2, 5, 1, 4	0, 2, 1, 4, 3	error

Hearts & Flowers Task. The Hearts and Flowers task (H&F) is a version of the Dots task taken from the Directional Stroop Battery used to assess EF (adapted from Wright & Diamond 2014). This was administered on day 1. Students were presented with either hearts (congruent) or flowers (incongruent) on each trial (Figure 6). For incongruent trials, the correct response is aligned with students' natural inclination – “press the button on the same side (left or right) as the heart.” For incongruent trials, the correct response goes against what comes naturally – “press the button on the opposite side (left or right) of the flower.” Trials are presented in 3 phases. Phase 1 – congruent trials only, phase 2 – incongruent trials only, phase 3 – mixed trials presented randomly.

To perform this task students are expected to hold each task in mind (short-term memory), switch between tasks to choose the right answer (task switching), and inhibit their prepotent response (see Wright and Diamond, 2014). The dependent measure was the difference in time it took to respond to a trial when participants had to change the rule versus a trial when participants did not have to change the rule to respond – known as switch cost response time.

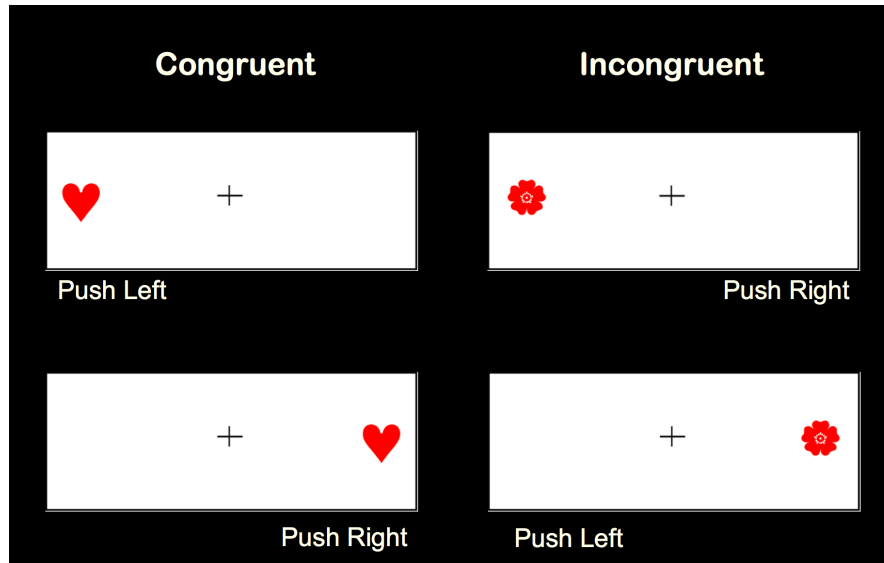


Figure 6. The hearts & flowers task. Congruent trials are shown on the left and incongruent trials are shown on the right (Wright and Diamond, 2014).

Stop-Signal Task. Stop-Signal Task (Administered Day 3; Figure 7). The Stop-Signal task (SST) measured participants' response inhibition. Students are presented with a fish for 850ms (go stimulus) or a fish followed by a manta ray (stop-signal, occurring on 40% of the trials). Students were instructed to press a button ("A" or "L") to send the fish home (within 850ms) unless the Manta Ray appeared, in which case they had to withhold from pressing any buttons. The sooner the Stop-Signal appears after the go signal, the easier it is to inhibit a response – this temporal difference is known as the Stop-signal Delay (SSD). SSDs are initially short but are increased following accurate trials. An adaptive algorithm ensured SSD increased to Final SSD length was used as a dependent measure (Bissett and Logan, 2012).

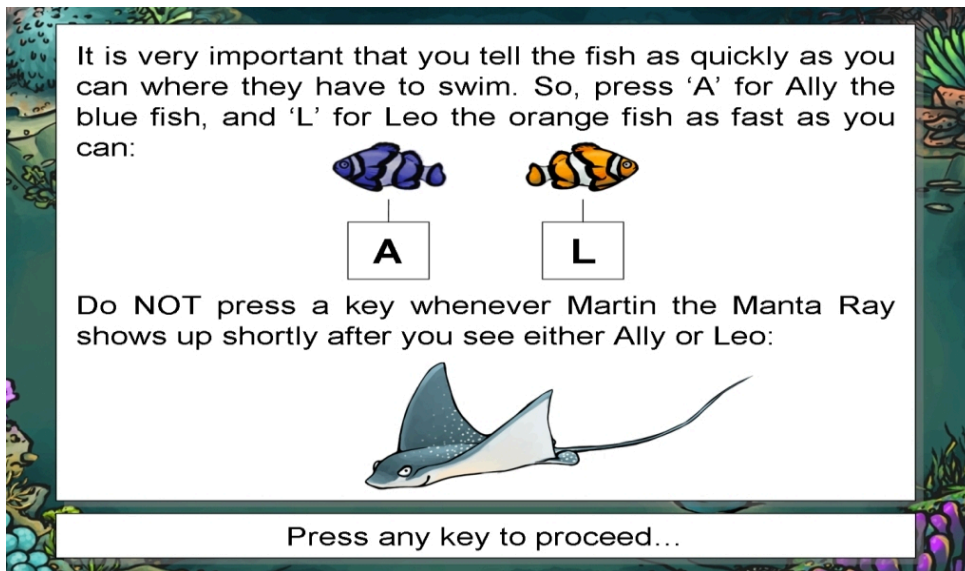


Figure 7. The Stop-Signal “game” instruction screen. The task of the students is to press the corresponding key quickly enough to “send” the fish home, but to not press the key if the manta ray fish appears. The manta ray appeared at random on 40% of the trials. Adapted from Bissett & Logan (2012).

Task Impurity of EF measures. IC/EF measures may suffer from *task impurity*: the inability of a task to measure a cognitive process in isolation from other cognitive processes (Miyake et al., 2000). Task impurity raises important concerns for understanding the contribution of IC processes on learning from instructional analogies. First, it is unclear how much overlap exists between WM & IC processes required by the tasks at hand and learning from instructional analogies. To address this, researchers have created factor variables by extracting common variance between measures of multiple EF/IC tasks. For instance, several measures of IC have in common that they assess IC performance, but they do not share the task specific demands that dilute the measurement of IC. By creating a factor variable, only the commonalities between the measures (i.e. common variance) are used to account for learning from instructional comparisons.

On the other hand, WM and IC are complex and isolating the contribution of each task also provides a lens in the contribution of each process within WM/IC. While the H&F task may recruit broader IC process, the Stop Signal task, targets only Inhibition. Thus, a separate approach for isolating the contribution of distinct IC processes on learning is to use tasks that target specific processes in IC, such as *Inhibition*: the intentional suppression of automatic, pre-potent responses. For instance, for the hearts and flowers task students are required to switch between two rules (“press on the same side” vs. “press on the opposite side”). The ability to switch between rules may share commonalities with IC, but is distinct from *Inhibition*. Thus, it is unclear how much distinct processes within IC contribute to learning, separately from common IC processes.

To address these concerns, I utilized the Stop-Signal task to target specific processes of *Inhibition* to target common IC processes within task switching. The Stop-Signal measures *Inhibition*, where participants have to respond (e.g. press a key) when presented with a “go” signal stimulus, but inhibit their response when the “go” signal is presented in conjunction with another “stop” signal stimulus. Similarly, to isolate processes of maintaining information in mind in the process of *updating* both though to be common to WM I used two measures, one for short-term memory and another as domain general working memory, described next.

Results

Baseline Analyses

Executive Functions share *commonalities*, but also have *diverse* functions, for controlling thought and behavior (Miyake et al., 2000). To understand whether the contribution of each cognitive measure was separable or unitary I conducted a factor analysis with a varimax rotation

on all measures (FDS & BDS, H&F, and SST; Table 11). Combining measures also reduces task specific variance and allows examination on a construct level, rather than on an individual task level. The theoretical expectation was to derive two distinct factors sharing common variance, thus, the analysis was restricted to two factors. A WM factor to account for the common contribution of short-term and domain general working memory processes (comprised of the FDS & BDS) and an IC factor accounting for the common contribution of response inhibition and task switching processes (comprised of the H&F and SST). The results of the factor analyses confirmed these predictions with both factors displaying similar loadings, which explained 65.1% of the total variance. The raw scores for the WM and IC measures were converted into z-scores for subsequent analyses.

Table 11
Factor Loadings and Descriptives

<i>N</i> = 96	WM Factor	IC Factor	Mean	SD
FDS	0.83	0.01	6.05	1.11
BDS	0.78	0.06	5.36	1.08
H&F	0.15	0.77	110*	169
SSD	-0.07	0.82	284	164
% of Variance	33.2%	31.8%		

*Median RT used & reported here. H&F and SSD are in ms

To examine the contribution of broader WM and IC as well as to unpack the contribution of each cognitive process, I conducted separate regressions on each mathematics construct (PK, PF, CK, and CM) for three models at pretest, immediate, and delayed test, summarized in Table 12.

The first model examines the role of WM & Inhibition as key processes in EF on each mathematical construct separately. Model 2 unpacks the role of WM by examining the individual contribution of short-term (FDS) and domain general WM processes (BDS). Model 3 unpacks the role of IC by examining the individual contribution of task switching (H&F) and response inhibition (SST).

Table 12
Regression models conducted in analyses. A separate regression was conducted for each mathematical construct

	Model 1 (<i>N</i> = 80)	Model 2 (<i>N</i> = 82)	Model 3 (<i>N</i> = 82)
Dependent Measure	Math Construct	Math Construct	Math Construct
Indicators	WM Factor IC factor Pretest score*	FDS BDS Pretest score*	H&F SST Pretest score*

*Only utilized at immediate and delayed posttests.

Tables 11 and 13 summarize the mean scores of the cognitive measures and mathematical constructs. The math mean varied slightly depending on completeness of measures, Table 13 summarizes math scores for students who completed all measures. The overall trend between the Sequentially Visible condition in Experiment 1 and Experiment 2 is similar (see Table 4), though students in Experiment 2 show an advantage of about 18% in conceptual knowledge at pretest. This could reflect historical changes due to the implementation of common core standards the year following Experiment 1, or an overall higher SES population.

Relationship between EF and Mathematics

Regression results with beta values for all Models on each mathematical construct are summarized in Table 14 and the respective effect sizes in Table 15.

Irrespective of cognitive ability, students improved from pretest to immediate and delayed posttest on PK, PF, and CK ($F > 10, p < .001$), but no overall difference on how much students' used the misconception ($F = 1.04, p = .23$). But differences in students' EF may reflect distinct patterns in their math outcomes. The remaining results will be discussed by presenting the relationship between WM and each math construct from Model 1, then I discuss Model 2 to unpack the contribution of each component within WM, *Short-Term Memory* (FDS) and *domain general WM* (BDS) on each construct. Similarly, I discuss the relationship between IC and each math construct from Model 1, and then in Model 3, unpack the contribution of *Response Inhibition* (SST) and *Task Switching* (H&F) within IC. Effect sizes for each component ranged from small ($\eta_p^2 = .02$) to moderate ($\eta_p^2 = .14$). Model 2 and Model 3 analyses were exploratory and thus there was no family-wise error correction, however the results should be interpreted with caution.

Table 13
Mean scores per construct for students who completed all measures
($N = 80$), SD in parenthesis.

	Pretest	Immediate	Delayed
Procedural	22% (0.25)	49% (0.35)	44% (0.36)
Flexibility	11% (0.12)	28% (0.22)	24% (0.22)
Conceptual	31% (0.26)	43% (0.29)	44% (0.31)
Misconcept.	26% (0.24)	20% (0.23)	24% (0.26)

Working Memory. Students' WM ability does not seem to predict pretest performance, though when unpacking the WM factor, BDS performance was positively related with higher uses of the common misconception ($\eta_p^2 = .06$).

Table 14

Beta values from regression models described in Table 12. Pretest beta values were always significant, $p < .05$; not shown here.

Model 1	Pretest		Immediate		Delayed	
	WM	IC	WM	IC	WM	IC
Procedural	4.08	-0.02	2.65	5.70†	8.49**	10.76***
Flexibility	0.76	-0.14	3.70	4.96*	4.58*	5.30*
Conceptual	0.65	7.18*	6.20*	3.20	6.99*	6.68*
Misconcep.	2.38	3.27	-3.76	-6.56*	-3.58	-6.91*
Model 2	FDS	BDS	FDS	BDS	FDS	BDS
Procedural	3.01	2.14	2.34	0.64	7.18*	3.04
Flexibility	1.11	-0.26	1.92	3.12	2.12	4.15
Conceptual	0.82	-0.63	8.11**	0.03	2.96	5.70†
Misconcep.	-3.00	6.63*	-3.91	-0.62	-2.62	-1.10
Model 3	SST	H&F	SST	H&F	SST	H&F
Procedural	-3.92	3.85	4.66	3.23	8.01*	6.74*
Flexibility	-0.80	0.80	2.87	3.68	5.41*	2.35
Conceptual	5.26†	4.29	2.08	2.84	3.68	5.43†
Misconcep.	4.73†	-0.62	-1.86	-5.14*	-1.23	-7.41**

† $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

At immediate test, overall WM ability was positively related with conceptual knowledge performance, which seems to be largely driven by advantages in students' FDS scores ($\eta_p^2 = .09$).

At delayed test, students' with higher WM factor scores had overall higher outcomes in procedural knowledge ($\eta_p^2 = .09$), procedural flexibility ($\eta_p^2 = .05$), and conceptual knowledge ($\eta_p^2 = .08$). When looking at the individual contribution of each WM component, only students with higher FDS scores had higher scores in procedural knowledge ($\eta_p^2 = .05$). While the relationship between students with higher BDS scores and conceptual knowledge scores was not significant, it suggested a positive trend ($p = .052$; $\eta_p^2 = .05$).

Table 15
Effect sizes for each predictor from Model 1, on each mathematics outcome, shown as partial eta-squared. Effect sizes in bold type reflect significant predictors.

Pretest	Pretest Score	WM	IC
Procedural	-	0.03	0.00
Flexibility	-	0.00	0.00
Conceptual	-	0.00	0.08
Misconception	-	0.01	0.02
Immediate Posttest			
Procedural	0.29	0.01	0.04
Flexibility	0.18	0.03	0.06
Conceptual	0.23	0.06	0.02
Misconception	0.18	0.03	0.06
Delayed Posttest			
Procedural	0.32	0.09	0.14
Flexibility	0.12	0.07	0.07
Conceptual	0.29	0.08	0.07
Misconception	0.08	0.02	0.07

Inhibitory Control. At pretest, students' with higher overall IC scores reflect an advantage only in their conceptual knowledge performance ($\eta_p^2 = .07$). At immediate test, students' with better IC scores use the misconception less ($\eta_p^2 = .06$), which is positively related to their task switching performance measured by the H&F task ($\eta_p^2 = .06$). IC students were also better in procedural flexibility ($\eta_p^2 = .06$). At delayed test, students with higher scores in IC display an advantage in their procedural knowledge ($\eta_p^2 = .14$), procedural flexibility ($\eta_p^2 = .07$), conceptual knowledge ($\eta_p^2 = .07$) constructs and a reduction in their use of misconceptions ($\eta_p^2 = .07$). In these cases, students with higher SST scores are better in procedural knowledge ($\eta_p^2 = .07$) and procedural flexibility ($\eta_p^2 = .07$). Students with higher H&F scores are better in procedural knowledge ($\eta_p^2 = .06$) and use the misconception less ($\eta_p^2 = .08$). The relationship

between IC and conceptual knowledge could be driven by students' task switching performance, though this relationship is not significant ($p = .057$; $\eta_p^2 = .05$).

To help interpret the data from another perspective and for illustration purposes, I divided students based on their WM & IC scores into high (top 25%), medium (middle 50%), & low (bottom 25%) performers (Figure 8).

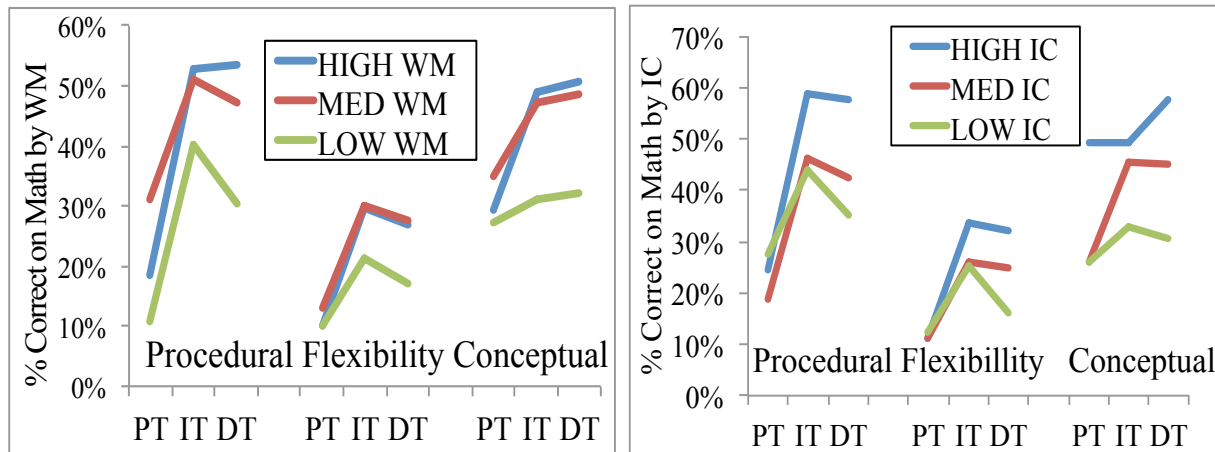


Figure 8. Mean scores for Pretest (PT), Immediate (IT) and Delayed (DT) tests by WM (left) and IC (right) score.

The regression results suggest a continuous progression between low, medium, and high performers, such that students with a 1-point advantage in WM or IC score have advantages on their mathematics performance ranging from roughly 20%-29% higher than the overall mean. A qualitative examination of the WM data suggests that this effect is driven by a difference between low and medium/high WM performers, with the largest differences on the conceptual understanding measures. Procedural knowledge has been proposed to be a preliminary step for attaining conceptual knowledge, though both reinforce each other iteratively (Rittle-Johnson, Siegler, & Alibali, 2001). It appears that low WM students show some improvements in their procedural and flexible knowledge at immediate test, but these gains decrease by delayed test, perhaps reflecting a level of procedural knowledge that is insufficient for retention, nor for

attaining a broader schema for ratio – reflected in conceptual understanding measures. This perspective reinforces the role of domain general WM processes (in contrast to short-term processes) as critical for durable schema formation, but may also indicate that there is a certain threshold for WM ability required. Thus students must have adequate WM for schema formation, but performance may not be affected if their WM ability passes this threshold.

In contrast, qualitative examinations of the IC measures lend themselves towards interpreting a more continuous relationship between IC ability and students' mathematics outcomes. Students with high and medium IC scores use the misconception more before the lesson (pretest $M = 28\%$) than after the lesson (delayed $M = 19\%$), while medium IC students remain about the same (pretest $M = 28\%$, delayed $M = 25\%$). An inverse relationship seems true for low IC students (pretest $M = 17\%$, delayed $M = 27\%$). Further research is needed in order to clarify the interpretations stemming from these exploratory perspectives.

In sum, both WM and IC predict procedural, flexible, and conceptual knowledge at delayed test and IC also predicts a reduction in misconceptions. These effects are less apparent at immediate test, suggesting that WM and IC may be particularly important for gaining a deeper, more schematic understanding of concepts, which in turn may promote flexible knowledge and retention of procedures. Negative correlations between IC and the misconception implies that inhibitory processes may be required to reduce misconceptions.

Discussion

This study clarifies the contribution of EF abilities for schema formation of mathematics concepts through instructional analogies. Many studies have examined the relationship between

EF and broader mathematics achievement (St Clair-Thompson & Gathercole, 2006), but there is little work done investigating the specific role of EF on *learning* mathematics through structure-mapping. Unlike previous accounts that have shown relationships between EFs and structure-mapping (Krawczyk et al., 2008; Morrison et al., 2011; Richland & Burchinal, 2013; Waltz et al., 2000; Zelazo, Muller, Frye, & Marcovitch, 2003) the pretest, intervention, posttest design in combination with the EF measures gives insight into the role of specific EFs throughout the trajectory of schema-formation by structure-mapping.

In addition, the WM & IC data align with current views that WM & IC are separate processes within EF, each explaining distinct variance (Miyake et al., 2000). These data reveal that within WM, the short-term and general working memory processes share commonalities, but each also accounts for distinct variance in an everyday analogical learning context. Similarly, within IC, response inhibition and task switching share commonalities, but also account for distinct variance in learning from analogy.

The results reveal that broader WM and IC processes predict learning in this instructional context. Both WM and IC predicted the retention of procedural knowledge, procedural flexibility, and conceptual knowledge, and IC also predicted the reduction of students' use of the misconception, reflected by delayed test results. EF resources (WM and IC) may matter most for durable schema formation, while their effect may be less evident for short-term learning, as evidenced by their more limited prediction of performance at immediate test.

In the short term, it seems that the relationship between WM and conceptual knowledge is largely influenced by short-term memory processes, whereas the relationship between students' use of misconceptions and their IC ability seems to be driven by students' task switching performance. A possible explanation is that in the structure mapping process, short-term memory

processes facilitate the representation of systems of relations, whereas task switching processes (which include inhibition) help reasoners attend to structural dissimilarities between these systems. However, at immediate test, it may be hard to distinguish between recency effects/object-level encoding and successful schema formation, which could also obstruct from understanding the role of WM and IC (and/or individual functions within WM and IC) in the long term. Thus, examining delayed test results provides better data on the role of WM and IC on successful structure-mapping.

A closer inspection of delayed tests results suggests that WM and IC components have the most predictive power when considered in tandem, as their individual contributions wane when considered separately. In terms of WM components, short-term storage seems to be related to only procedural knowledge, whereas general WM seems somewhat related to the attainment of conceptual knowledge, though not significant perhaps due to low power ($p = .052$).

A closer examination of separate IC processes allows for hypothesis generation about the specific EF resources and their relations to learning. While these data should be interpreted with caution, the data patterns suggest that students that are better on the response inhibition task (SST) gain more in procedural knowledge and are more flexible with procedures. Further, students who perform better on task switching (H&F) continue to use misconceptions less – and there is somewhat of a relationship between task switching and conceptual knowledge, though only marginally significant ($p = .057$).

In sum, in an ecologically valid learning context, the data provide evidence that individual differences in EF may impact whether students successfully benefit from a structure-mapping opportunity comparing a misconception to correct solutions. Teachers wishing to confront students' misconceptions, thus, may be helping students with high EF resources while

harming those with low EF resources when sequentially presenting these analogs in their lessons. Simultaneous presentations of analogs may reduce the disparity in schema-formation due to individual differences in EF, but this remains to be tested.

Chapter 4: General Discussion of Theoretical and Practical Implications

The findings from these studies have the potential to positively shape U.S. teaching practices as well as contribute to several areas of cognitive scientific literatures. From a theoretical standpoint, these findings extend previous laboratory-based results on analogical learning, conceptual change, and EF using a video-based methodology with high ecological validity.

Implications for Theory

The studies described in use a novel theoretical perspective and methodological approach to bridge research in laboratory and classroom settings. Experiments 1 and 2 capitalize on the advantages of video to arrive at causal relationships between specific teaching practices and student learning in everyday classroom settings, ideally allowing for greater generalizability of the findings. Theoretically, findings from Experiment 1 support previous laboratory-based results indicating that visual representations can support schema formation and learning from analogy (Gick and Holyoak, 1983), and extend them to an applied setting. In addition, the work extends studies of visual representations to examine the role of visual representations on schema formation when relational analogs include a misconception, which are mostly unexplored in prominent structure-mapping models (Gentner & Forbus, 2011; Gentner, 1983). Conceptual change literature (Carey & Spelke, 1994; Chi, 2013; Vosniadou, 2013) has investigated how people overcome misconceptions in the context of science education (Brown & Clement, 1989; Brown, 2013; Chinn & Brewer, 1993), and recently in mathematics (Vamvakoussi & Vosniadou, 2012) but the influence of misconceptions on structure-mapping models remains to be fully

defined. So, this study has potential to contribute to both the conceptual change and analogy literatures.

Findings from Experiment 2 reveal that students' working memory and inhibitory control are critically involved in durable learning from misconceptions embedded in instructional comparisons. This was investigated within the context of presenting analogs sequentially, which may place a significant strain on students' EF resources, but provides a lens into the contribution of WM and IC in the process of learning, in contrast to achievement. Many studies have investigated the effects of EF on broader mathematics achievement (St Clair-Thompson & Gathercole, 2006), but this represents one of few studies that investigates the role of EF on a more proximal level in the trajectory of change, using a pretest, intervention, posttest design. As such, it also contributes to a more nuanced understanding of the processes responsible for performance failures in analogical reasoning, attributed to EF (Krawczyk et al., 2008; Morrison et al., 2011; Richland & Burchinal, 2013; Waltz et al., 2000; Zelazo et al., 2003).

Working memory and inhibitory control data suggest these processes are separate from each other, confirming current models of EF (Miyake et al., 2000). While both WM & IC explained distinct variance for long-term change in procedural knowledge, procedural flexibility, and conceptual knowledge, only IC was related to the reduction of misconceptions. Both WM & IC involve sub processes, but isolating them within WM and IC may be more difficult due to task impurity and the practicality of administering multiple tests within already tight classroom schedules. Nevertheless, the data point towards the interpretation that short-term memory is related to learning procedures while domain general working memory is related to changes in conceptual understanding. Within IC, students' ability to inhibit pre-potent responses and switch between tasks predicts changes in procedural knowledge. Further, task switching ability accounts

for a reduction in the use of misconceptions and only inhibition ability accounts for changes in the flexible use of procedures.

In light of delayed test data, it appears that short-term memory processes (FDS) may be important for initial schema-formation and for later recall of the appropriate procedures, whereas general WM processes as measured by the BDS are more important for long-term generalizable knowledge. Perhaps students with better FDS scores had greater resources to represent the systems of relations during the structure-mapping processes, but only those students with better BDS scores were able to re-represent these systems for appropriate alignment and mapping between the source and target relations, leading to a more durable schema.

On the other hand, students who were better at response inhibition (SST) and task switching (H&F) may not notice their advantage immediately, but these processes may be crucial for long-term schema formation. It could be the case that better response inhibition during the structure-mapping process aids students' WM to attend to appropriate representations by reducing interference from competing and inappropriate representations (in this case ratio concepts over subtraction – the common misconception). Thus, leading towards increased procedural knowledge and flexibility. Another interpretation, though not mutually exclusive, is that response inhibition is responsible for reducing competing representations and selecting the correct representation at the time of the assessment 1-week later (though this may also imply reductions in the use of misconception, not reflected in the data).

It appears that task-switching processes operate at a higher level such that at every switching point, response inhibition may be required to select the appropriate task. This process seemed likely to lead towards an increase in conceptual understanding and a reduction of misconceptions. A possible explanation is that in order to identify relations that structurally align

in the source and target representations, the reasoner has to repeatedly switch between these representations while inhibiting distracting information in order to successfully map their structural relations. Overall, the data from WM and IC measures align with previous neurological and behavioral data, and computational models suggesting a similar role for inhibitory control (e.g. LISA; Morrison et al., 2011; Zelazo et al., 2003; Waltz et al., 2000; Krawczyk et al., 2008; Richland & Burchinal, 2013). Previous behavioral data have suggested that increases in relational complexity within analogs would place a higher demand on children's EF resources (Halford, Andrews, Dalton, Boag, & Zielinski, 2002). Also, populations with compromised EF resources (e.g., damaged PF cortex) or strained EFs (students performing dual-tasks during structure mapping; Waltz et al., 2000), and younger children (Richland et al., 2006) are more likely to fail at structure-mapping. Broader EF and IC at 54-months have been found to predict analogical reasoning at age 15 (Richland & Burchinal, 2013). Thus, there is mounting evidence converging on the importance of WM and IC as underpinnings of analogical reasoning.

Implications for Practice

From an instructional perspective, utilizing teaching by comparison is critical for learning deep mathematical conceptual knowledge (Rittle-Johnson and Star, 2007, 2009; Star and Rittle-Johnson, 2009; National Mathematics Panel, 2008). While showing visual representations and making mathematical comparisons is common to everyday mathematics instruction, teachers in the U.S. rarely scaffold instructional comparisons adequately (Richland, Zur, and Holyoak, 2007; Hiebert et al., 2005). This may be partly due to a lack in specificity of recommendations on how to improve these practices (Hiebert et al., 2005) and the difficulty in sustaining major changes

due to the culture of instructional routines (e.g. Stigler & Hiebert, 1999). The results from these studies provide specificity in moving towards recommendations for teachers regarding optimal use of visual representations during instructional comparisons – particularly for leading discussions about multiple ways of solving single problems. Thus shifting to leave all source and target representations visible throughout a full mathematical discussion, rather than only while they are first being presented, requires only a re-organization of existing routines rather than a large time investment and modification of current practice. Thus, the intervention described in Experiment 1 is feasible for integration into current teaching practices.

One must note that we cannot interpret the results of Experiment 1 to indicate that making analogs visible simultaneously will *always* lead to successful structure-mapping and mathematical schema formation. Only that if the analogs being compared are informative and the learner notices their relationship does making them visible simultaneously aid in abstraction. Thus, key to improving educational practice is certainly ensuring the instruction uses optimal structured analogs, and ensuring that any misconceptions are identified and compared well with an alternative and more accurate representation.

Making teachers aware of the cognitive processes that children undergo when comparing multiple solutions to a single problem may help teachers scaffold students throughout each step of the comparison. Namely, teachers can highlight elements within each solution as a system of relationships, then guide students to notice the relevant elements in each solution strategy. Next, teachers can help students map between the solutions by highlighting elements in each solution and helping students notice whether these are in alignment and/or misalignment.

Executive functions develop from early childhood throughout adulthood, thus, teacher awareness may be particularly important in understanding the limitations of children's EF

resources in relationship to structure-mapping. The metacognitive awareness about the reasons *why* comparisons are difficult, *what* needs to be aligned and/or misaligned, and *how* to support structure-mapping may positively influence adoption and potential sustainability of the recommended practice: making compared solutions visible simultaneously.

A primary constraint to implementation of making representations visible throughout lessons is space. When co-designing this lesson with teachers I faced the challenge that teachers often use their presentation space (e.g. white boards) for many purposes including daily schedules and reminders, which may reduce the amount of space available to leave multiple representations visible. This challenge is compounded by the trend to reduce presentation space through the use of such technologies as electronic whiteboards, such as *innovative white board technologies* (IWB; De Vita, Verschaffel, & Elen, 2014). These innovations enable teachers to control the board from their computer in a dynamic fashion, allowing for advanced preparation or careful design of visual representations, which can be a great strength. However there is also typically less room to make multiple representations visible, because these boards are about a third of the size of typical classroom chalk or white boards. These data suggest that IWB's (e.g. Smart boards) have the potential to be highly effective at instantiating single visual representations at a time, much as in the sequentially visible condition, which led to the lowest learning gains and greatest rate of misconceptions. Findings from Experiment 2 suggest this could be particularly true for low inhibitory control students who may already be at a disadvantage. Thus, the data imply that teachers could enhance learning by invoking by creative solutions utilizing their awareness of structure-mapping processes and children's EF limitations in using these technological options to make a record of multiple visual representations and draw explicit connections between them.

Limitations and Future Directions

The current study provides important findings on the role of visual representations for challenging a common misconception through structure-mapping in the mathematical area of proportional reasoning. While this is an area that is critical for students' future attainment of algebra (NRC, 2001), a broader variety of mathematical domains need to be tested to examine the universality of these results before making a clear guideline for teachers.

A strength of the studies is that the instructional stimuli derive from videodata of a real classroom lesson, leading to a simulation of an everyday classroom learning experience, with the aim to increase the study's generalizability to teaching practices. While video-lessons are an increasing trend with the heightened use of methodologies such as "flipped classrooms" (Jinlei, Ying, & Baohui, 2012) in higher education, elementary students generally interact with live teachers, instead of recordings of a teacher. Despite this, video can convey emotion, body language and other non-verbal cues, thus offering a more realistic medium than text-based or computerized materials. Further, teacher actions within a video-lesson are more translatable to a true lesson.

Thus, this technology has high potential for maximizing internal and external validity for testing findings evidenced in laboratory contexts and translate them to teaching practices as well as isolating the efficiency of instructional methods that teachers routinely use in their classrooms. At the same time, there are limits to the simulation, so a future direction for this work would be to extend the methodology into testing teacher delivered material. Additional future directions include using the video methodology to test the efficacy of additional aspects of the instructional

routines to provide additional explicit guidelines, including use of teacher gestures or order of presenting contrasting representations.

At present, further studies are being conducted to examine the impact of the following instructional practices: (a) the teacher's gestures when presenting and linking key ideas and (b) the visual organization of solutions on the board. These two practices were observed by Richland, Zur, and Holyoak (2007) to correlate with the practices used in Experiment 1, but they remain to be tested experimentally. Other studies are examining the role of stereotype threat and EF in connection with teaching supports when learning from instructional analogies. These studies attempt to uncover the EF mechanisms underlying the effectiveness of teaching strategies or the hindrances induced by stereotype anxiety. Within this realm, it may be possible that presenting solutions sequentially is inefficient for students with lower WM & IC resources, but presenting solutions simultaneously may prove favorable for all students. A natural extension of Experiments 1 & 2 is to examine the role of EF when a teacher visualizes all solutions throughout the lesson.

Despite the strong experimental evidence presented in this dissertation and supporting theoretical accounts, the relationships between EF and learning through structure-mapping are not causal. Determining a causal link may be achieved by training specific EF processes and examining whether such training improves learning through structure-mapping. WM & IC training, in general, has seen excitement in recent years for its promises that it may lead to benefits in other thinking tasks or transfer. Yet, its effectiveness remains a hotly debated topic with variable outcomes, with most evidence arising from WM training interventions (see Diamond & Lee, 2011 for exceptions). Some have argued against benefits for WM training (Melby-Lervåg & Hulme, 2012; Shipstead, Redick, & Engle, 2012), but others have found

successful transfer effects (Diamond & Lee, 2011; Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Jaeggi, Buschkuhl, Shah, & Jonides, 2014; Klingberg, 2010; Morrison & Chein, 2011) and even neural changes as a result of training (Buschkuhl, Hernandez-Garcia, Jaeggi, Bernard, & Jonides, 2014). While some of the tested transfer tasks, such as Raven's Progressive Matrices are considered visuospatial analogy tasks (Raven, 1938; Sternberg, 1977), there seems to be only one study that has examined the effects of WM training on semantic analogies (Richey, Phillips, Schunn, & Schneider, 2014).

Richey et al., (2014) did not find an effect on analogy performance after successful WM training. Interestingly, WM training transferred to non-trained WM tasks, and WM performance was significantly correlated with analogy performance. The authors provide three possible explanations. First, the training effects were not large enough to transfer to other tasks. Second, WM may be more strongly related to acquiring background knowledge, which is a significant predictor of analogical reasoning (Novick & Holyoak, 1991), but WM may not be a "critical bottleneck" in analogical performance (Richey et al., 2014). The exploratory results suggest that IC processes may play a significant role as well, but data on IC training is scant. Thirdly, it may take time for people to adapt to higher WM resources and use strategies associated with improved WM. The authors seemed to favor the latter two explanations, yet an explanation not considered by the authors is that not all types of WM training lead to transfer (Jaeggi & Buschkuhl, 2014) and training varies between individuals (Jaeggi et al., 2014). A potential way to account for background knowledge may be to investigate analogical processes as they operate when learning a novel concept, while accounting for previous knowledge with a pretest. This may also be more relevant than performance on an abstract analogy task. The methodology presented in this dissertation may provide the ideal conditions to examining whether the link

between WM training and learning is causal and whether it generalizes to educational settings. Critically, within the learning context, WM may play a different role than during performance, since reasoners need to integrate sequences of events to form a complex schema. Recent data suggest that a combination of WM & reasoning training can be successfully integrated in school curricula and improve reasoning on educationally relevant tests (Ariès, Groot, & van den Brink, 2014). In addition, the exploratory results from Experiment 2 reflect that students WM scores that were in the top 25% and middle 50% had strikingly similar mathematics scores, suggesting students in the bottom 25% account for most of the difference in the regression analysis. This also suggests, that a minor improvement in WM score may result in significant gains in mathematics. Further work is needed to investigate whether WM training leads to gains in learning mathematics through structure-mapping and may also provide another data point in understanding whether the link between EF and analogy is causal.

General Summary

In summary, these findings suggest that instructional recommendations should emphasize the utility of making compared representations visible simultaneously, but more broadly to highlight the importance of supporting learners in aligning, mapping, and drawing inferences about the similarities and differences across representations such as multiple solution strategies for a problem. Teachers should also be made aware of the challenges inherent in making such comparisons when one of the representations is a misconception. In such cases, students may need additional support to control their attentional responses to the misconception in order to engage in more productive knowledge re-representation and new schema formation. In the context of instructional analogies, it is important to consider that visual representations should

highlight relationships between representations, not just increase salience, memorability, and clarity of one representation. The latter has the potential to support deeper encoding of a misconception, rather than desired schema formation that leads to generalizable learning.

Differences between students' cognitive resources point towards working memory and inhibitory control as mechanisms not only contributing to performance, but also learning from instructional analogy. This could be the first study to experimentally link learning of an everyday in classroom task with EF. This provides promising avenues for advancing our understanding of learning in everyday settings and optimizing learning conditions for all students. Instructional attention should be paid to carefully considering the role of instructional supports to account for student learning differences and balancing the benefits for improved encoding of relational structure with ensuring that all students align, map, and compare these structured representations to ensure broader generalization, misconception revision, and appropriate schema formation.

References

- Abedi, J. (2008). Classification System for English Language Learners : Issues and Recommendations.
- Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: a meta-analytic review. *Educational Psychologist, 48*(2), 87–113.
- Alibali, M. W., Flevares, L. M., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology, 89*(1), 183–193. <http://doi.org/10.1037/0022-0663.89.1.183>
- Alloway, T. P. (2007). *Automated Working: Memory Assessment: Manual*. Pearson.
- Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20–29. <http://doi.org/10.1016/j.jecp.2009.11.003>
- Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Development, 80*(2), 606–621.
- Ariës, R. J. F. J., Groot, W., & van den Brink, H. M. (2014). Improving reasoning skills in secondary history education by working memory training. *British Educational Research Journal, 41*(2), 210–228. <http://doi.org/10.1002/berj.3142>
- Atchley, P., & Kramer, A. F. (2001). Object and space-based attentional selection in three-dimensional space. *Visual Cognition*. <http://doi.org/10.1080/13506280042000009>
- Balslev, T., De Grave, W. S., Muijtjens, A. M. M., & Scherpbier, A. J. J. A. (2005). Comparison of text and video cases in a postgraduate problem-based learning format. *Medical Education, 39*(11), 1086–1092. <http://doi.org/10.1111/j.1365-2929.2005.02314.x>

- Barnier, a. J. (2012). Memory, ecological validity and a barking dog. *Memory Studies*, 5(4), 351–359. <http://doi.org/10.1177/1750698012461243>
- Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, 22(4), 577–609; discussion 610–660. <http://doi.org/10.1017/S0140525X99252144>
- Begolli, K. N., & Richland, L. E. (2013). Visual Support for Instructional Analogy : Context Matters. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1863–1868). Austin, TX: Cognitive Science Society.
- Begolli, K. N., & Richland, L. E. (2015). Teaching Mathematics by Comparison : Analog Visibility as a Double-Edged Sword. *Journal of Educational Psychology*, 107(3). <http://doi.org/http://dx.doi.org/10.1037/edu0000056>
- Begolli, K. N., Richland, L. E., & Jaeggi, S. M. (2015). The Role of Executive Functions for Structure-Mapping in Mathematics. In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warfaumont, & J. Yoshimi (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (Vol. 1, pp. 1–6). Austin, TX: Cognitive Science Society.
- Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: mechanisms, alleviation, and spillover. *Journal of Experimental Psychology. General*, 136(2), 256–276. <http://doi.org/10.1037/0096-3445.136.2.256>
- Bender, A., & Beller, S. (2013). Cognition is ... Fundamentally Cultural. *Behavioral Sciences*, 3(1), 42–54. <http://doi.org/10.3390/bs3010042>
- Berry, G., & Klein, R. (1993). Does motion-induced grouping modulate the flanker compatibility effect?: A failure to replicate Driver & Baylis. *Canadian Journal of Experimental*

Psychology = Revue Canadienne de Psychologie Experimentale, 47(4), 714–729.

<http://doi.org/10.1037/h0078867>

- Bindemann, M., Mike Burton, A., & Langton, S. R. H. (2008). How do eye gaze and facial expression interact? *Visual Cognition*, 16(6), 708–733.
- Bissett, P. G., & Logan, G. D. (2012). Post-stop-signal slowing: Strategies dominate reflexes and implicit learning. *Journal of Experimental Psychology: Human Perception and Performance*, 38(3), 746–757. <http://doi.org/10.1037/a0025429>
- Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In *Practical Aspects of Memory: Current Research and Issues, Vol. 1: Memory in everyday life* (Vol. 1, pp. 396–401). Retrieved from <http://psycnet.apa.org/psycinfo/1988-97682-062>
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings.
- Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, 25, 24–34.
<http://doi.org/10.1016/j.learninstruc.2012.11.002>
- Brown, A. L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *Journal of the Learning Sciences*, 2(2), 141–178. http://doi.org/10.1207/s15327809jls0202_2
- Brown, D. E. (2013). Students' Conceptions as Dynamically Emergent Structures. *Science & Education*, 23, 1463–1483. <http://doi.org/10.1007/s11191-013-9655-9>
- Brown, D. E., & Clement, J. (1989). Overcoming misconceptions via analogical reasoning: abstract transfer versus explanatory model construction. *Instructional Science*, 18, 237–261.

- Bunce, D. M., Flens, E. A., & Neiles, K. Y. (2010). How long can students pay attention in class? A study of student attention decline using clickers. *Journal of Chemical Education*, 87(12), 1438–1443. <http://doi.org/10.1021/ed100409p>
- Buschkuohl, M., Hernandez-Garcia, L., Jaeggi, S. M., Bernard, J. a, & Jonides, J. (2014). Neural effects of short-term training on working memory. *Cognitive, Affective & Behavioral Neuroscience*, 14(1), 147–60. <http://doi.org/10.3758/s13415-013-0244-9>
- Carey, S., & Spelke, E. S. (1994). Domain-specific knowledge and conceptual change. *Mapping the Mind: Domain Specificity in Cognition and Culture*, 169–200.
- Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using Spacing to Enhance Diverse Forms of Learning: Review of Recent Research and Implications for Instruction. *Educational Psychology Review*, 24(3), 369–378. <http://doi.org/10.1007/s10648-012-9205-z>
- Carraher, T. N., Schliemann, A. D., & Carraher, D. W. (1988). Mathematical concepts in everyday life. *New Directions for Child and Adolescent Development*, 1988(41), 71–87.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20(6), 633–642. <http://doi.org/10.3758/BF03202713>
- Chi, M. T. H. (2013). Two kinds and four sub-types of misconceived knowledge, ways to change it and the learning outcomes. *International Handbook of Research on Conceptual Change*, 49–70.
- Chinn, C. A., & Brewer, W. F. (1993). The Role of Anomalous Data in Knowledge Acquisition: A Theoretical Framework and Implications for Science Instruction. *Review of Educational Research*, 63(1), 1–49. <http://doi.org/10.3102/00346543063001001>

- Cho, S., Holyoak, K. J., & Cannon, T. D. (2007). Analogical reasoning in working memory: resources shared among relational integration, interference resolution, and maintenance. *Memory & Cognition*, 35(6), 1445–1455. <http://doi.org/10.3758/BF03193614>
- Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, 50(1), 66–71.
- Clement, J., & Brown, D. E. (2008). Using Analogies and Models in Instruction to Deal with Students' Preconceptions. *Creative Model Construction in Scientists and Students*, (1989), 139–155. http://doi.org/10.1007/978-1-4020-6712-9_10
- Collins, A. (1999). The changing infrastructure of education research. *Issues in Education Research*, 198–289.
- Collins, A., Joseph, D., & Bielaczyc, K. (2009). Design Research: Theoretical and Methodological Issues. *Journal of the Learning Sciences*, 13(August 2014), 15–42. <http://doi.org/10.1207/s15327809jls1301>
- Common Core State Standards Initiative. (2010). Common Core State Standards for. *Development*, 31, 93. Retrieved from <http://www.corestandards.org/>
- Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development*, 84(0), 1863–1871. <http://doi.org/10.1111/cdev.12097>
- Correa-Chavez, M., & Rogoff, B. (2005). Cultural research has transformed our ideas of cognitive development.pdf. *International Journal of Behavioral Development*.
- D'Eath, R. B. (1998). Can video images imitate real stimuli in animal behaviour experiments? *Biological Reviews*, 73(3), 267–292. <http://doi.org/10.1111/j.1469-185X.1998.tb00031.x>

- Daniel, D. B. (2012). Promising principles: Translating the science of learning to educational practice. *Journal of Applied Research in Memory and Cognition*, *1*(4), 251–253.
<http://doi.org/10.1016/j.jarmac.2012.10.004>
- De Vita, M., Verschaffel, L., & Elen, J. (2014). Interactive whiteboards in mathematics teaching: A literature review. *Education Research International Volume 2014 (2014)*, Article, 2014(Article ID 401315), 1–16.
- DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of Experimental Child Psychology*, *113*(4), 552–568.
<http://doi.org/10.1016/j.jecp.2012.06.009>
- Derry, S. J. (2007). *Guidelines for Video Research in Education: Recommendations from an Expert Panel. Report submitted to the National Science Foundation*. Retrieved from <http://drdc.uchicago.edu/what/video-research-guidelines.pdf>
- Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., ... Sherin, B. L. (2010). Conducting Video Research in the Learning Sciences: Guidance on Selection, Analysis, Technology, and Ethics. *Journal of the Learning Sciences*.
<http://doi.org/10.1080/10508400903452884>
- Design-Based Research Collective. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher*, *32*(1), 5–8.
<http://doi.org/10.3102/0013189X032001005>
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.
<http://doi.org/10.1146/annurev-psych-113011-143750>

- Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science (New York, N.Y.)*, 333(6045), 959–964.
<http://doi.org/10.1126/science.1204529>
- Dietrich, E. (2000). Analogy and Conceptual Change, or You Can't Step Into the Same Mind Twice. *Cognitive Dynamics : Conceptual and Representational Change in Humans and Machines*, 265–294.
- Donovan, M. S. (2013). in *Education Systems*, 317(April), 317–320.
<http://doi.org/10.1126/science.1236180>
- Dreyfus, A., Jungwirth, E., & Eliovitch, R. (1990). Applying the “cognitive conflict” strategy for conceptual change—some implications, difficulties, and problems. *Science Education*, 74(5), 555–569.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Brooks-Gunn, J. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6), 1428.
- Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*. [http://doi.org/10.1016/S0166-2236\(00\)01633-7](http://doi.org/10.1016/S0166-2236(00)01633-7)
- Dunlosky, J., & Rawson, K. a. (2012). Despite their promise, there's still a lot to learn about techniques that support durable learning. *Journal of Applied Research in Memory and Cognition*, 1(4), 254–256. <http://doi.org/10.1016/j.jarmac.2012.10.003>
- Dunlosky, J., Rawson, K. a., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From

- Cognitive and Educational Psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <http://doi.org/10.1177/1529100612453266>
- English, L. D., & Halford, G. S. (1995). Mathematics education. *Mahwah, NJ: LEA*.
- Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, 51(1), 136–158. <http://doi.org/10.1016/j.jml.2004.03.008>
- Fyfe, E. R., Decaro, M. S., & Rittle-Johnson, B. (2014). An alternative time for telling: When conceptual instruction prior to problem solving improves mathematical knowledge. *British Journal of Educational Psychology*, 1–18. <http://doi.org/10.1111/bjep.12035>
- Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The Effects of Feedback During Exploratory Mathematics Problem Solving: Prior Knowledge Matters. *Journal of Educational Psychology*. <http://doi.org/10.1037/a0028389>
- Gaba, D. M. (2004). The future vision of simulation in health care. *Quality & Safety in Health Care*, 13 Suppl 1, i2–i10. <http://doi.org/10.1136/qshc.2004.009878>
- Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A.-M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology*, 93(3), 265–281. <http://doi.org/10.1016/j.jecp.2005.08.003>
- Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, 18(October 2003), 1–16. <http://doi.org/10.1002/acp.934>
- Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy*. *Cognitive Science*, 7(2), 155–170. [http://doi.org/10.1016/S0364-0213\(83\)80009-3](http://doi.org/10.1016/S0364-0213(83)80009-3)

- Gentner, D., & Forbus, K. D. (2011). Computational models of analogy. *Wiley Interdisciplinary Reviews: Cognitive Science*. <http://doi.org/10.1002/wcs.105>
- Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*. <http://doi.org/10.1037/0022-0663.95.2.393>
- Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In *Perspectives on thought and language: Interrelations in development* (pp. 225–277). Retrieved from <http://groups.psych.northwestern.edu/gentner/papers/GentnerRattermann98.pdf>
- Gibson, J. J. (1947). *Motion picture testing and research*. Washington, DC: US Government Printing Office. Retrieved from <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=AD0651783> http://scholar.google.ch/scholar?q=motion+picture+testing&hl=en&btnG=Search&as_sdt=1,5&as_sdt=on#4
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Book (Vol. 40). <http://doi.org/10.2307/989638>
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*. [http://doi.org/10.1016/0010-0285\(80\)90013-4](http://doi.org/10.1016/0010-0285(80)90013-4)
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1–38. [http://doi.org/10.1016/0010-0285\(83\)90002-6](http://doi.org/10.1016/0010-0285(83)90002-6)
- Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction*, 39, 72–87. <http://doi.org/10.1016/j.learninstruc.2015.05.001>

- Goldman, D. R. (2007). A Framework for Video Annotation , Visualization , and Interaction. *Framework*.
- Goswami, U. (2001). Analogical reasoning in children. *The Analogical Mind: Perspectives from Cognitive Science*, 437–470.
- Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist*, 53(1), 5–26. <http://doi.org/10.1037/0003-066X.53.1.5>
- Halford, G. S., Andrews, G., Dalton, C., Boag, C., & Zielinski, T. (2002). Young children’s performance on the balance scale: the influence of relational complexity. *Journal of Experimental Child Psychology*, 81(4), 417–445. <http://doi.org/10.1006/jecp.2002.2665>
- Hart, K. M. (1984). *Ratio: Children’s strategies and errors*. NFER-Nelson Windsor, England.
- Heath, C., Luff, P., & Sanchez Svensson, M. (2007). Video and qualitative research: Analysing medical practice and interaction. *Medical Education*, 41(1), 109–116. <http://doi.org/10.1111/j.1365-2929.2006.02641.x>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, 33(2-3), 61–83; discussion 83–135. <http://doi.org/10.1017/S0140525X0999152X>
- Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M., ... Gallimore, R. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis*, 27(2), 111–132.
- Hill, H. C., Ball, D. L., Blunk, M. L., Goffney, I. M., & Rowan, B. (2007). Validating the Ecological Assumption: The Relationship of Measure Scores to Classroom Teaching and Student Learning. *Measurement: Interdisciplinary Research & Perspective*. <http://doi.org/10.1080/15366360701487138>

- Houdé, O., Pineau, A., Leroux, G., Poirel, N., Perchey, G., Lanoë, C., ... Simon, G. (2011). Functional magnetic resonance imaging study of Piaget's conservation-of-number task in preschool and school-age children: a neo-Piagetian approach. *Journal of Experimental Child Psychology*, *110*(3), 332–346.
- Hunting, R. P. (1983). Emerging methodologies for understanding internal processes governing children's mathematical behaviour. *Australian Journal of Education*.
- Jaeggi, S. M., & Buschkuhl, M. (2014). New Frontiers of Multidisciplinary Research in STEAM-H (Science, Technology, Engineering, Agriculture, Mathematics, and Health), *90*, 19–43. <http://doi.org/10.1007/978-3-319-07755-0>
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–6833. <http://doi.org/10.1073/pnas.0801268105>
- Jaeggi, S. M., Buschkuhl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, *42*(3), 464–80. <http://doi.org/10.3758/s13421-013-0364-z>
- Jinlei, Z., Ying, W., & Baohui, Z. (2012). Introducing a New Teaching Model: Flipped Classroom. *Journal of Distance Education*. Retrieved from http://en.cnki.com.cn/Article_en/CJFDTOTAL-YCJY201204007.htm
- Jonassen, D. H. (2008). *Meaningful learning with technology*. Prentice Hall.
- Kapon, S., & diSessa*, A. a. (2012). Reasoning Through Instructional Analogies. *Cognition and Instruction*, *30*(May 2014), 261–310. <http://doi.org/10.1080/07370008.2012.689385>
- Kaput, J., & West, M. M. (1994). Missing value proportional reasoning problems: Factors affecting informal reasoning patterns.

- Karplus, R., Pulos, S., & Stage, E. K. (1983). Proportional reasoning of early adolescents. *Acquisition of Mathematics Concepts and Processes*, 45–90.
- Kersting, N. B. (2008). Using Video Clips of Mathematics Classroom Instruction as Item Prompts to Measure Teachers' Knowledge of Teaching Mathematics. *Educational and Psychological Measurement*, 68(5), 845–861. <http://doi.org/10.1177/0013164407313369>
- Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive Ethology: a new approach for studying human cognition. *British Journal of Psychology (London, England : 1953)*, 99(Pt 3), 317–340. <http://doi.org/10.1348/000712607X251243>
- Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., & Eastwood, J. D. (2003). 1 ATTENTION RESEARCHERS! IT IS TIME TO TAKE A LOOK AT THE REAL WORLD Alan Kingstone 1 , Daniel Smilek, Jelena Ristic, 12(5), 1–17.
- Klahr, D. (2013). What do we mean? On the importance of not abandoning scientific rigor when talking about science education. *Proceedings of the National Academy of Sciences of the United States of America*, 110 Suppl (2003), 14075–80. <http://doi.org/10.1073/pnas.1212738110>
- Klahr, D., & Li, J. (2005). Cognitive research and elementary science instruction: From the laboratory, to the classroom, and back. *Journal of Science Education and Technology*, 14(2), 217–238. <http://doi.org/10.1007/s10956-005-4423-5>
- Klein, R. a., Ratliff, K. a., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., ... Nosek, B. a. (2014). Investigating variation in replicability: A “many labs” replication project. *Social Psychology*, 45(3), 142–152. <http://doi.org/10.1027/1864-9335/a000178>
- Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, 14(7), 317–324. <http://doi.org/10.1016/j.tics.2010.05.002>

- Koedinger, K. R., Alevan, V., Roll, I., & Baker, R. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. *Handbook of Metacognition in Education*, 42(1983), 647–51. <http://doi.org/10.1002/lsm.20954>
- Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional Complexity and the Science to Constrain It. *Science*, 342(22 november 2013), 935–937.
<http://doi.org/10.1126/science.1238056>
- Kornell, N., Rabelo, V. C., & Klein, P. J. (2012). Tests enhance learning-Compared to what? *Journal of Applied Research in Memory and Cognition*, 1(4), 257–259.
<http://doi.org/10.1016/j.jarmac.2012.10.002>
- Krawczyk, D. C., Hanten, G., Wilde, E. A., Li, X., Schnelle, K. P., Merkley, T. L., ... Chapman, S. B. (2010). Deficits in analogical reasoning in adolescents with traumatic brain injury. *Frontiers in Human Neuroscience*, 4.
- Krawczyk, D. C., Morrison, R. G., Viskontas, I. V, 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, 2020–2032.
<http://doi.org/10.1016/j.neuropsychologia.2008.02.001>
- Kvavilashvili, L., & Ellis, J. a. (2004). Ecological validity and the real-life/laboratory controversy in memory research: a critical and historical review. Retrieved from <http://centaur.reading.ac.uk/13968/>
- Lamon, S. J. (1993a). Ratio and proportion: Children’s cognitive and metacognitive processes. *Rational Numbers: An Integration of Research*, 131–156.
- Lamon, S. J. (1993b). Ratio and proportion: Connecting content and children’s thinking. *Journal for Research in Mathematics Education*, 41–61.

- Lauth, G. W., Heubeck, B. G., & Mackowiak, K. (2006). Observation of children with attention-deficit hyperactivity (ADHD) problems in three natural classroom contexts. *The British Journal of Educational Psychology*, 76(Pt 2), 385–404.
<http://doi.org/10.1348/000709905X43797>
- Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge University Press.
- Lo, J.-J., & Watanabe, T. (1997). Developing ratio and proportion schemes: A story of a fifth grader. *Journal for Research in Mathematics Education*, 216–236.
- Lyden, P., Brott, T., Tilley, B., Welch, K. M., Mascha, E. J., Levine, S. C., ... Marler, J. (1994). *Improved reliability of the NIH Stroke Scale using video training. NINDS TPA Stroke Study Group. Stroke; a journal of cerebral circulation* (Vol. 25).
- Marcus, A., & Wilder, D. A. (2009). A comparison of peer video modeling and self video modeling to teach textual responses in children with autism. *Journal of Applied Behavior Analysis*, 42(2), 335–341. <http://doi.org/10.1901/jaba.2009.42-335>
- Mayer, R. E. (2008). Applying the science of learning: evidence-based principles for the design of multimedia instruction. *The American Psychologist*, 63(November), 760–769.
<http://doi.org/10.1037/0003-066X.63.8.760>
- Mayer, R. E. (2012). Advances in applying the science of learning to education: An historical perspective. *Journal of Applied Research in Memory and Cognition*, 1(4), 249–250.
<http://doi.org/10.1016/j.jarmac.2012.10.001>
- McCandliss, B. D., Kalchman, M., & Bryant, P. (2003). Design Experiments and Laboratory Approaches to Learning: Steps Toward Collaborative Exchange. *Educational Researcher*, 32(1), 14–16. <http://doi.org/10.3102/0013189X032001014>

- McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, *14*(2), 200–206. <http://doi.org/10.3758/BF03194052>
- Melby-Lervåg, M., & Hulme, C. (2012). Is Working Memory Training Effective? A Meta-Analytic Review. *Developmental Psychology*, *49*(2), 270–291. <http://doi.org/10.1037/a0028228>
- Metcalfe, J., Kornell, N., & Son, L. K. (2007). A cognitive-science based programme to enhance study efficacy in a high and low risk setting. *European Journal of Cognitive Psychology*, *19*, 743–768. <http://doi.org/10.1080/09541440701326063>
- Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. <http://doi.org/10.1177/0963721411429458>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, a H., Howerter, a, & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cognitive Psychology*, *41*, 49–100. <http://doi.org/10.1006/cogp.1999.0734>
- Monin, B., Oppenheimer, D. M., Ferguson, M. J., Carter, T. J., Hassin, R. R., Crisp, R. J., ... Kahneman, D. (2014). Commentaries and Rejoinder on Klein et al. (2014). *Social Psychology*, *1*(-1), 1–13. <http://doi.org/10.1027/1864-9335/a000202>
- Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, *18*(1), 46–60. <http://doi.org/10.3758/s13423-010-0034-0>

- Morrison, R. G., Dumas, L. A. A. L. a a, & Richland, L. E. (2006). The Development of Analogical Reasoning in Children : A Computational Account Multiple Factors in Analogical Development. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 603–608.
- Morrison, R. G., Dumas, L. A. A. L. a a, & Richland, L. E. (2011). A computational account of children’s analogical reasoning: Balancing inhibitory control in working memory and relational representation. *Developmental Science*, 14(3), 516–529.
<http://doi.org/10.1111/j.1467-7687.2010.00999.x>
- Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., & Knowlton, B. J. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *Journal of Cognitive Neuroscience*, 16(2), 260–271. <http://doi.org/10.1162/089892904322984553>
- National Mathematics Advisory Panel. (2008). The Final Report of the National Mathematics Advisory Panel. *Foundations*, 37(9), 595–601. Retrieved from <http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf>
- National Research Council. (2001). *Adding It Up: Helping Children Learn Mathematics*. The National Academies Press. Retrieved from <http://www.nap.edu/catalog/9822/adding-it-up-helping-children-learn-mathematics>
- Nosek, B. a. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, 7(6), 657–660.
<http://doi.org/10.1177/1745691612462588>

- Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 17(3), 398–415.
<http://doi.org/10.1037/0278-7393.17.3.398>
- Nunnally, J. C. (1967). *Psychometric Methods* (McGraw-HillBook Company, New York).
- Oliveira, R. F., Rosenthal, G. G., Schlupp, I., McGregor, P. K., Cuthill, I. C., Endler, J. A., ...
Waas, J. R. (2000). Considerations on the use of video playbacks as visual stimuli: the
Lisbon workshop consensus. *Acta Ethologica*. <http://doi.org/10.1007/s102110000019>
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design :
Recent Developments Cognitive Load Theory and Instructional Design : Recent
Developments. *Educational Psychologist*, 38(1), 1–4.
<http://doi.org/10.1207/S15326985EP3801>
- Pardos, Z. a., Baker, R., San Pedro, M. O. C. Z., Gowda, S. M., & Gowda, S. M. (2013).
Affective states and state tests. *Proceedings of the Third International Conference on
Learning Analytics and Knowledge - LAK '13*, 117.
<http://doi.org/10.1145/2460296.2460320>
- Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A. C., Koedinger, K. R., McDaniel, M. A., &
Metcalf, J. (2007). *Organizing Instruction and Study to Improve Student Learning. IES
Practice Guide. NCER 2007-2004. National Center for Education Research*.
- Peugh, J. L., & Enders, C. K. (2004). Missing Data in Educational Research: A Review of
Reporting Practices and Suggestions for Improvement. *Review of Educational Research*.
<http://doi.org/10.3102/00346543074004525>

- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227.
- Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence*. London: HK Lewis.
- Reed, E. . (1996). *Encountering the world: toward an ecological psychology*. *Choice Reviews Online* (Vol. 34). <http://doi.org/10.5860/CHOICE.34-5983>
- Rich, P. J., & Hannafin, M. J. (2008). Decisions and reasons: Examining preservice teacher decision-making through video self-analysis. *Journal of Computing in Higher Education*, 20(1), 62–94. <http://doi.org/10.1007/BF03033432>
- Richey, J. E., Phillips, J. S., Schunn, C. D., & Schneider, W. (2014). Is the Link from Working Memory to Analogy Causal? No Analogy Improvements following Working Memory Training Gains. *PLoS ONE*, 9(9), e106616. <http://doi.org/10.1371/journal.pone.0106616>
- Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development. *Psychological Science*, 24(1), 87–92.
- Richland, L. E., & Hansen, J. (2013). Reducing Cognitive Load in Learning by Analogy. *International Journal of Psychological Studies*, 5(4), 69–80. <http://doi.org/10.5539/ijps.v5n4p69>
- Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied; Journal of Experimental Psychology: Applied*, 15(3), 243.
- Richland, L. E., Linn, M. C., & Bjork, R. A. (2007). Instruction. *Handbook of Applied Cognition*, 553–583.

- Richland, L. E., & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology*, 35(1), 28–43.
<http://doi.org/10.1016/j.cedpsych.2009.09.001>
- Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2004). Working Memory and Inhibition as Constraints on Children's Development of Analogical Reasoning. *Current*, 14, 1149–54.
Retrieved from <http://www.cogsci.northwestern.edu/cogsci2004/papers/paper471.pdf>
- Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, 94(3), 249–273.
- Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the Conceptual Structure of Mathematics. *Educational Psychologist*, 47(3), 189–203.
<http://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-NEW YORK THEN WASHINGTON-*, 316(5828), 1128.
- Risko, E. F., Buchanan, D., Medimorec, S., & Kingstone, A. (2013). Everyday attention: Mind wandering and computer use during lectures. *Computers and Education*, 68, 275–283.
<http://doi.org/10.1016/j.compedu.2013.05.001>
- Rittle-Johnson, B., & Schneider, M. (2013). Developing Conceptual and Procedural Knowledge of Mathematics. *Oxford Handbook of Numerical Cognition*.
<http://doi.org/10.1093/oxfordhb/9780199642342.013.014>
- Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*. <http://doi.org/10.1037/0022-0663.93.2.346>

- Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561–574. <http://doi.org/10.1037/0022-0663.99.3.561>
- Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*(3), 529–544. <http://doi.org/10.1037/a0014224>
- Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning. *Psychological Science, 17*(3), 249–255. <http://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Roediger, H. L., & Karpicke, J. D. (2006b). The Power of Testing Memory Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science, 1*(3), 181–210. <http://doi.org/10.1111/j.1745-6916.2006.00012.x>
- Roediger, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition, 1*(4), 242–248. <http://doi.org/10.1016/j.jarmac.2012.09.002>
- Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <http://doi.org/10.1037/0278-7393.15.3.456>
- Rotello, C. M., Heit, E., & Dubé, C. (2014). When more data steer us wrong: replications with the wrong dependent measure perpetuate erroneous conclusions. *Psychonomic Bulletin & Review, 944–954*. <http://doi.org/10.3758/s13423-014-0759-2>
- Santagata, R., Gallimore, R., & Stigler, J. W. (2005). The use of video for teacher education and professional development. Past experiences and future directions. *Current Perspectives on*

- Applied Information Technologies (Volume 2): Preparing Teachers to Teach with Technology.*, 151–167.
- Saxe, G. B. (1988). Candy Selling and Math Learning. *Educational Researcher*.
<http://doi.org/10.3102/0013189X017006014>
- Schmidt, R. a., & Bjork, R. a. (1992). New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training. *Psychological Science*, 3(4), 207–217. <http://doi.org/10.1111/j.1467-9280.1992.tb00029.x>
- Schwartz, D. L., & Bransford, J. D. (1998). A Time For Telling. *Cognition and Instruction*, 16(4), 475–5223. http://doi.org/10.1207/s1532690xci1604_4
- Schwartz, D. L., & Bransford, J. D. (2005). Efficiency and Innovation in Transfer. *Transfer of Learning from a Modern Multidisciplinary Perspective*, 1–51. <http://doi.org/10.1111/j.1365-2133.2005.06492.x>
- Schwartz, D. L., Chase, C. C., Opezzo, M. a., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103(4), 759–775. <http://doi.org/10.1037/a0025140>
- Shimizu, Y. (2003). Problem solving as a vehicle for teaching mathematics: A Japanese perspective. *Teaching Mathematics through Problem Solving: Prekindergarten-Grade, 6*, 205–214.
- Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, 138(4), 628–654. <http://doi.org/10.1037/a0027473>
- Shweder, R. A. (2012). Anthropology's disenchantment with the cognitive revolution. *Topics in Cognitive Science*, 4(3), 354–361. <http://doi.org/10.1111/j.1756-8765.2012.01199.x>

- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, 28(9), 1059–1074. <http://doi.org/10.1068/p2952>
- Smith III, J. P., diSessa*, A. a., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, 3(2), 115–163.
- Soble, J. R., Spanierman, L. B., & Liao, H.-Y. (2011). Effects of a brief video intervention on White university students' racial attitudes. *Journal of Counseling Psychology*, 58(1), 151–157. <http://doi.org/10.1037/a0021158>
- Spiro, R. J., Feltovich, P. J., Coulson, R. L., & Anderson, D. K. (1989). Similarity and analogical reasoning.
- St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology (2006)*, 59(4), 745–759.
<http://doi.org/10.1080/17470210500162854>
- Stein, M. K., Engle, R. a., Smith, M. S., & Hughes, E. K. (2008). Orchestrating Productive Mathematical Discussions: Five Practices for Helping Teachers Move Beyond Show and Tell. *Mathematical Thinking and Learning*, 10(793217298), 313–340.
<http://doi.org/10.1080/10986060802229675>
- Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review*, 84(4), 353–378. <http://doi.org/10.1037/0033-295X.84.4.353>
- Stigler, J. W., & Hiebert, J. (1999). The teaching gap.
- Suchman, L. A. (1987). *Plans and Situated Actions*. Cambridge University Press. Retrieved from http://books.google.com/books?id=AJ_eBJtHxmsC

- Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, 55(December), 661–699.
<http://doi.org/10.1111/j.0023-8333.2005.00320.x>
- Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America*, 110(16), 6313–7.
<http://doi.org/10.1073/pnas.1221764110>
- Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers’ gestures facilitate students’ learning: A lesson in symmetry. *Contemporary Educational Psychology*, 28, 187–204.
[http://doi.org/10.1016/S0361-476X\(02\)00007-3](http://doi.org/10.1016/S0361-476X(02)00007-3)
- Vamvakoussi, X., & Vosniadou, S. (2012). Bridging the Gap Between the Dense and the Discrete: The Number Line and the “Rubber Line” Bridging Analogy. *Mathematical Thinking and Learning*, 14(May 2014), 265–284.
<http://doi.org/10.1080/10986065.2012.717378>
- Van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers’ “learning to notice” in the context of a video club. *Teaching and Teacher Education*, 24(2), 244–276.
<http://doi.org/10.1016/j.tate.2006.11.005>
- 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), 581–591. <http://doi.org/10.1037/0882-7974.19.4.581>
- Vosniadou, S. (2013). Conceptual change in learning and instruction: The framework theory approach. *International Handbook of Research on Conceptual Change*, 2, 11–30.

- Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. *International Handbook of Research on Conceptual Change*, 3–34.
- Waltz, J. A., Lad, A., Grewal, S. K., & Holyoak, K. J. (2000). of working, 28(7), 1205–1212.
- Webb, A., Knott, A., & MacAskill, M. R. (2010). Eye movements during transitive action observation have sequential structure. *Acta Psychologica*, 133(1), 51–56.
<http://doi.org/10.1016/j.actpsy.2009.09.001>
- Wilson, K., & Korn, J. H. (2007). Attention During Lectures: Beyond Ten Minutes. *Teaching of Psychology*, 34(2), 85–89. <http://doi.org/10.1080/00986280701291291>
- Wright, A., & Diamond, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Frontiers in Psychology*, 5(March), 1–9.
<http://doi.org/10.3389/fpsyg.2014.00213>
- Zelazo, P. D., Muller, U., Frye, D., & Marcovitch, S. (2003). VI. The development of executive function: Cognitive complexity and control-revised. *Monographs of the Society for Research in Child Development*, 68(3), 93–119. <http://doi.org/10.1111/j.0037-976X.2003.00266.x>
- Zook, K. B. (1991). Effects of analogical processes on learning and misrepresentation. *Educational Psychology Review*, 3(1), 41–72. <http://doi.org/10.1007/BF01323662>
- Zook, K. B., & Di Vesta, F. J. (1991). Instructional analogies and conceptual misrepresentations. *Journal of Educational Psychology*, 83(2), 246.

APPENDIX A

Let's go back to our original problem and pull this altogether. I'd like us to think about how these strategies are different and how they are similar. This is really important part of what we are doing today.

Let's start with reviewing why Ryan's strategy is not the best way to solve this problem.

Remember in Ryan's strategy, he tried subtracting "total shots tried" from "shots made" and tried to compare the missed shots (point to the subtraction results of 8 and 9) to figure out who was the better free-throw shooter. But we found out that this strategy does not work. Remember when Ken made 0 shots in the counterexample (point to the counterexample), but still missed less shots? From this example, we learned that we cannot subtract shots tried from shots made (point to the subtraction results of 8 and 9) and then compare the shots missed.

Subtraction is not the right way to solve this problem.

Now, **how is this different from Carina's strategy?**

To find out who is better she first setup Ken and Yoko's shots made and shots tried as a fraction.

Without looking at the numbers (cover the numbers with your palm), how is comparing the fractions of shots made and shots tried for Ken and Yoko, in Carina's strategy (point to the shots made and shots tried ratio) fundamentally different from comparing shots missed only in Ryan's way (point to the subtraction results of 8 and 9)?

Brief Pause**** (perhaps the video should stop here and the students should be prompted to answer this question on their computer) Right after they answer the question the video follows.

Well, Ryan only compared one unit, shots missed (point to shots missed), whereas Carina compared two units (shots made and shots tried).

Why did she do that? Because they shot a different amount. So, if we want to know who is better at shooting free throws when they do not shoot the same number of shots, we have to compare the number of shots made and the number of shots tried.

This relationship of comparing *shots made to shots tried* is called a RATIO.

Thus,

WRITE: A relationship between two quantities is a RATIO.

So, after Carina set it up as a ratio, she made the *shots tried* of Ken and Yoko, the 20 and the 25, equal to each other. She did this by finding the LCM of, 20 and 25 , which was 100 (*point to the ratio of 64/100 and 60/100*). and then she multiplied 12 by 5 and 16 by 4 to get 60 and 64 respectively (*point to the part where Carina did the calculations on the board*). Remember, she multiplied 12 by 5 because that's the number of times she had to multiply 20 to make 100, and she multiplied 16 by 4 because that's the number of times it takes 25 to make 100. Therefore, since we found the LCM and now the *shots tried* for both Ken and Yoko are equal (*point to shots tried*), we can compare their *shots made*. (*point to shots made*). So the point here is that we have to make the shots made equal in order to compare who is better.

This was a good strategy, but the problem with this strategy was when we tried to find the LCM for harder numbers like 19 and 25 (*point to these numbers*) we had a hard time.

We found out from Maddie's strategy that we could just divide shots made with shots tried. Let's try to figure out why Maddie's strategy works by comparing it to Carina's. This is the really important part of what we are doing today.

Something that is similar between Carina's and Maddie's which is different from Ryan's strategy is that they both take into account two labels shots tried and shots made. So they use the same units to compare who is better. What else is similar between Carina's and Maddie's strategy

besides the labels? Numbers? Are there numbers in Maddie's strategy that correspond to the number 60 and the number 100 from Carina's strategy? – **point at the original numbers**, 60 and 100, what did we call the relationship between the two quantities in Carina's strategy? A ratio.

Can anyone see a RATIO in Maddie's strategy? Take a look at Ken's numbers in Maddie's strategy, 12 and 20 . Remember that 12 and 20 stand for 12 shots made and 20 shots tried. Carina found that Ken's had 60 shots made and 100 shots tried by doing her LCM calculations with 12 and 20. So, the 12 and 20 in Maddie's is really like the 60 and 100 in Carina's strategy. To prove that these numbers are the same, I'll ask you to divide 60/100.

Students take 1-min to divide.

If we divide the numbers in Ken's RATIO what number would we get?

Ask for a student answer and point to Maddie's results that it's the same number

Point out: Maddie's strategy and Carina's strategy give us the same answer don't they? **Use hand to point to each number on the board showing they're the same.**

Therefore, we can think of the division symbol as a sign for a ratio between the numbers 12 and 20 from Maddie's strategy. A ratio is like a division.

This new RATIO corresponds to Ken's ratio *60 shots made/ 100 shots tried* in Carina's strategy.

We can do the same for Yoko's ratio. **Write** the word ratio next to Carina.

It seems like we went through all the trouble of finding the LCM in Carina's method when we could just divide the numbers and compare those.

The great advantage to Maddie's method is that by dividing the two units she compares both how many times they shot and how many they made, regardless of whether Yoko shot more than Ken.

So, she doesn't have to set the shots made equal to each other. (There has to be a better way to say this - please help - I hope you get the point I am trying to get across).

A ratio or the numbers we got from our division? In our case .60 and .64 represent something special for us when they are next to the ratio of shots made to shots tried.

So, what's different from Carina's method is that now we have a single number (point to the .60 and .64) that represents a ratio of two specific units. In fact, in certain cases when we attach these special units to ratios we call it a RATE. So, a rate is a special case of ratio. It's when both the numerator and denominator of the RATIO have different units assigned to them. when we use ratios this way we call it a RATE.

WRITE: A ratio that has specific units is a RATE

We use these units for measuring things. In our case what are we measuring? Who is better? Could we say we are measuring success?

Another way to think about this success rate is to say Ken makes .60 shots for every shot tried, or Ken makes .60 shots PER shot tried. Just remember that a RATE will always have these units associated with it so that you can think of it as something PER something, while a RATIO doesn't have this requirement. You could compare a ratio of boys to girls in our class, but you wouldn't have a rate of boys to girls.

Where else do we have a single number that represents a ratio of two units? Does 45 Miles per hour sound familiar? In our case we have .60 shots made per shot tried. Just like we go 45 miles per 1 hour, Ken makes .60 for each shot he tries. So, if we he shots 1 time he makes a little more than half of a shot, but if he shoots 10 shots then he makes 6. How many would he make if he shot 100?

So, we can use Ken's rate to figure out how many shots he makes if we know how many shots he tries.

Some of you may be familiar with these terms and some of you may not. In the future we will talk about specific instances of rates, like the ones some of you mentioned earlier – mph – tax rate – etc. and with more practice, it will make more sense when something is a rate and when it is a ratio. What is important now is that you understand what strategy is best when you have to figure out problems like the ones we just solved. Problems that involve comparing **(point to Ken's numbers)** two numbers for each person and then comparing that person to someone else. You should also know that you can use these strategies to figure out who or what goes faster? What is more expensive? Etc.

Now let's all focus on solving a new problem. Now that you have seen all of these strategies I want you to turn the page on your package and solve the following TWO problems.

There are TWO problems and I want you guys to solve both in any way you like. Please do not share with your neighbor. Take as much time as you need. Here too we want to see how you are thinking, so, please do not erase anything, instead use the other side of the paper.

APPENDIX B

Problems Used in the Immediate Posttest

Construct Type	Items	Scoring and α scale																		
<p>Procedural Knowledge</p>		<p>Posttest $\alpha = .89$ Pretest $\alpha = .86$</p>																		
<p>Produce correct Procedures: Familiar</p>	<p>Problem 1. In Cambridge, Sue and Joan played in a free-throw tournament. The results of their shooting are shown in the table below. Who is the better free throw shooter?</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th></th> <th>Shots Made</th> <th>Total Shots Tried</th> </tr> </thead> <tbody> <tr> <td>Sue</td> <td>7</td> <td>11</td> </tr> <tr> <td>Joan</td> <td>11</td> <td>15</td> </tr> </tbody> </table> <p><i>Please show all your work.</i></p> <p>Who is better? _____</p> <p>Problem 2. In Nashville, Miguel and Amos played in a free-throw tournament. The results of their shooting are shown in the table below. This time, please use TWO different ways to find who is the better free throw shooter.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th></th> <th>Shots Made</th> <th>Total Shots Tried</th> </tr> </thead> <tbody> <tr> <td>Miguel</td> <td>9</td> <td>15</td> </tr> <tr> <td>Amos</td> <td>16</td> <td>25</td> </tr> </tbody> </table> <p><i>Please show all your work.</i></p> <p>Part 1) Way 1. Who is better _____? Way 2. Who is better _____?</p>		Shots Made	Total Shots Tried	Sue	7	11	Joan	11	15		Shots Made	Total Shots Tried	Miguel	9	15	Amos	16	25	<p>Setup: 1 point for producing a correct solution strategy (e.g. division between $7 \div 11$ and $11 \div 15$)</p> <p>Contender: 1 point for selecting the correct contender, if students produced a correct strategy. This ensured that students chose the correct contender by using a correct solution method, and not by chance.</p> <p>Setup: 1 point for producing at least one correct solution strategy. OR ½ of a point when producing a correct strategy and using subtraction as an alternative strategy. This score accounted for students who believed subtraction (a common misconception) was one of the correct strategies.</p> <p>Contender: 1 point for selecting at least one correct contender and producing</p>
	Shots Made	Total Shots Tried																		
Sue	7	11																		
Joan	11	15																		
	Shots Made	Total Shots Tried																		
Miguel	9	15																		
Amos	16	25																		

Produce correct
Procedures:
Transfer

Problem 3.

Joe and Kai played video games at boomers and then went to turn in their tickets for prizes. For every game they won, they got 1 ticket. Joe played 27 games and won 11 tickets. Kai played 11 games and won 4 tickets. Who is a better video game player? Please show all your work.

at least one correct strategy.

OR

½ of a point when selecting a correct contender, producing a correct and 1 incorrect strategy (subtraction).

Setup:

1 point for producing the correct strategy.

Contender:







1 point for selecting the correct contender and using a correct strategy.

Produce correct
Procedures: Transfer

Problem 4.

Mr. Perez, Mr. Lopez, and Mr. Smith are giving out cookies to their students. The table below shows the number of cookies to students in each classroom.

- a) Write in the number (ratio) of cookies to students in each classroom.

	Cookies	Students	Number (ratio) of Cookies to Students
Mr. Perez's Classroom			
Mr. Lopez's Classroom			
Mr. Smith's Classroom			

- b) Which two classrooms have the same amount (ratio) of cookies to students?

Omitted from analyses, because the problem was not sensitive to the intervention. Students' scores did not differ significantly between pretest and posttest (average change pre- to posttest of -5%)

a)

1 point when correct for all three teachers.

0.667 of a point when correct for two teachers.

0.334 of a point when correct for one teacher.

Identify correct procedures on three sub-problems (parts), each scored on this evaluation question:

Is this a correct way to solve this problem?

- a) This is a correct way to solve this problem. b) This is a right way to solve it but the wrong answer c) No, this is NOT a correct way to solve this problem d) I don't know

(Identify subtraction as an incorrect solution)

Problem 5
Part 1)

Yoko and Ken shot several free-throws in their basketball games. The result of their shooting is shown in the table.

One point for choice “c”

	Shots Made	Total Shots Tried
Ken	7	11
Yoko	3	5

Chloe solved it this way:

Ken: $11 - 7 = 4$ missed shots

Yoko : $5 - 3 = 2$ missed shots

Now that Chloe found who missed more, she compared only the shots missed, and decided that Yoko was better because she missed less shots

(Identify LCM as a correct solution strategy)

Part 2)

	Correct Guesses	Coin Tossed
Jess	15	20
Charlie	18	25

One point for choice “a”

Steven tried to solve it this way:

First he found the least common multiple for the denominators:

$$\text{Jess: } \frac{\quad}{100} \text{ and Charlie: } \frac{\quad}{100}$$

Then he found the numerators:

$$\text{Jess: } \frac{75}{100} \text{ and Charlie: } \frac{70}{100}$$

Then he compared the two fractions and decided Jess was better at guessing.

(Identify division as a correct solution)

Part 3)

One point for choice “a”

	Tickets Won	Games Played
Joe	11	27
Kai	4	9

Kevin tried to solve it this way:

First he divided:

Joe: $11 \div 27$

Kai: $4 \div 9$

and found that: Joe: 0.407 and Kai: 0.44

**Procedural
Flexibility
Construct**

**Posttest $\alpha = .67$
Pretest $\alpha = .57$**

Produce and evaluate more than one strategy for a problem

Problem 2

Please show your work.

Way 1

Way 2

Who is better _____?

Who is better _____?

Part 2)

Which way is better?

a) Way1 b) Way2

Why? a) *Less steps* b) *More steps but easier*,
c) *I don't know* d) *It's the only way I know*

Part 1)

Setup:

1 point for producing two different correct strategies.

Contender:

1 point for selecting two correct contenders and producing two correct strategies.

Part 2)

1 point if choice “way1/way2” referred to division and the student produced at least one correct strategy, but not subtraction.

1 point if choice a) describes division as having “less steps” given the student produced two correct strategies.

Identify the most efficient solution method: familiar

Problem 6

A weather channel in California (**TWC**) and a weather channel in New York (**KTL**) both tried to predict all the rainy days last month. The California weather channel (**TWC**) correctly predicted 5 rainy days out of 8 rainy days total. The New York weather channel (**KTL**) correctly predicted 14 rainy days out of 21 total rainy days.

1 point for choice a)

Which strategy will tell us which channel was more **accurate**, in the **least** number of steps:

Circle your answer:

- a) Divide $5 \div 8$ and $14 \div 21$
- b) Multiply $5 * 8$ and $14 * 21$
- c) Find the least common multiple for 21 and 8
- d) Subtract $8 - 5$ and $21 - 14$

Identify the most efficient solution method: Transfer

Problem 5 Part 3) was used with the following question:

1 point for choice a)

After thinking about it, Steven realized that he could also find out who is better by finding the least common multiple for the numerators at the start of the problem:



Charlie: $\frac{90}{\quad}$ Jess: $\frac{90}{\quad}$

Is this a correct way to solve this problem?

- a) This is a correct way to solve this problem.
 - b) This is a right way to solve it but the wrong answer
 - c) No, this is NOT a correct way to solve this problem
 - d) I don't know
-

Adapting procedure
to a novel problem
type and context

Problem 7

Adelina's Lemonade	Marcos' Lemonade
	

1 point for producing a
correct solution strategy

1 point for selecting the
correct contender, but
only if they produced a
correct strategy

Whose lemonade tastes more “lemony?”

Show all your work.

Recognizing the
correct solution &
setup on novel
problem type and
context:

Problem 8.


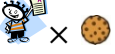
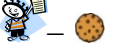
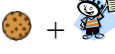
Yoko decided to divide her cookie jar
amongst her friends.

	Cookies	Friends
Yoko		

Problem omitted due to
ceiling effects (averages
ranged between 81%-
89%) and they were not
sensitive to the
intervention (average
change pre- to posttest
-3%)

Part 1) To figure out how many cookies each
friend gets, how should you set up the
problem?

Circle your answer

- a. 
- b. 
- c. 
- d. 

Part 1)
1 point for choice a)

Part 2) What units go with your answer in Part 1?

- a) cookies per friend
- b) friends times cookies
- c) cookies
- d) friends

Part 2)
1 point for choice a)

Evaluating explanation for solution strategy

Problem 1 (see above) was followed by the following question:

How do you know?

- a) *I compared the number of shots made*
- b) *I compared the number of shots tried*
- c) *I compared the shots made to shots tried*
- d) *I compared the number of shots missed*

1 point for choice c) only if student used a correct strategy and selected the correct contender for problem 1.

Knowledge of units: **Problem 5 Part 2) and Part 3)** were used with the following questions:

Write the labels that go with these numbers:

Part 2)

Jess: $\frac{75}{100}$ ----- Charlie: $\frac{72}{100}$ -----

Part 2)

1 point for writing „coins tossed „ correct guesses

Part 3)

Joe: 0 .407 -----

Kai: 0.44 -----

Part 3) Omitted from analysis; Only 6 out of 97 students responded correctly.

1 point for writing „games played „ tickets won

Problem 5 Part 3) was used with the following question:

What do the numbers .407 and .44 represent?
(Circle your answer)

- a) *The number of games played for each ticket won*
- b) *The number of tickets won*
- c) *The number of games played*
- d) *The number of tickets won for each game played*

1 point for choice a)

Problem 4 part b) was used.

- a) Which two classrooms have the same amount (ratio) of cookies to students?

Omitted from analyses (see Problem 4)

1 point for correct answer.

<p>Produced Misconception</p> <p><i>Used strategy shown to be invalid during instruction</i></p>		<p>Posttest $\alpha = .66$ Pretest $\alpha = .46$</p>
	<p>Problems 1, 2, 3, and 7</p>	<p>1 point if students used subtraction as a solution strategy.</p>
<p>Use of Most Efficient Strategy Construct</p>		<p>Posttest $\alpha = .66$ Pretest $\alpha = .83$</p>
<p>All problems taking the form of the instruction were scored to evaluate use of the most efficient strategy</p>	<p>Problems 1, 2, 3, and 7</p>	<p>1 point for using division to solve the problem.</p>
<p>Negative Transfer Construct</p> <p>Problems in which strategy shown to be invalid during instruction is correct</p>		<p>Posttest $\alpha = .68$ Pretest $\alpha = .58$</p>

Correctly producing subtraction

n=2

Problem 9.

Ken and Yoko participated in a swimming competition. They each swam several times. Who won more times in all?

	Number of Losses	Number of Swims
Ken	8	20
Yoko	12	25

Please show all your work.

Who won more _____

How do you know?

- a) *I compared the number of swims*
- b) *I compared the number of losses*
- c) *I compared the swims won to swims lost*
- d) *I compared the number of wins*

n=1

Problem 10.

A weather channel in California (TWC) and a weather channel in New York (KTL) both tried to predict all the rainy days last month. The California weather channel (TWC) correctly predicted 5 rainy days out of 8 rainy days total. The New York weather channel (KTL) correctly predicted 14 rainy days out of 21 total rainy days.

Which strategy will tell us which channel was more accurate, in the least number of steps:

Circle your answer:

- a) *Divide $5 \div 8$ and $14 \div 21$; b) Multiply $5 * 8$ and $14 * 21$; c) Find the least common multiple for 21 and 8; d) Subtract $8 - 5$ and $21 - 14$*

1 point for producing the correct strategy (subtracting 20-8 and 25-12).

1 point for selecting the correct contender and using a correct strategy.

The multiple choice question was omitted due to floor effects.

1 point for choice d)