Creating Mappings When Learning From Multiple Representations:

The Role of Self-Generation

BY

KEVIN D. DIETZ
B.A., University of Michigan, 2007

THESIS

Submitted as partial fulfillment of the requirements for the degree of Master of Arts in Psychology in the Graduate College of the University of Illinois at Chicago, 2012

Chicago, Illinois

Defense Committee:

Susan R. Goldman, Chair and Advisor
James W. Pellegrino
Jennifer Wiley
Mike Stieff, Chemistry
This thesis is dedicated to my parents, Doug and Debbie Dietz, for all their support.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Learning from Multiple Representations</td>
<td>1</td>
</tr>
<tr>
<td>Improving Mappings Between Multiple Representations</td>
<td>5</td>
</tr>
<tr>
<td>The Generation Effect</td>
<td>7</td>
</tr>
<tr>
<td>Aims for the Present Study</td>
<td>8</td>
</tr>
<tr>
<td>METHOD</td>
<td>11</td>
</tr>
<tr>
<td>Participants</td>
<td>11</td>
</tr>
<tr>
<td>Design</td>
<td>11</td>
</tr>
<tr>
<td>Materials</td>
<td>12</td>
</tr>
<tr>
<td>Probability Theory</td>
<td>12</td>
</tr>
<tr>
<td>Worked-examples</td>
<td>13</td>
</tr>
<tr>
<td>Learning Measures</td>
<td>13</td>
</tr>
<tr>
<td>Common Measures Between Pre-test and Post-test</td>
<td>15</td>
</tr>
<tr>
<td>Unique Pre-test Items</td>
<td>16</td>
</tr>
<tr>
<td>Unique Post-test Items</td>
<td>17</td>
</tr>
<tr>
<td>Procedure</td>
<td>18</td>
</tr>
<tr>
<td>Pre-test</td>
<td>18</td>
</tr>
<tr>
<td>Study Phase</td>
<td>18</td>
</tr>
<tr>
<td>Post-test</td>
<td>20</td>
</tr>
<tr>
<td>Scoring</td>
<td>21</td>
</tr>
<tr>
<td>Procedural Knowledge Items</td>
<td>21</td>
</tr>
<tr>
<td>Conceptual Knowledge Items</td>
<td>21</td>
</tr>
<tr>
<td>Think-aloud Coding</td>
<td>22</td>
</tr>
<tr>
<td>Justification Statements</td>
<td>22</td>
</tr>
<tr>
<td>Mappings</td>
<td>24</td>
</tr>
<tr>
<td>RESULTS</td>
<td>25</td>
</tr>
<tr>
<td>Conceptual Knowledge</td>
<td>26</td>
</tr>
<tr>
<td>Procedural Knowledge</td>
<td>26</td>
</tr>
<tr>
<td>Justification Statements</td>
<td>28</td>
</tr>
<tr>
<td>Mappings</td>
<td>30</td>
</tr>
<tr>
<td>Relationship Between Knowledge Outcomes and Think-aloud Measures</td>
<td>33</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>37</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>41</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>44</td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS (continued)

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPENDIX B</td>
<td>52</td>
</tr>
<tr>
<td>APPENDIX C</td>
<td>59</td>
</tr>
<tr>
<td>APPENDIX D</td>
<td>63</td>
</tr>
<tr>
<td>APPENDIX E</td>
<td>69</td>
</tr>
<tr>
<td>APPENDIX F</td>
<td>71</td>
</tr>
<tr>
<td>IRB APPROVAL</td>
<td>73</td>
</tr>
<tr>
<td>VITA</td>
<td>76</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. CORRELATIONS BETWEEN MAPPING TYPES, JUSTIFICATION STATEMENTS, AND POST-TEST PERFORMANCE</td>
<td>34</td>
</tr>
<tr>
<td>II. SCORES AND CRITERIA FOR INDEPENDENT EVENTS AND MUTUAL EXCLUSIVITY</td>
<td>49</td>
</tr>
<tr>
<td>III. SCORES AND CRITERIA FOR ORDER RELEVANCE VERSUS IRRELEVANCE</td>
<td>49</td>
</tr>
<tr>
<td>IV. SCORES AND CRITERIA FOR WITH REPLACEMENT VERSUS WITHOUT REPLACEMENT</td>
<td>50</td>
</tr>
<tr>
<td>V. SCORES AND CRITERIA FOR MULTIPLICATION RULE</td>
<td>51</td>
</tr>
<tr>
<td>VI. PROCEDURAL JUSTIFICATIONS AND EXAMPLES</td>
<td>65</td>
</tr>
<tr>
<td>VII. CONCEPTUAL JUSTIFICATIONS AND EXAMPLES</td>
<td>66</td>
</tr>
<tr>
<td>VIII. TYPES OF MAPPINGS</td>
<td>68</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>14</td>
</tr>
<tr>
<td>2.</td>
<td>19</td>
</tr>
<tr>
<td>3.</td>
<td>27</td>
</tr>
<tr>
<td>4.</td>
<td>28</td>
</tr>
<tr>
<td>5.</td>
<td>29</td>
</tr>
<tr>
<td>6.</td>
<td>32</td>
</tr>
<tr>
<td>7.</td>
<td>36</td>
</tr>
<tr>
<td>8.</td>
<td>62</td>
</tr>
<tr>
<td>9.</td>
<td>62</td>
</tr>
<tr>
<td>10.</td>
<td>70</td>
</tr>
<tr>
<td>11.</td>
<td>72</td>
</tr>
<tr>
<td>12.</td>
<td>72</td>
</tr>
</tbody>
</table>

1. Example of how an item was presented during the study phase
2. Procedure of the current study
3. Adjusted mean percentage correct of conceptual knowledge items at post-test by learning condition
4. Adjusted mean percentage correct of procedural knowledge items by learning condition and procedural knowledge measures
5. Mean number of justification statements by learning condition and justification type as evidenced by participants’ think-alouds during the study phase
6. Number of mapping types by learning condition and mapping category
7. Path diagram of mediation model for post-test conceptual items
8. Mean percentage correct of conceptual knowledge items at pre-test and post-test by learning condition in pilot study
9. Mean percentage correct of procedural knowledge items by learning condition and test type in pilot study
10. Mean number of token mappings made during the study phase by learning condition and mapping category
11. Path diagram of mediation model for isomorphic transfer procedural items
12. Path diagram of mediation model for far transfer procedural items
LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANCOVA</td>
<td>Analysis of Covariance</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>BCA</td>
<td>Bias Corrected and Accelerated</td>
</tr>
<tr>
<td>HSD</td>
<td>Honestly Significant Different</td>
</tr>
</tbody>
</table>
SUMMARY

Students are often able to learn a concept effectively if it is presented in multiple formats and they map common elements across the representations (Ainsworth, 2006). However, low-knowledge learners often fail to do this. Recent research has demonstrated some benefits to learning for low-knowledge individuals who explain commonalities in representations that are provided to them (Berthold & Renkl, 2009). The present study tested whether there was an advantage to having learners generate their own scaffolded mappings and explain them compared to explaining provided mappings, especially with respect to conceptual, as opposed to procedural, understanding of the to-be-learned material. Seventy-five low-knowledge participants studied worked-examples of probability word problems as they generated explanations under one of three conditions: when no mappings were provided, when researcher-produced mappings were provided, or when the participant generated the mappings with scaffolded assistance. Participants were also instructed to think-aloud during the study process, in contrast to prior work on this topic. Results indicate the condition that required explaining scaffolded learner-generated mappings led to higher conceptual, but not procedural, knowledge at post-test compared to the researcher-provided mappings or no mapping conditions. Analyses of think-aloud protocols collected during the solution process suggest that these findings may be due to the higher number of conceptual justification statements produced when learners generate their own mappings. Additionally, there is evidence that these types of justification statements mediated the relationship between the acquisition of conceptual knowledge and the types of mappings learners made between the multiple representations.
INTRODUCTION

When learning about a new concept it is not uncommon for a student to see information about that concept presented in many different formats. For example, when algebra students learn about rate of change, they may encounter the exact same data across multiple representations. The rate of change of an object could be represented algebraically in an equation, visually in a static or animated graphic, or perhaps diagrammatically on a chart. Although the same essential information is presented in each representation (i.e. the rate of change), some empirical evidence suggests that having learners engaging with multiple representations can yield greater learning gains compared to providing the learner with only one representation of the information (e.g., Kolloffel, Eysink, de Jong, & Wilhelm, 2009). But there is also evidence to suggest that multiple representations may not be beneficial to all learners and may even harm performance (Ainsworth, Bibby, & Wood, 2002). Often the effectiveness of multiple representations is determined by how learners use them and the supports provided for their use.

Learning From Multiple Representations

The prevalence of multiple representations in learning environments and the mixed results concerning their impact on learning have produced a number of accounts of the processes and outcomes of learning from multiple representations. For example, Paivio (Paivio, 1986; suggest that multiple representations are most effective in learning if representations tap into different means of cognitive processing, that is, the visual and verbal processing channels. Each processing channel has a limited capacity, so representations that rely too heavily on one mode of processing will not be as effective as if the processing is more evenly distributed across both channels. According to Paivio and Clark (1991), when two representations—one visual and
and one verbal—convey similar information, the relevant information from each representation is used to create separate verbal and visual mental models and connections are made between the two models, thereby mapping the visual and verbal representations to one another. Similarly, Schnitz and Bannert (2003) propose that multiple representations allow for the construction of complementary mental models that emphasize the similarities and differences between the representations, thereby making salient the analog structure common across representations.

Often success of learning from multiple representations is measured by increased procedural and conceptual knowledge. Procedural knowledge refers to how to solve a problem and often is reflected in the performance of a series of steps in the problem-solving process. Conceptual knowledge refers to a deeper understanding of why certain procedures are used and is often reflected in explanations of problem solutions. In mathematics, conceptual understanding should demonstrate knowledge about the underlying mathematical principles or rules (Rittle-Johnson, Siegler, & Alibali, 2001). Ainsworth (1999; 2006) postulated that multiple representations provide three key benefits to procedural and conceptual learning over singular representations. First, multiple representations serve a complementary function; each representation may provide different information to the learner, or the same information but through different processing modalities. Second, one representation may constrain the ways in which another representation can be interpreted. That is, an easy-to-understand or already-known representation can facilitate the interpretation of a more difficult or unknown representation. Finally, multiple representations create opportunities for learners to construct links across representations. Therefore, multiple representations can be beneficial to learning because they allow for flexibility in how information is presented and because the learner is often required to integrate information, through mapping conceptual commonalities and differences across
representations (Ainsworth & VanLabeke, 2004). Successful mapping results in conceptual understanding of the underlying principles and allows for the transfer of learned knowledge to new situations (Ainsworth, 2006).

Learning from multiple representations is only advantageous to learners who actually engage in mapping between representations. Low-knowledge learners, or novices, do not spontaneously engage in mapping between representations (Ainsworth, et al., 2002; Ainsworth & VanLabeke, 2004). Low-knowledge learners may have difficulty with multiple representations for several reasons. First, Seufert and Brunken (2006) suggest that when low-knowledge learners—in contrast to more knowledgeable learners—engage with multiple representations, they are making connections at only a surface level, rather than at a conceptual level. To illustrate what the authors mean, consider low-knowledge learners studying a mathematics problem solved using two different representations (i.e., an equation and a diagram). Low-knowledge learners connect the words and numbers in one representation to the corresponding words and numbers in the other representation. Low-knowledge learners do this instead of focusing on why the underlying mathematical principles conveyed in each representation relate to one another. Therefore, when low-knowledge learners do actually engage with multiple representations they receive little to no benefit to conceptual understanding of the to-be-learned material because they often only make surface level connections between the representations; they do not relate the material on the basis of the underlying principles. The authors argue that relating underlying principles is a highly demanding process and if it could be made less process-intensive then low-knowledge learners would have the cognitive resources to relate multiple representations at conceptual rather than surface levels (Seufert & Brunken, 2006).
Second, low-knowledge learners’ poor performance with multiple representations may be due to their failure to recognize how various pieces of information are being communicated through the different representations. That is, they fail to recognize the important connections between the representations and may often conclude that the representations provide redundant information. As a result, low-knowledge learners frequently decide to focus on only one representation (Ainsworth, 2006).

Third, low-knowledge learners may not understand why it is important to make connections between representations. Schownke and colleagues (Schownke, Berthold, & Renkl, 2009) eye-tracked low-knowledge learners as they studied worked-examples of complex probability events. Each worked-example was represented as a diagram and as an equation. After studying the worked-examples, participants watched a playback of their eye movements and verbally reported what they thought they had been thinking during the eye-tracked phase of the study. From analyzing the participants’ retrospective think-alouds, the authors concluded that low-knowledge learners did not understand the importance of mapping commonalities between representations and, as a consequence, the learners mostly mentioned surface features and rarely mentioned conceptual similarities between the representations. That is, low-knowledge learners appeared to be unaware of the functional benefits that multiple representations afford or why it might be beneficial to relate conceptual information in one representation to conceptual information in another representation. A number of researchers have explored whether providing additional guidance for low-knowledge learners through scaffolding, training, or additional instruction facilitates conceptual connections across representations.
Improving Mappings Between Multiple Representations

Because low-knowledge learners tend not to connect conceptually-related elements between representations spontaneously, a number of studies have attempted to scaffold this process. One of the most studied scaffolds is prompting learners to explain their reasoning, in part because it can be beneficial for mapping visual and verbal elements across representations (Aleven & Koedinger, 2002). Indeed, based on a review of the impact of explanation on multimodal learning, Roy and Chi (2005) concluded that prompting explanations during study might be particularly helpful for low-knowledge learners because the explanation process facilitates mapping between representations by requiring the learner to focus on conceptual similarities and differences. That is, by engaging in explanation learners are required to integrate information in order to make sense of the correspondence between representations. Roy and Chi’s (2005) conclusions are also supported by more recent research on mathematics learning by Berthold, Eysink, and Renkl (2009). Berthold et al. (2009) used print-based questions to prompt some participants to provide written explanations as they studied worked examples; others were simply provided with a text box in which they could “take notes”. Two important findings emerged. First, learning outcomes were greater when participants were prompted for written explanations as compared to “note taking.” Second, explanation prompts produced more justifications about the procedures and concepts underlying the worked examples compared to the “note taking” condition.

Furthermore, certain types of representations may better promote generating conceptual justifications about the underlying concepts when learners are prompted to explain. Conceptual justifications refer to statements that go beyond stating how a series of steps can be used to arrive at a correct answer, and explain why particular concepts are important and useful to arrive at the
correct answer. In contrast, procedural justifications are statements that explain how a series of steps can be used to arrive at the correct answer. Ainsworth and Loizou (2003) provide evidence that including diagrammatic representations with text-based representations of the to-be-learned material prompted learners to generate more conceptual justifications about their understanding of the material compared to providing purely text-based representations (Ainsworth & Loizou, 2003). This finding suggests it may be important to provide diagrammatic representations of the material, in addition to text-based representations, when learning from multiple representations in order to promote greater conceptual understanding through learners’ increased production of conceptual justification statements.

A second type of scaffold, referred to by Berthold and Renkl (2009) as “relating aids,” has shown beneficial effects for conceptual understanding. Relating aids make salient the connections between representations by drawing attention to corresponding features in different representations. For example, a relating aid may consist of highlighting parts of two different representations that correspond to the same underlying element. Berthold and Renkl (2009) presented multiple representations of worked examples either with or without highlighting. They found that highlighting led to a slight increase in conceptual understanding of the to-be-learned mathematical content compared to when there was no highlighting. At the same time there was no effect of highlighting on procedural knowledge. Based on this pattern, Berthold and Renkl (2009) argued that the positive effect for conceptual understanding was due to the highlighting having promoted conceptually-relevant mappings between representations. In this same study, the authors also examined the impact of relating aids on the quality of explanations generated by the learners. Contrary to their expectations, the number of explanations related to procedural or conceptual knowledge was not enhanced by the presence of relating aids. One interpretation of
this finding is that the mappings conveyed by experimenter-provided highlighting were not procedurally or conceptually meaningful to the learners and, as a result, learners focused on surface-level commonalities rather than searching for deeper conceptual justifications. Thus, although the relating aids provided low-knowledge learners with mappings that they were not spontaneously making, this did not enhance explanations or overall procedural learning performance.

**The Generation Effect**

The lack of an effect on explanations and procedural knowledge, as well as the modest effect on conceptual knowledge, may be due to the fact that the relating aids provide too much scaffold assistance to the learner. What may be needed to enhance learning in low-knowledge learners is a scaffolded prompt that assists learners in creating mappings that go beyond surface relations. In this instance, a scaffolded prompt would provide the learner with one half of the mapping and prompt the learner to generate the corresponding second half of the mapping. That is, the relating aid would be present in one representation (i.e. highlighting in only the diagram) and the learner would need to generate a mapping to the corresponding element in the other representation. In effect, such a prompt creates a “self-generation” condition, an instance of the generation effect. The generation effect is a well-established and robust psychological phenomenon in which learners are better able to retain information that they generate themselves compared to information that is presented to them (Bertsch, Pesta, Wiscott, & McDaniel, 2007). The generation effect has been observed in a variety of laboratory-based memory studies (e.g., Slamecka, & Graf, 1978) as well as in subject matter learning. For example, in mathematics Dahlberg and Housman (1997) found that college-level mathematics students who self-generated examples and connections between mathematical ideas better retained the information compared
to students who did not self-generate. Likewise, it may be the case that connections across representations that learners themselves generate are more effective than those generated for the learner.

**Aims For The Present Study**

The present study explored the effects on conceptual and procedural knowledge of prompting learners for explanations under three conditions of relating aids: learner-generated, provided (and thus generated by the researcher), and none. A modified version of the Berthold and Renkl (2009) paradigm was used as the task in which low-knowledge learners studied worked-example probability problems that used two representations - a diagram and an equation of the correct solution. Participants either (1) generated a scaffolded mapping between the representations and explained the mapping (Generate + Explain), (2) explained researcher-provided mappings (Provide + Explain), or (3) explained the worked-example solutions with no guidance about mappings (Explain Only). In the Generate + Explain condition, the generation was scaffolded, or guided, after pilot-testing revealed that without it the generation task was too open ended to be productive for learning. The scaffold provided the learner with one half of the mapping and the learner generated the corresponding second half of the mapping. Specifically, the diagrammatic worked-example contained highlighting on the important elements for problem solution and the learner generated a mapping to the corresponding elements in the equation-based worked-example.

We hypothesized that the learner-generated scaffolded mapping condition (Generate + Explain) would be more beneficial to conceptual understanding than when provided mappings were explained (Provided + Explain) or when there were no cues for mapping (Explain Only). However, no differences among conditions were expected on procedural knowledge. These
predictions were based on previous findings indicating that relating aids facilitate modest conceptual understanding but not procedural understanding, and that explanation enhances procedural understanding and conceptual understanding (Berthold, et al., 2009; Berthold & Renkl, 2009). Because all three conditions engaged in explanation, we expected an increase in procedural understanding from pre-test to post-test, and did not expect the increase to be differentially affected by learning condition.

Also of interest in the present study were the processes by which people solved these probability problems and how they interpreted the worked examples. Past research has attempted to examine processing data using eye-tracking in conjunction with retrospective think-alouds while watching playbacks of the tracked eye-movements (Schownke, et al., 2009). Other work has looked at prompted written justifications for part of the solution process. For example, Berthold, et al. (2009) had participants provide written explanations to text-based prompts as they studied multiple worked-examples. But these prompts were rather specific in nature (e.g. “Why do you calculate the total possible outcomes by multiplying”). Neither of these studies provided the sort of online processing data that concurrent think aloud protocols can provide (Ericsson & Simon, 1980; Young, 2005). In the present study, we were interested in what representations and parts of the representations participants processed as well as how they talked about the solutions that were depicted in the worked examples. In particular, we were interested in (1) whether manipulating the presence of relating aids led to differences in what parts of the representations participants said they examined and mappings between representations that they voiced; and (2) if there were differences among the experimental conditions in the discussion of the mathematics involved in the problems. Accordingly, we derived indices relevant to these questions from the comments participants made when thinking aloud while studying the worked
examples. However, it was not our intent to create process traces of the actual solution process. Thus, the information derived from the think-aloud protocols provides insights into the mechanisms that underlie observed condition differences and allows us to examine the relationships between the justification statements (i.e., procedural and conceptual) made by participants, the types of mappings between representations learners are engaging in, and differences in procedural and conceptual learning at post-test. Specifically, conceptual justifications should be related to participants’ conceptual understanding. Additionally, because Ainsworth and Loizou (2003) suggested that mapping commonalities between different types of representations (i.e. a diagrammatic worked example and an equation-based worked example of the same mathematical problem) may increase the number of conceptual justifications produced by learners, we expected diagram-equation mappings to be positively associated with conceptual justifications, which, in turn, would be positively associated with conceptual understanding. We would not expect to observe a similar relationship with other types of mappings—notably between a word problem and a diagrammatic worked example of the problem, or between a word problem and an equation-based worked example that problem—because these types of mappings do not focus on making connections between two representations that are conveying similar conceptual information. Therefore, if the above-mentioned hypotheses regarding conceptual understanding are supported, we would expect the Generate + Explain condition to also produce more conceptual justifications and more mappings between the diagrammatic-based worked example and the equation-based worked example compared to the other two learning conditions.
METHOD

Participants

Eighty-one UIC undergraduates (46 females, $M_{\text{age}} = 19.67; SD_{\text{age}} = 2.39$) with low-knowledge about complex probability theory were recruited from the psychology subject pool to participate in this study. Low-knowledge was defined as failing to score above 50% correct on a pre-test for both procedural and conceptual knowledge about complex probability events. Six participants were excluded from analysis because either they did not answer all pre-test and post-test questions ($n = 3$) or they failed to correctly answer screening questions about basic probability theory ($n = 3$). The remaining 75 participants were randomly assigned to one of the three study conditions. Participants in the three learning conditions did not differ with respect to their knowledge about basic probability concepts (Generate + Explain: $M = 2.60$, $SD = .50$; Provided + Explain: $M = 2.72$, $SD = .46$; Explain Only = 2.71, $SD = .41$), $F(2, 72) = 1.21$, ns. Additionally, participants in the various learning conditions did not spend different amounts of time studying the worked examples during the study phase, $F(2, 72) = 2.39$, ns. The Generate + Explain condition ($M = 13.42$ minutes, $SD = 4.5$ minutes) spent approximately the same amount of time studying the worked examples as the Provided + Explain condition ($M = 10.95$ minutes, $SD = 4.01$ minutes) and the Explain Only condition ($M = 11.60$ minutes, $SD = 3.80$ minutes).

Design

The basic design is a repeated measure mixed-model analysis of variance (ANOVA), with the learning condition (Generate + Explain, Provided + Explain, Explain Only) as the between-subjects variable. Learning condition was the focal independent variable. Justification statements (procedural, conceptual) and types of mappings (between the problem and diagram, between the problem and equation, between the diagram and equation) were within-subjects
variables. Post-test scores (procedural isomorphic-transfer items, procedural near-transfer items, procedural far-transfer items, conceptual post-test items), number of justification statements, and the number of mapping types were the repeated measures in these analyses. These terms will be elaborated on in the following subsections. Where appropriate, pre-test scores were entered as a covariate. Items in the pre-test and post-test were counterbalanced, and no effects of item set were observed, $F$s < 1.

**Materials**

**Probability Theory**

The learning material came from the domain of probability theory. Probability theory has often been used in previous studies on learning from multiple representations (see Berthold & Renkl, 2009; Große & Renkl, 2006; Kolloffel, et al., 2009; Schwonke, et al., 2009) because university students generally have basic knowledge about probability (e.g. the probability of rolling a “2” on a six-sided die), yet lack a more advanced understanding of complex probability events (e.g. permutations and combinations). Therefore, the mathematical concepts presented to participants are relatively unfamiliar to most participants and neither too easy nor too difficult. The unfamiliarity and difficulty about complex probability events has been supported by previous research (e.g. Berthold & Renkl, 2009) and by pilot data collected at UIC for the study problem sets. Specifically within probability theory, problem sets focusing on permutations and combinations were used, inline with material used in past research (e.g. Berthold & Renkl, 2009, Große & Renkl, 2006; Kollofel, et al., 2009; Schownke, et al., 2009; Seufert & Brunken, 2006). Essentially, in permutations the order in which objects are selected matters; in combinations, as two objects are selected simultaneously, the order is irrelevant. Permutations and combinations can be further broken down into whether an item, once selected, has the possibility of being
selected again or not. This is termed “with replacement” and “without replacement,” respectively. For this study, only three problem types were used: permutations with replacement, permutations without replacement, and combinations without replacement. Combinations with replacement problems were not used because this type of probability event is not frequently taught in conjunction with the other three types of complex probability events due to its abstract nature (i.e. the same object can be simultaneously selected twice, despite only one such object existing). Additionally, based on pilot data, participants had more difficulty understanding worked-examples of combination with replacement problems than any other problem type. This result may be because combination with replacement problems do not easily lend themselves to word problems and may be more difficult to represent diagrammatically without providing additional details as the meaning of various elements in the diagram.

**Worked-examples**

Participants were presented with six worked-example probability word problems, two of each type described above. The worked-examples were presented one at a time on a computer screen. At the top of the screen a probability word problem was presented. Below the word problem, an arithmetic solution and a pictorial tree diagram solution were displayed. The answer to the word problem was displayed at the bottom right-hand side of the screen (see Figure 1).

**Learning Measures**

Participants’ procedural and conceptual knowledge was assessed both before and after the study phase using a pre-test/post-test design. The pre-test and post-test asked comparable questions, with the exception of basic probability (pre-test only) and isomorphic- and far-transfer items (post-test only), as elaborated below. The learning measures consisted of items that tapped into procedural and conceptual knowledge.
Figure 1. Example of how an item was presented during the study phase. From top to bottom: how the item was presented in the Generate + Explain, Provided + Explain, and Explain Only conditions.
Procedural knowledge items asked subjects to solve problems by using certain probability principles (i.e. with replacement, without replacement, order relevance, order irrelevance). An example of a procedural knowledge item is: “There are 5 different colored balls in a bag: red, blue, yellow, pink, and green. Two balls are selected at random, one at a time. After a ball is selected it is placed back in the bag. What is the probability that the green ball will be picked first and the red ball picked second?” Conceptual knowledge items asked participants to explain the importance of certain probability principles for solving the problems they encountered. That is, why a certain procedure or principle was used to arrive at the correct solution. For example, a conceptual knowledge item asked: “In order to calculate the probability of two or more events occurring the probabilities of each event are multiplied, rather than added together. Why do you think this is?” The complete set of procedural and conceptual knowledge items can be found in Appendix A.

**Common Measures Between Pre-Test and Post-test**

There were two common learning measures between pre-test and post-test, one assessing procedural knowledge and one assessing conceptual knowledge. The common procedural knowledge learning measure was a question set of six word problems about complex permutation and combination events. Items in the pre-test and post-test procedural sets were counterbalanced to control for any differences in problem set difficulty. The six items in each question set were distributed as follows: two permutations without replacement problems, two permutation with replacement problems, and two combination without replacement problems. In the pre-test this question set was referred to as the “procedural pre-test” and in the post-test as “near-transfer post-test.” Near-transfer post-test questions can be directly compared to the pre-test procedural questions because both question sets asked participants to solve structurally
similar problems, yet did not contain the same surface story of previously encountered problems. Therefore, near transfer problems were able to assess participants’ procedural understanding of about how to solve structurally similar, with dissimilar surface features, probability problems.

The common conceptual knowledge learning measure was a question set consisting of four open-ended questions designed to assess participants’ deeper understanding of complex probability events. For example, to assess conceptual knowledge, participants were asked to explain why two probability fractions are multiplied rather than added together. Post-test questions about conceptual knowledge can be directly compared to pre-test questions, as both question sets ask participants to explain their knowledge about underlying probability concepts (i.e. order relevance vs. irrelevance, replacement vs. no replacement, mutual exclusivity, and the multiplication rule). These question sets were counterbalanced across pre-test and post-test to control for any differences in problem set difficulty.

**Unique Pre-test Items**

In addition, the pre-test assessed participants’ prior knowledge about basic probability concepts (e.g. the probably of rolling a “2” on a 6-sided die). These questions about basic probability were necessary to ascertain whether participants had a basic understanding of probability events. There were three questions that assessed basic probability. If a participant was not able to answer at least two of these questions correctly then that participant was excluded from analysis. This exclusion is based on the high likelihood that an individual with limited knowledge about simple probability events would not be able to learn about complex probability events, as the former is necessary to understand the latter.
**Unique Post-test Items**

Additionally, the post-test assessed the extent to which participants could apply the procedural knowledge they learned from the study phase. This was achieved through isomorphic-transfer items and far-transfer items. Six isomorphic-transfer items assessed whether participants were able to apply what they learned to solve problems very similar to those they studied. This is accomplished through presenting the same word problems they encountered as worked-examples during the study phase except with the values of the variables in the problem changed in ways that still preserve the surface structure and solution method of the problem. Isomorphic-transfer problems at post-test were counterbalanced with the studied worked examples to control for any differences in difficulty between the problem sets.

Far transfer items assessed a participant’s ability to apply procedural knowledge to problems that differed in both surface and underlying structure from those they studied. That is, far-transfer items required combining knowledge about multiple probability principles in order to arrive at the correct solution. For example, one far-transfer item focused on permutation without replacement and combination without replacement, whereas an item from the other types of procedural knowledge questions would focus on permutation without replacement or combination without replacement, but not both. In other words, these items were constructed by “blending” two probability principles together. Therefore, no direct comparison can be made between far-transfer and pre-test items. However, far-transfer items are useful in determining whether participants could apply their knowledge to problems beyond those they directly studied. There were six far-transfer problems, two of each blended possibility. In sum, these two types of transfer problems allowed for a better understanding of how far participants could apply their procedural knowledge learned from studying the worked-examples.
**Procedure**

The study was administered to participants in individual sessions lasting 90 minutes. Participants used a computer to work through the pre-test and post-test questions, study the worked-examples, type written responses to the open-ended conceptual knowledge questions, and submit their answers. All instructions were presented on the computer screen. Participants’ written work, verbalizations, and gestures were recorded using the *CaptureFlux* video and audio capture software (Glagla, 2008). Answers to items and response times were recorded through *E-Prime* (Version 2.0). The study was self-paced by the participant. Additionally, paper and pencil were provided to participants. An overview of the procedure can be found in Figure 2. Upon beginning the experiment participants first received general instructions about the study (see Appendix B for all instructions). The intent of these instructions was to orient the participant to the general concepts of the tasks they would be engaging in and to provide guidance about how to submit their answers. After the initial instructions the participants completed the pre-test.

**Pre-test**

The pre-test consisted of two parts in a fixed order. First the procedural knowledge pre-test was administered followed by the conceptual knowledge pre-test. Items in each question set were presented in a randomized order.

**Study Phase**

Following the pre-test, participants entered the study phase of the experiment. First, participants received instructions about the types of problems they were going to study, the presentations and format of the worked-examples (i.e. diagrammatic and equation-based worked-examples of complex probability word problems), how to explain, and how to think-aloud. Participants were also presented with an example item similar to those they were about to study.
Figure 2. Procedure of the current study.

**Pre-test**
- Procedural knowledge item set (6 questions)
- Conceptual knowledge item set (4 questions)

**Study Phase**
- Receive condition-specific instructions
- Study 6 worked examples
  - Each worked example is presented diagrammatically (as a tree diagram) and arithmetically (as an equation).
  - All participants instructed to explain as they studied
  - Important elements were highlighted in just the diagram (Generate + Explain), in both the diagram and equation (Provided + Explain), or in neither representation (Explain Only)

**Post-test**
- Isomorphic transfer item set (6 questions)
- Near transfer item set (6 questions)
- Far transfer item set (6 questions)
- Conceptual knowledge item set (4 questions)
The instructions differed slightly depending on the condition. That is, in addition to the above instructions, the Generate + Explain condition was also instructed to pay particular attention to the important elements highlighted in the tree diagram\(^1\) and map those elements to parts in the equation; the Provided + Explain condition was instructed to explain why the corresponding highlights in the tree diagram and the equation were important; the Explain only condition was instructed to explain why one needs to multiply the two probability fractions together to solve the problem. See Appendix B for the exact instructions each learning condition received.

During the study phase participants studied six worked-examples. The worked examples were presented in a blocked randomized order, so that both problems of the same type were presented back to back. This type of blocking has typically been used in previous studies (e.g. Berthold & Renkl, 2009).

All participants were required to think aloud as they studied the worked-examples. The think-aloud protocols were audio and video recorded then later transcribed. The transcripts, in addition to capturing verbalizations, also captured gestures (e.g. pointing to specific parts of the worked-examples) made by the participants in an effort to better understand how participants made use of and understood the worked-examples.

**Post-test**

Following the study phase, participants were given a post-test. In the post-test, participants were asked to solve isomorphic-transfer problems, followed by near-transfer problems, then far-transfer problems, and finally, open-ended conceptual knowledge questions. Questions within each subset of the post-test were presented in a randomized order. Before the

---

\(^1\) Pilot data indicated that participants performed better on post-test conceptual knowledge items when the highlights appeared in the tree diagram and participants were instructed to map to the equation than when highlights appeared in the equation and participants were instructed to map to the tree diagram. Therefore, the decision was made to place the highlights in the tree diagram for the Generate + Explain condition in order to maximize performance on conceptual understanding (see Appendix C for more details).
post-test began, participants were informed that they would be asked to solve problems similar to the ones they had just studied and to use what they learned to answer these questions.

**Scoring**

**Procedural Knowledge Items**

Procedural pre-test and post-test knowledge items were scored as either correct (1 point) or incorrect (0 points). Probability fractions equivalent to the correct answer were scored as correct, as was the corresponding decimal value. For example, if the answer to an item was “2/10” then “1/5,” “0.20,” and other equivalent responses were also scored as a correct.

**Conceptual Knowledge Items**

Each conceptual knowledge item was scored on a three-point scale: no credit (0 points), partial credit (0.5 points), and full credit (1 point). An item was given a score of “0” if the item was not answered or if no part of the answer was correct. An item received a score of “0.5” points if a partially correct answer was provided. A score of “1” was given for completely correct answers. For instance, when scoring a conceptual question focused on the principle of order relevance/irrelevance, a participant received 0.5 points for explaining how order irrelevance allows for desirable events to occur in multiple ways (i.e., Event A or Event B could occur first) compared to only one acceptable outcome when order is relevant (i.e., Event A must occur first). A participant would receive an additional 0.5 points for explaining that in either instance the total possible number of outcomes remains constant, but when the order is irrelevant it increases the total number of desirable outcomes, thus increasing the overall probability, compared to when the order is relevant. A scoring rubric for conceptual items can be found in Appendix A. Two raters scored conceptual knowledge items independently, blind to learning condition. Each rater scored each conceptual knowledge item for each participant (4 pre-test
conceptual knowledge items, 4 post-test conceptual knowledge items for 75 participants), for a total of 600 units coded for reliability. According to Landis and Koch (1977), inter-rater reliability was considered “substantial” (Cohen’s Kappa = .75). For differently scored items the raters discussed their scores with each other to reach consensus.

**Think-aloud Coding**

Students were instructed to think-aloud during the study phase of the experiment. Think-aloud protocols were analyzed to determine what aspects of the representations participants focused on. Two indices were derived from participants’ think-alouds: the types of justification statements produced and the number of mappings made between the various representations. Think-alouds were segmented into utterances that were separated by pauses of at least two seconds and each utterance was coded for what representations participants were focusing on and what type of justification statements, as described below, were used.

**Justification statements**

The think-aloud protocols were analyzed for statements participants made about the various probability principles presented in the worked examples and the justifications they provided for the importance of these principles. A statement that related a solution step or series of steps in the worked examples to a probability principle (i.e. order relevance, order irrelevance, with replacement, without replacement, multiplication rule) was considered a procedural justification statement. For example, a procedural justification for the probability concept of without replacement is: “when you take one [marble] out it is not put back in for the second event.” Sometimes participants provided justifications that went beyond stating that a solution step related to a probability principle and explained why particular probability principles were important and useful at certain solution steps. These statements were referred to and coded as
conceptual justifications. The production of conceptual justifications was dependent on first producing a procedural justification, as a participant could not explain why a principle was important without first identifying how that principle manifested in the worked examples. An example of a conceptual justification for the probability principle without replacement is: “You have to draw again, but the marble you picked before isn’t replaced so the sample size is reduced to five…but you still have two green marbles to pick, so the probability of picking any of the two marbles [on the second draw] is 2/5” (A detailed coding rubric for principle and justification statements can be located in Appendix D). Because each worked example focused on two principles of probability (e.g. order relevance and replacement) and the multiplication rule, participants could receive a maximum score of three correct procedural justifications and three correct conceptual justifications per worked example. Thus, across all six worked-examples a participant could receive a maximum of 18 correct procedural justifications and 18 correct conceptual justifications. The types of justification statements were taken as indications of procedural and conceptual understanding, respectively. Note that the procedural and conceptual justifications are similar to what Berthold and Renkl (2009) referred to as “principle-based self-explanations” and “rationale-based self-explanations,” respectively.

Fifteen think-aloud transcripts were randomly selected (5 from each learning condition) and coded separately by two coders. In coding for justification statement types, each of the six worked-example study items for each of the 15 selected transcripts were coded for how many procedural and conceptual justification statements were made by the participant, with a maximum possible of three correct procedural and three correct conceptual justifications per worked example. Therefore, a total of 270 coding opportunities were analyzed for reliability for both procedural and conceptual justifications. Inter-rater reliability was high for both procedural
justifications (Cohen’s Kappa = .87) and conceptual justifications (Cohen’s Kappa = .81).

Disagreements were resolved through discussion and mutual agreement of both raters. After an acceptable level of inter-rater agreement was reached, one coder coded the remaining think-aloud transcripts for procedural and conceptual justifications.

**Mappings**

Think-aloud protocols were also analyzed for what representations were mapped onto one another, and how often these mappings occurred. In each study item there were three distinct representations present: the word problem, the tree diagram worked-example, and the equation worked-example. An analysis of the think-alouds was conducted to examine what types of mappings—between the word problem and the diagram, the word problem and the equation, or the diagram and the equation—were occurring. Mapping can be defined as the process of relating an element from one representation to the corresponding element of another representation. For example, the following would be considered a mapping between the tree diagram and the equation: “Once the green [marble] has been picked, the highlighted ones [points to the diagram] show that there are two green left to be picked. Which is how they got 2/5 [pointing to 2/5 fraction in equation]” Both utterances and gestures made by the participants were used to make determinations of mappings between representations. A complete mapping rubric can be found in Appendix D.

Fifteen transcripts were randomly selected (5 from each learning condition) and coded for mappings separately by two coders. In coding for types of mappings, each of the six worked-example study items for each of the 15 selected participants were coded for whether the participant mentioned each of the three mapping types. In coding for mapping tokens, the number of times each mapping was mentioned in each mapping was indicated. Therefore, a total
of 270 coding opportunities were coded for reliability for both types and tokens of mappings. According to Landis and Koch (1977), Inter-rater reliability was at an acceptable level when coding for types of mappings (Cohen’s Kappa = .94) and tokens of mappings (Cohen’s Kappa = .72). Disagreements were resolved though mutual agreement of both raters. After the initial 15 transcripts were coded, one coder coded the remaining think-aloud transcripts for types and tokens of mappings.

**RESULTS**

The present study investigated how different learning conditions that paired prompted explanations with various techniques to help map information among representations facilitated conceptual and procedural understanding of mathematical concepts. Specifically, it was hypothesized that the combination of scaffolded self-generated mappings between representations and explaining those mappings (Generate + Explain) would be more beneficial to conceptual understanding than explaining mappings provided by the researcher (Provided + Explain), and more beneficial than engaging in explanation without specific instructions to focus on mappings between representations (Explain Only). Increased procedural understanding from pre-test to post-test were expected in all three conditions. Furthermore, we expected that if participants in the Generate + Explain condition showed a greater level of conceptual understanding at post-test, then they were also expected to produce more conceptual justification statements and engaged in more mappings between the diagrammatic and equation-based representations as they studied the worked-examples compared to those in the Provided + Explain and Explain Only conditions. In all analyses, Learning Condition represents the between-subjects variable.
**Conceptual Knowledge**

Conceptual knowledge was assessed at pre-test and again at post-test. An analysis of covariance (ANCOVA) with Learning Condition as the independent variable, post-test conceptual knowledge as the dependent variable, and pre-test conceptual knowledge as a covariate revealed a main effect of Learning Condition at post-test, $F(2, 71) = 3.70, p < .05$ (see Figure 3). Orthogonal contrasts revealed that the Generate + Explain condition showered higher conceptual knowledge at post-test than the Provided + Explain condition, $F(1, 71) = 6.47, p < .05$. However, the Provided + Explain and the Explain Only conditions did not differ in post-test conceptual knowledge, $F < 1$. As expected, the covariate (pre-test conceptual knowledge) was a significant predictor of conceptual knowledge at post-test, $F(1, 71) = 15.59$. Furthermore, the homogeneity of variance assumption was not violated, $F(2,69) = 1.82, ns$. That is, differences between the three learning conditions on post-test conceptual knowledge did not vary as a function of the covariate, pre-test conceptual knowledge. In sum, the Generate + Explain condition demonstrated higher conceptual knowledge at post-test than the Provided + Explain condition and the Explain Only condition. This result supports our hypothesis that the Generate + Explain condition would show higher post-test conceptual knowledge than the other learning conditions.

**Procedural Knowledge**

Procedural knowledge was assessed at pre-test and on three types of transfer items at post-test. The three types of items at post-test were: isomorphic transfer knowledge items (same questions encountered during study but with different numerical values), near transfer knowledge items (same principles encountered during study but entirely new questions), and far transfer knowledge items (new questions that blended two probability principles encountered during the
study phase). ANCOVA was utilized to assess the effects of learning condition on the three types of post-test procedural knowledge, with Learning Condition (Generate + Explain, Provided + Explain, Explain Only) as the independent measure, the three post-test item types (isomorphic transfer, near transfer, far transfer) as the repeated measure, performance on the three post-test procedural knowledge item types as the dependent variable, and pre-test procedural knowledge scores treated as a covariate. No differences emerged between the learning conditions on any post-test procedural knowledge measure, $F$’s < 1 (see Figure 4). As expected, the covariate (pre-test procedural knowledge) was a significant predictor of procedural transfer knowledge for isomorphic transfer items, $F(1, 71) = 15.81, p < .001$, near transfer items, $F(1, 71) = 38.99, p < .001$, far transfers items, $F(1, 71) = 11.43, p < .01$. Tukey Honestly Significant Different (HSD)
tests revealed that more isomorphic transfer items were answered correctly than near transfer items, and more near transfer items were answered correctly than far transfer items. Furthermore, the homogeneity of variance assumption was not violated for any ANCOVA analysis \( (F’s < 1 \) for isomorphic and near transfer items; \( F[2, 69] = 1.63, \text{ns} \), for far-transfer items). In sum, no differences between the learning conditions were observed on the various post-test procedural knowledge measures.

**Justification statements**

A 3(Learning Condition: Generate + Explain, Provided + Explain, Explain Only) X 2(Justification Type: Procedural, Conceptual) repeated measure ANOVA was performed with Learning Condition as the between-subjects variable, Justification Type as the within-subjects
variable, and the number of statements produced as the dependent variable. The assumption of sphericity was not violated. A main effect for Learning condition was not observed, $F(2, 72) = 1.20, ns$. There was, however, a main effect for Justification Type; participants made more procedural justifications than conceptual justifications, $F(1,72) = 306.58, p < .001, \eta^2_p = .81$. These results were qualified by a Learning Condition X Justification Type interaction, $F(2,72) = 6.27, p < .01, \eta^2_p = .15$ (see Figure 5).

Orthogonal contrasts on the interaction revealed that the Generate + Explain condition produced more conceptual justifications than the Provided + Explain condition, $F(1, 72) = 13.75, p < .001$, and that the Provided + Explain and Explain Only conditions did not differ in their total

![Figure 5. Mean number of justification statements by Learning Condition and Justification Type as evidenced by participants’ think-alouds during the study phase. Error bar represent standard error.](image-url)
number of conceptual justifications, $F(1, 72) = 2.62$, ns. Furthermore, Tukey HSD tests revealed that all three learning conditions produced the same number of procedural justifications.

As hypothesized, our results suggest that the Generate + Explain condition produced more conceptual justifications than the two other learning conditions. However, the conditions did not differ in the number of procedural justifications produced. It also should be noted that by examining the means found in Figure 5, the total number of procedural and conceptual justifications made by participants across learning conditions were relatively low in comparison to the maximum number (18) of each type of justifications possible. Therefore, although the Generate + Explain condition produced more conceptual justifications than the other conditions, the average number of conceptual justifications produced (approximately five) by the Generate + Explain condition was fairly low compared to the total number of possible conceptual justifications.

**Mappings**

The mappings participants made between the various representations during the study phase were analyzed using a types analysis and a token analysis. During the study phase, participants could make mappings between the word problem and the tree diagram, between the word problem and the worked equation, and between the tree diagram and the worked example. The criterion for determining whether a type of mapping was used was whether a participant used the particular mapping on each study problem. In other words, in the study phase a participant could be credited as using a particular mapping type once for each of the six study problems, for a maximum total score of 6 types per mapping. In contrast, the token analysis indexed the total number of times a particular mapping type was mentioned. Because both
analyses yield similar results, only the type analysis is presented here. The analysis of mapping tokens can be found in Appendix E.

A 3(Learning Condition: Generate + Explain, Provided + Explain, Explain Only) X 3(Mapping Category: Problem-Diagram, Problem-Equation, Diagram-Equation) repeated measures ANOVA was run with Learning Condition as the between-subjects variable and Mapping Category as the within-subjects variable. The assumption of sphericity was not violated. A main effect for Learning Condition was observed, $F(2, 72) = 5.76, p < .01, \eta^2_p = .14$. A main effect for Mapping Category was also observed, $F(2, 144) = 23.46, p < .001, \eta^2_p = .25$. These results were qualified by a Learning Condition X Mapping Category interaction, $F(4, 144) = 5.03, p < .01, \eta^2_p = .12$ (see Figure 6).

Following up the main effect for Learning Condition, orthogonal contrasts revealed that, although the Generate + Explain condition did not produce more overall types of mappings than the Provided + Explain condition, $F(1, 72) = 1.72, ns$, the Generate + Explain condition did produce more types of mappings than the Explain Only condition, $F(1, 72) = 11.26, p < .01$.

Following up the main effect for mapping category, orthogonal contrasts revealed that participants produced more Problem-Equation mapping types than Problem-Diagram mapping types, $F(1, 144) = 19.96, p < .001$. Additionally, participants produced more Problem-Diagram mapping types than Diagram-Equation mapping types, $F(1, 144) = 5.25, p < .05$.

Orthogonal contrasts within the interaction focused on comparing differences in the number of diagram-equation mapping types across the three learning conditions, as this was of main interest to the study. Because no specific hypotheses were made about the Problem-Diagram mapping types and Problem-Equation mapping types in the three learning conditions,
Figure 6. Number of mapping types by Learning Condition and Mapping Category. Error bars represent standard error. Note: for each mapping category, a maximum score of 6 would indicate that participants mentioned mapping between the two representations on all six worked-examples.

these analyses were followed up post-hoc with Tukey HSD tests. Participants in the Generate + Explain condition did not use more diagram-equation mapping types than participants in the Provided + Explain condition, $F(1, 144) = 2.64, \ ns$. However, the Generate + Explain and Provided + Explain conditions both produced more diagram-equation mapping types than the Explain Only condition, $F(1, 144) = 13.55, p < .001; F(1, 144) = 4.18, p < .05$, respectively. Tukey HSD tests showed that the Generate + Explain and Provided + Explain also produced more problem-diagram mapping types than the Explain Only condition, however, the three conditions did not differ with respect to problem-equation mapping types. Therefore, it appears
that presence of highlighting allowed for more mappings between the word problem and the worked tree diagram, but not between the word problem and the worked equation.

**Relationship Between Knowledge Outcomes and Think-Aloud Measures**

Think-aloud data were also analyzed to understand how mappings between representations and justification types predicted performance on the various post-test measures. Of particular interest was how conceptual justifications and diagram-equation mappings influenced one another and how they related to post-test conceptual knowledge. Prior research by Berthold, et al. (2009) suggests that we should observe a positive relationship between conceptual justifications and conceptual knowledge as measured at post-test. That is, an increase in the number of conceptual justifications produced should positively correlate with post-test conceptual knowledge. Additional research by Ainsworth and Loizou (2003) and Schwonke, et al. (2009) suggests that those learners who integrate information between the multiple worked examples by mapping common elements between the representations should also demonstrate higher conceptual knowledge at post-test. Therefore, we would also expect mappings between the diagrammatic worked example and the equation worked example to be positively correlated with post-test conceptual knowledge. Correlations among the various dependent measures from the present study lend support to both of the conclusions of the prior research. As shown in Table 1, the number of conceptual justifications and diagram-equation mappings were each positively correlated with post-test conceptual knowledge and with each other (relationships between other variables are explored further in Appendix F). Given this pattern of correlations, it is possible that a mediation relationship exists between these three variables. A temporal order to the variables can also be established based on the experimental instructions. First participants were
## TABLE I

**CORRELATIONS BETWEEN MAPPING TYPES, JUSTIFICATION STATEMENTS, AND POST-TEST PERFORMANCE**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Problem-Diagram Mapping Type</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Problem-Equation Mapping Type</td>
<td>-.49**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Diagram-Equation Mapping Type</td>
<td>.73**</td>
<td>-.31*</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Procedural Justifications</td>
<td>.10</td>
<td>.29*</td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Conceptual Justifications</td>
<td>.25*</td>
<td>.14</td>
<td>.25*</td>
<td>.76**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Post-test Isomorphic Transfer Knowledge</td>
<td>-.20</td>
<td>.24*</td>
<td>-.15</td>
<td>.30**</td>
<td>.29*</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Post-test Near Transfer Knowledge</td>
<td>-.07</td>
<td>.22</td>
<td>-.03</td>
<td>.21</td>
<td>.21</td>
<td>.61**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>8. Post-test Far Transfer Knowledge</td>
<td>-.13</td>
<td>.23*</td>
<td>-.01</td>
<td>.32**</td>
<td>.28*</td>
<td>.27*</td>
<td>.38**</td>
<td>-</td>
</tr>
<tr>
<td>9. Post-test Conceptual Knowledge</td>
<td>.16</td>
<td>-.08</td>
<td>.24*</td>
<td>.25*</td>
<td>.44**</td>
<td>.32**</td>
<td>.21</td>
<td>.31**</td>
</tr>
</tbody>
</table>
instructed to generate or examine a provided mapping and then engage in explaining that mapping, so temporally a mapping needs to be created or examined before it can be explained.\(^2\)

A meditational analysis would allow for the examination of the relationship between the three variables of interest. A multiple regression approach would allow for the examination of the unique and shared effects conceptual justifications and diagram-equation mappings have on post-test conceptual performance, but would not help clarify the effects they may have on one another.\(^3\) Therefore, it seems appropriate to examine the relationship between these variables through a mediation analysis with diagram-equation mappings entered as the more distal variable, conceptual justifications entered as the more proximal mediating variable, and post-test conceptual knowledge entered as the outcome variable.

To test whether linkages between post-test conceptual knowledge and diagram-equation mappings might be mediated by differences in the number of conceptual justifications produced, we analyzed whether higher use of diagram-equation mapping types were associated with higher use of conceptual justification, that in turn, were associated with higher post-test conceptual knowledge. This hypothesis was tested using the Preacher and Hayes (2008) bootstrapping method for testing direct and indirect effects. Results are based on 1000 bootstrap samples. The total mediation model accounted for significant variance in post-test conceptual knowledge, \(R^2 = .23, F(2, 72) = 9.88, p < .001\). As shown in Figure 7, higher use of the diagram-equation mapping type was associated with higher use of conceptual justifications. Additionally, higher

\(^2\) Examination of a random sampling of transcripts revealed that explanations always came after the mention of a mapping. However, even if every participant did not adhere to this temporal order, causality does not have to be strictly followed for this type of mediation. Of most importance is that there is a more distal, weaker association between a predictor variable and the outcome variable and a stronger association between another more proximal predictor variable and the outcome variable.

\(^3\) A multivariate regression analysis was also conducted and revealed similar findings as the mediation analysis, namely that conceptual justifications accounted for more variance in post-test conceptual knowledge measures than the mappings learners made between the representations. Therefore, only the mediation analysis will be discussed further.
use of conceptual justifications was associated with higher post-test conceptual knowledge. Finally, the analysis revealed, with 95% confidence, the direct effect of diagram-equation mapping types on post-test conceptual knowledge was mediated by conceptual justifications, with a Bias Corrected and Accelerated (BCA) bootstrap confidence interval of .01 to .56. In summary, the number of conceptual justification can account for the pathway between diagram-equation mapping types and post-test conceptual knowledge.4

Secondary analyses investigated possible meditational relationships between mappings, procedural justifications, and post-test procedural knowledge. No relationship was observed between diagram-equation mappings, procedural justifications, and any of the post-test procedural measures. Further meditational relationships between other types of mappings (e.g.

---

4 Testing for a possible moderated mediation (or mediated moderation), with Learning Condition as the moderating variable, using the process outlined by Preacher, Rucker, and Hayes (2007), failed to reveal any significant results for a moderating affect between diagram-equation mapping types and conceptual justifications (i.e. Model 2), between conceptual justifications and conceptual post-test performance (i.e. Model 3), or between both pathways (i.e. Model 5).
between the word problem and the equation), procedural justifications, and post-test procedural measures are fully investigated in Appendix F.

**DISCUSSION**

Regardless of learning condition, all learners showed higher levels of conceptual and procedural knowledge at post-test compared to pre-test. However, based on the results from the conceptual knowledge analysis, the Generate + Explain condition demonstrated higher conceptual knowledge at post-test than the other learning conditions. Also, based on the analysis of the justification statements, the Generate + Explain condition produced more conceptual justifications than the two other conditions. This supports our hypothesis that providing a scaffolded prompt that asked learners to generate mappings allows for greater conceptual understanding of the to-be-learned material compared to providing learners with complete mappings or with no mappings and asking them to explain. As expected, no differences were observed between the three conditions for procedural understanding.

Scaffolding learners to generate mappings between representations provided conceptual learning benefits. However, the benefits were somewhat small—an approximate 10% increase in conceptual knowledge scores at post-test compared to the other two learning condition—and low compared to the maximum score possible (40% of conceptual items answered correctly at post-test). Therefore, there is room for improvement in both the relative and absolute benefits of generating mappings between representations. Perhaps improvement could be achieved through requiring the mapping process to be even more generative (for example, by explicitly requiring learners to map each highlighted mapping in the diagram to some element in the equation), providing additional training on how to generate mappings between representations, or providing learners with feedback about the appropriateness of their mappings.
With regard to how the different learning conditions mapped between the various representations, participants in the Generate + Explain produced more diagram-equation mappings than the Explain Only condition but not significantly more diagram-equation mappings than the Generate + Provided Condition. We hypothesized that if the Generate + Explain condition demonstrated higher conceptual knowledge at post-test compared to the other two conditions this would lead to learners producing more mappings between the diagram and the equation. However, both the Generate + Explain and Provided + Explain conditions produced more diagram-equation mappings than the Explain only condition. Thus, it may be the case that any instruction that draws attention to mapped similarities and differences between the diagrammatic worked-example and the equation-based worked-example may be beneficial to producing mappings between the two representations.

However, the role of mapping between representations may be less important to acquiring conceptual knowledge than previously thought. When we examined how diagram-equation mappings and conceptual justifications predicted post-test conceptual knowledge, our mediation model suggests that the number of conceptual justifications could account for any significant role that diagram-equation mappings has on post-test conceptual knowledge. That is, even though mappings between the two types of worked examples were directly predictive of conceptual knowledge on post-test, the number of conceptual justifications produced by learners mediated these effects. Conceptual justifications accounted for the pathway between the diagram and equation mappings and post-test conceptual knowledge. Thus, the number of conceptual justifications generated by the learner appears to be an important mediating factor in explaining the acquisition of conceptual knowledge when mapping between representations.
There are some limitations to this study. First, the learning environment of this study may not necessarily reflect how learners usually learn about complex probability events in a typical educational setting. The study session lasted approximately one hour, with participants studying the multiple representations of the worked-examples for approximately ten minutes. This may be far less time and in a different situation than how students typically learn about these concepts in a mathematics classroom. Yet, given the short time frame, participants in the study still demonstrated significant improvement from pre-test to post-test measures, albeit still well below completely accurate performance.

Second, it is unclear whether the positive benefits to conceptual understanding for the Generate + Explain condition were due to the combined effects of scaffolded mapping generation along with explanation or simply due to the scaffolded mapping generation. That is, a “Generate Only” condition without explanation might produce similar results. We believe this is unlikely based on the previously documented benefits of explanation (e.g. Berthold & Renkl, 2009) and findings from the present study regarding the combined benefits of generating mappings and explanations and for conceptual knowledge understanding. Without the requirement to explain and justify their mappings, learners will likely engage only in surface level mapping and thus no benefit to conceptual understanding.

Finally, the examination of mappings between the representations was conducted based on verbal think-aloud responses and videotaped observations of the participants’ gestures. Perhaps a more accurate method to ascertain mappings would be through using eye-tracking technology to monitor eye movements while participants think aloud as they make mappings between elements in the two types of worked-examples during the study phase. Eye tracking in conjunction with think-aloud protocols could allow for a more accurate account of exactly where
in the representations participants are looking as they make mappings, how often they move between representations, and how certain justification statements are associated with the different types of mappings. Future studies could address these limitations.

Our findings provide further insight into how to improve learning for low-knowledge individuals. First, because low-knowledge learners do not often engage in mapping non-surface level features (e.g. principles of probability in this study) between representations, a generative mapping process that requires low-knowledge learners to actively make connections and explain the importance of those connections may be important in improving understanding of conceptual knowledge for these learners. Second, low-knowledge learner's typically poor conceptual understanding when learning through multiple representations can be improved by having them explain and justify how the various conceptual elements represented relate to one another by having them map between representations. Requiring learners to map and make connections between representations compels them to attend to both representations rather than focusing on just one of the representations. Finally, having low-knowledge learners explain the connections they are making between the representations may assist them in better understanding the importance of the connections they are generating. Our results support the notion that assisting low-knowledge learners in becoming more aware of the functional benefits multiple representations can provide—through requiring them to map elements between representations and explaining the importance of those mappings—is beneficial to the acquisition of conceptual knowledge.
REFERENCES


APPENDIX

APPENDIX A

Procedural and Conceptual Knowledge Items
The follow are the procedural and conceptual items that participants encountered during the pre-test, the worked-example study phase, and the post-test. Pre-test procedural items and near-transfer post-test items were counterbalanced to control for any differences in difficulty between the problem sets. The worked-example study items and the isomorphic-transfer items were also counterbalanced to control for any differences in difficulty between the two item sets. What follows is a list of the items contained in the different item sets. Following the item sets in Tables II, III, IV, and V are the conceptual scoring rubrics that details how conceptual pre-test and post-test items were scored.

Basic Probability Knowledge Questions
You are playing a game with a regular six-sided die. If you roll a 4 you win. What is the probability that you roll a 4? Answer: 1/6

A card is selected at random from a standard deck of 52 cards. Assuming all cards are equally likely to be selected, what is the probability of picking an ace (there are 4 aces in a deck)? Answer: 4/52

A coin is tossed in the air. It lands on tails. What is the probability that the next toss will land on tails? Answer: 1/2

Procedural Pre-test/Near-Transfer Items
You and your friend are running a race with 4 other people. A computer randomly determines starting positions. What is the probability that you will be placed in the first starting position and your friend in the second starting position? Answer: 1/30

In your Introduction to Psychology Class, the instructor decides to randomly draw names to determine the order of student presentations. The first name draw will present first, the second name draw will present second, and so on until all the names have been selected. There are only 12 students in your class, which includes yourself and your friend. Both you and your friend want to be the first to present. What is the probability that you or your friend will be selected to present first and the remaining person selected to present second? Answer: 2/132

Rebecca has 7 coats, each a different color. She is always running late and will quickly grab a coat without looking as she leaves her home. What is the probability that she wears her red coat on Monday and her white coat on Tuesday? Answer: 1/49

You flip a quarter 8 times. What is the probability that the quarter lands on heads on the first four flips and on tails the last four flips? Answer: 1/256
There are 7 players on a basketball team. Only 5 players may be on the court at one time and any player can play any position. If the coach randomly selected 5 players to be on the court, what is the probability she would select the 5 tallest players? *Answer: 1/21*

For History class, Hank must answer a combination of 3 essay questions from a list of 5 questions. Hank only knows how to answer 3 of the questions. If Hank randomly selects a combination of 3 questions, what is the probability that he selects the 3 questions he knows (the order that he picks the questions does not matter)? *Answer: 1/10*

"The Mathletes," a newly formed club at UIC consists of only four members: Albert, Betty, Carl, and David. The club decides to randomly draw names out of a hat to determine who will be president, vice president, and treasurer. President is drawn first, followed by vice president, and then treasurer. What is the probability that Albert will be picked to be president, Betty as vice president, and Carl as treasurer? *Answer: 1/24*

A museum curator randomly hung 4 paintings in a row. What is the probability that she arranged the paintings in order from oldest to newest? *Answer: 1/24*

A card is drawn from a deck of 13 cards numbered 1-13. After the card is drawn it is replaced in the deck. The deck is shuffled and a second card is then drawn and replaced. What is the probability that an even card is drawn each time? *Answer: 36/169*

Allison's carpool randomly draws names from a hat to determine who will drive the carpool each week to work. After a name is picked, it is placed back in the hat, so it is possible for someone to drive more than one week in a row. There are 5 people total in her carpool. What is the probability that Allison will have to drive the carpool for two weeks in a row? *Answer: 1/25*

After eating most of his Halloween candy, Johnny has 6 pieces of candy left: a tootsie roll, a sucker, a chocolate bar, sweet-tarts, taffy, and milkduds. Johnny closes his eyes and randomly picks out two pieces of candy at the same time. What is the probability that he picks the chocolate bar and the taffy? *Answer: 1/15*

Your friend chooses a combination of two different numbers between 1 and 5 (1 and 5 are included). You try to guess the combination. What is the probability that you guess the combination correctly on the first attempt? *Answer: 2/20*

**Procedural Worked-Example/Isomorphic-Transfer Items**

Hillary and Megan, along with 4 other people enter a charity raffle. The first prize drawn is a pair of socks, and the second prize drawn is a t-shirt. Hillary really hopes to win the socks and Megan wants the t-shirt. Each person can only win one prize. What is the probability that Hillary wins the first prize drawn and Megan wins the second prize drawn? *Answer: 1/30*

Hillary and Megan, along with 3 other people enter a charity raffle. The first prize drawn is a movie ticket, and the second prize drawn is a giftcard. Hillary really hopes to win the movie ticket and Megan wants the giftcard. Each person can only win one prize. What is the probability that Hillary wins the first prize drawn and Megan wins the second prize drawn? *Answer: 1/20*
There are 7 slips of paper in a bag numbered 1 to 7. Two slips are selected at random, one at a time. After a slip is selected it is not replaced back in the bag. What is the probability 5 or 7 will be selected first and 4 picked second? Answer: 2/42

There are 8 slips of paper in a bag numbered 1 to 8. Two slips are selected at random, one at a time. After a slip is selected it is not replaced back in the bag. What is the probability 2 or 3 will be selected first and 7 picked second? Answer: 2/56

At the carnival you spin a wheel of fortune twice. The wheel is divided into 6 colored segments. Of the 6 segments, one is colored gold and one is colored silver. You win if your spin lands on the gold or the silver segment. What is the probability that you spin the wheel twice and win both times? Answer: 4/36

At the carnival you spin a wheel of fortune twice. The wheel is divided into 5 colored segments. Of the 5 segments, one is colored gold and one is colored silver. You win if your spin lands on the gold or the silver segment. What is the probability that you spin the wheel twice and win both times? Answer: 4/25

There are 5 different colored balls in a bag: red, blue, yellow, pink, and green. Two balls are selected at random, one at a time. After a ball is selected it is place back in the bag. What is the probability that the green ball will be picked first and the red ball picked second? Answer: 1/25

There are 6 different colored balls in a bag: orange, blue, yellow, red, purple, and green. Two balls are selected at random, one at a time. After a ball is selected it is place back in the bag. What is the probability that the purple ball will be picked first and the orange ball picked second? Answer: 1/36

A basket contains 1 red, 2 black, and 3 green marbles. Two marbles are drawn at once. Find the probability that the sample contains 2 green marbles. Answer: 6/30

A basket contains 2 black, 2 red, and 3 green marbles. Two marbles are drawn at once. Find the probability that the sample contains 2 green marbles. Answer: 6/42

You and your friend take part in a two-day mountain bike course. Each day the instructor brings along 8 helmets (orange, silver, brown, red, green, yellow, purple, and white). The helmets are handed out randomly and given back to the instructor at the end of the day. What is the probability that you and your friend get the red and green helmet on the first day of the course (it does not matter who gets which color)? Answer: 2/56

You and your friend take part in a two-day mountain bike course. Each day the instructor brings along 5 helmets (orange, silver, brown, red, green). The helmets are handed out randomly and given back to the instructor at the end of the day. What is the probability that you and your friend get the red and green helmet on the first day of the course (it does not matter who gets which color)? Answer: 2/20
Procedural Far Transfer Items
Six Teams (A, B, C, D, E, F) are competing in an ice sculpting contest. The winning team receives $1000, second place receives $500, and third place receives $250. Teams finishing fourth and fifth get consolation prizes. You make a bet with your friend that Team A wins $1000, Team C wins $500, Team E wins $250, and Teams B and D receive consolation prizes. What is the probability you win your bet? Answer: 1/360

Katie has three pairs of shoes: sneakers, heels, and boots. On most days it doesn't matter which pair she wears, but today it is raining and she should really wear boots. She randomly picks a shoe from her closet. Then she randomly selects two more shoes from her closet at the same time. What is the probability she selected a boot on the first draw and other other boot on the second draw? Answer: 8/60

There are 5 different colored marbles in a bag: orange, blue, green, red, and yellow. Suzy grabs one marble from the bag, looks at it, and then places it back in the bag. She reaches into the bag again, grabs a marble, looks at it, and then places it back in the bag. Finally, she grabs two marbles at once, looks at them and returns them to the bag. Without knowing what marbles Suzy actually picked, what is the probability she grabbed the blue marble on all 3 pulls? Answer: 2/125

Toby the dog has 4 cans of dog food remaining. Toby's owner is in a hurry and, without looking, randomly grabs a can of dog food. He realizes the can is dented so he puts it back in the cabinet. Without looking again, he randomly grabs a can of dog food again. This one is also dented so he puts in back in the cabinet. Fed up because he keeps grabbing dented cans, he finally randomly grabs two cans of dog food at the same time from the cabinet, also without looking. What is the probability that on all 3 occasions Toby's owner grabbed the same dented can of dog food? Answer: 3/96

At a school raffle, there are two smaller prizes and one grand prize. When selecting winners for the smaller prizes, names are drawn at random and not replaced, so that each person can only win one smaller prize. However, for the grand prize drawing anyone who entered the raffle is eligible, even those who won a smaller prize. 10 students total enter the raffle. What is the probability Joe wins the first small prize drawing, Sarah wins the second small prize drawing, and either of them wins the grand prize? Answer: 2/900

John's fantasy football league uses a computer program to randomly determine the order in which each team gets to pick players. There are a total of 4 teams in the league. Each team gets one pick in Round 1 and one pick in Round 2. However, in Round 3, each team gets two picks. What is the probability that John's team gets the first picks in Rounds 1 and 2, and also the first two picks in Round 3? Answer: 2/896

Conceptual Pre-test/Post-test Items
In some situations, the order in which things are selected matters. For example, when selecting the order of class presentations. In other situations, for example, in the state lottery, the order in which the numbers are selected is not important. Based on what you studied, which event has a
greater probability of occurring, an event where order matters, or one where it does not? Explain your answer.

In some situations, the order in which things are selected matters. For example, when picking names to determine the starting order of a race. In other situations, for example, when selecting toppings for a pizza, the order in which toppings are selected is not important. Based on what you studied, which has a greater probability of occurring: a set of events where the order objects are selected matters, or a set of events where the order objects are selected does not matter? Explain your answer.

In certain situations when determining the probability of an event, an item can be selected more than once. For example, when deciding what pair of shoes to wear each day, the same pair of shoes can be worn more than once. In other situations an item cannot be selected more than once. For example, when determining the seating arrangement at a party, once a seat is assigned to a guest, that same seat cannot then be assigned to another guest. Generally, which has a greater probability of occurring: two events where items can be selected more than once, or two events where items can be selected only once? Explain your answer.

In certain situations when determining the probability of an event, an item can be selected more than once. For example, when selecting songs to play, the same song can be played twice in a row. In other situations an item cannot be selected more than once. For example, on a baseball team if player if assigned to be the pitcher, he can not also be assigned to play another position at the same time. Generally, which has a greater probability of occurring: two events where items can be selected more than once, or two events where items can be selected only once? Explain your answer.

Your friend flips a coin in the air and it lands on heads 10 times in a row. Your friend says that the coin is due to land on tails on the next flip because it has already landed on heads 10 times in a row. Is the coin more likely, less likely, or as likely to land on tails on the next flip compared to landing on heads? Explain your answer.

Roulette is a popular casino game. In roulette, a ball is rolled around a wheel containing many different numbers and players bet on which number they think the ball will land on. Often casinos will display the last 20 winning numbers on an electronic board. A player wants to place a bet on her favorite number, 8. But she sees that 8 is one of the 20 most recent winning numbers. If she places her bet on 8, is her probability of winning better, worse, or the same than if she placed her bet on another number? Explain your answer.

In order to calculate the probability of two or more events occurring, the probabilities of each event are multiplied, rather than added together. Why do you think this is? [Question appears in both pre-test and the post-test]
### Conceptual Items Coding Rubric

#### TABLE II
**SCORES AND CRITERIA FOR INDEPENDENT EVENTS / MUTUAL EXCLUSIVITY**

<table>
<thead>
<tr>
<th>Score</th>
<th>Criteria</th>
</tr>
</thead>
</table>
| 0     | o Does not answer the question  
|       | o Does not provide an explanation that mentions any of the criteria for 0.5 or 1 |
| .5    | Mentions **ONE** of the following:  
|       | o Each flip/roll is an **independent event** (previous events have no bearing on the current event)  
|       | o **Probability remains constant** on the next flip/roll (1/2 or 1/20 depending on the question) |
| 1     | Mentions **TWO** of the following:  
|       | o Each flip/roll is an **independent event** (previous events have no bearing on the current event)  
|       | o **Probability remains constant/equal chance** on the next flip/roll (1/2 or 1/20 depending on the question) |

#### TABLE III
**SCORES AND CRITERIA FOR ORDER RELEVANCE VS. IRRELEVANCE**

<table>
<thead>
<tr>
<th>Score</th>
<th>Criteria</th>
</tr>
</thead>
</table>
| 0     | o Does not answer the question  
|       | o Does not provide an explanation that mentions any of the criteria for 0.5 or 1  
|       | o Explanation consists of “because [choice] has more possibilities/higher probability” |
| .5    | Mentions **ONE** of the following:  
|       | o Order irrelevance allows for set of events to occur in two different ways (e.g. AB or BA). Conversely, order relevance only allows a set of events to occur in one manner (e.g. AB)  
|       | o In both order relevance and order irrelevance, the total number of outcomes remains the same, but order irrelevance increases how many outcomes are desirable (e.g. 2/30 vs. 1/30), or vise versa |
| 1     | Mentions **TWO** of the following:  
|       | o Order irrelevance allows for set of events to occur in two different ways (e.g. AB or BA). Conversely, order relevance only allows a set of events to occur in one manner (e.g. AB)  
|       | o In both order relevance and order irrelevance, the total number of outcomes remains the same, but order irrelevance increases how many outcomes are desirable (e.g. 2/30 vs. 1/30), or vise versa |
### TABLE IV
SCORES AND CRITERIA FOR WITH REPLACEMENT VS. WITHOUT REPLACEMENT

<table>
<thead>
<tr>
<th>Score</th>
<th>Criteria</th>
</tr>
</thead>
</table>
| 0     | o Does not answer the question  
        o Does not provide an explanation that mentions any of the criteria for 0.5 or 1  
        o Explanation consists of “because [choice] has more possibilities/higher probability” |
| 0.5   | Mentions **ONE** of the following:  
        o Replacement allows for choice to be picked again/no replacement allows for undesired object to be removed  
        o **Either**: Because placing a desired second-selection outcome back in the sample space increases the probably of drawing the same outcome a second time compared to no replacement. (e.g. picking two Aces from a deck of cards)  
        o **OR**: Because placing a non-desired second-selection outcome back in the sample space decreases the probably of drawing a desired outcome the second time compare to with replacement (e.g. picking an Ace then a King from a deck of cards) |
| 1     | Mentions **TWO** of the following:  
        o Replacement allows for choice to be picked again/no replacement allows for undesired object to be removed  
        o **Either**: Because placing a desired second-selection outcome back in the sample space increases the probably of drawing the same outcome a second time compared to no replacement. (e.g. picking two Aces from a deck of cards)  
        o **OR**: Because placing a non-desired second-selection outcome back in the sample space decreases the probably of drawing a desired outcome the second time compare to with replacement (e.g. picking an Ace then a King from a deck of cards) event to be higher than if the desired outcome pulled in event one was left out. |
TABLE V
SCORES AND CRITERIA FOR MULTIPLICATION RULE

<table>
<thead>
<tr>
<th>Score</th>
<th>Criteria</th>
</tr>
</thead>
</table>
| 0     | o Does not answer the question  
o Does not provide an explanation that mentions any of the criteria for 0.5 or 1 |
| 0.5   | o Mentions ONE of the following:  
o Probability fractions need to be multiplied when calculating the probability of Event A AND Event B both occurring to get the total number of outcomes  
o Mentions that probability fractions of two separate (single) events need to be multiplied when calculating the probability of Event A AND Event B both occurring |
| 1     | o Mentions TWO of the following  
o Probability fractions need to be multiplied when calculating the probability of Event A AND Event B both occurring to get the total number of outcomes  
o Mentions that probability fractions of two separate (single) events need to be multiplied when calculating the probability of Event A AND Event B both occurring |
APPENDIX B

Experimental Instructions for Running Participants

General Notes:
- Computer 3 is the station for the participant to sit. Computer 2 is for the researcher to record the study session.
- If you need to exit the experiment before it is completed, press Alt + Ctrl + Shift. This will abort the experiment. But it will also not record any participant data, so only do this as a last resort (e.g. the computer freezes).
- If at any time the participants look confused or unsure what to do, ask them if you can help clarify something for them.
- Sometimes the participants may ask questions about the problems they are studying/solving. If they are unclear about a word in the problem, you may help them (e.g. if they don’t know what “die” means you can tell them that it is what you call one dice), but do not provide help about understanding or solving the problem.
- Participants are not allowed to use a calculator.

Before participant arrives:
- Load the experiment on Computer 1
  - Start ➔ All Programs ➔ E-Prime 2.0 ➔ E-Run
    - Note: make sure you run eprime 2.0 and not eprime. The experiment will not run on e-prime 1.1
  - Select the appropriate condition from the dropdown box. If you don’t see the condition you need in the box, click the folder to open it. The experimental condition files are located in C:\DietzMasters.
  - Click “Run” and enter participant’s PIN (don’t click ok yet)
- Set up Audio/Video Recording on Computer 2
  - Launch “CaptureFlux” (program should be on the desktop)
  - Under “Sources” tab (press F2 if you can’t find it) make sure the Video Source is set to USB Video Device and Audio Source is set to SoundMAX Digital Audio AND Microphone.
    - Make sure “Capture format” is set to Wmv
  - In the “Directories” tab (press Alt+D to get to), under Video Files, and make sure the file save location is the Desktop (C:\Documents and Settings\colflesh\Desktop) (Really, you can save the file to wherever you want, but I find it easiest to save the file to the desktop)
    - Note: Don’t try to record the video directly to the external hard drive. The computer is old and trying to record directly to the hard drive causes the video and audio to become out of sync.
  - Name the video file using the following format: PIN_Condition_Date_Researcher
    - For example, a participant with the PIN 9999, in Condition 3, run on January 3, 2011 by Kevin, would be entered as: 9999_3_01032011_KD.wmv

When participant arrives:
1. Have participant read and sign consent form
2. Start audio/video recording by pressing the Capture button in CaptureFlux.

3. Provide the participant with a few sheets blank paper and a pencil.

4. Once the participant is ready, have him/her click “OK” twice on the computer to start the experiment.

5. After the participant completes the pre-test, there are condition-specific instructions that appear. The last line of the instructions states to “Tell the researcher when you are ready to continue”. This is to ensure that the participant understands the instructions and knows what s/he is expected to do. At this point, answer any questions the participant has about the instructions. When the participant is clear about what to do, have them press SPACEBAR to continue (Note: this is the only time they will be asked to press spacebar instead of enter to advance the study. This is to ensure that they don’t click through the instructions without fully understanding them). Different conditions ask participants to do different things, so make sure you know what is asked of the participants in different conditions. In short:
   - **Condition 1 (a & b):** make connections between the diagram and equation and explain those connections. If not making connections, ask, “**How do the highlight parts relate to the parts in the tree diagram/equation?**”. If not explaining connections, ask, “**Explain how those parts relate to each other.**”.
   - **Condition 2 (a & b):** explain already provided connections. If not explaining, ask, “**Explain how those parts relate to each other.**”
   - **Condition 3 (a & b):** explain why it is important to multiply. “**Explain why you have to multiply.**”

6. As the participant studies the worked examples, make sure s/he is thinking aloud. All conditions require that the participant think aloud. If s/he is not talking about what they are thinking about, prompt them with questions such as, “**What are you thinking about?**” and “**Keep talking**”.

7. After the worked examples, participants will be prompted to solve a series of post-test questions. They do not have to think aloud during these.

8. When the participant is finished with the experimenter on the computer, have him/her fill out a demographic questionnaire. Afterwards, the participant is free to go. You should offer them a debriefing form (found on the shelf). You may also be asked to sign their PEC form.

After the participant leaves:
- Press ENTER on Computer 1 to end the experiment (if the participant as not already done so)
- Fill out the Study Log
- If the participant used scratch paper, collect the scratch paper, write the participant’s PIN and condition number and date (e.g. 9999_3b_01032011) on the top right corner, and place it in the pile scratch paper pile in the filing cabinet
- Stop the video/audio recording
- Use the headphones or speakers to ensure that video and audio was recorded for the session.
  - If video and/or audio was not recorded, make a note of it in the Study Log
- Move the video file from the desktop to the external hard drive (Drive G:\)
  - If you don’t see G:\ make sure the hard drive is turned on. Also, make sure to turn off the hard drive when you are finished moving the file.
  - You may delete the file from the desktop after it is move to the external hard drive
• Assign credit on PECOLSUS
  o Go to https://www.psych.uic.edu/pecolsus/stu_login.asp
  o Enter “rrr” for NetID and password (without quotes). This takes you to the researcher log-in page
  o Enter Kevin’s NetID and PECOLSUS password
  o Click on the icon that looks like a sheet a paper to enter the Session List Page
  o Click on the “View Students” link next to the appropriate study session.
  o Under “Record PECs” select “Attended” then it submit.
  o Note: If a participant is a no-show, do not immediately mark them as “No-Show”. Sometimes they are running a minute or two late (which is fine, but we will not run them if they are over 10 minutes late) or have a legitimate excuse. If a participant doesn’t show or is more than 10 minutes late, let Kevin know and he will mark them as a “no show” a few days after the experimental session.
• Update the participant session list.
  o Note anything that was not “normal” with the session (e.g. participant clicked through the experiment and didn’t solve the problems, finished the study in less than 30 minutes, fell asleep, limited English abilities, did not think aloud).

**Experimenter Instructions for Transcribing Think-Aloud Protocols**

Make sure that the video as not already been transcribed by looking at the transcribing sheet. If the video has not been transcribed, add the video to the sheet and put your name under the “Transcribed By” column.

Open the transcription template on the external hard drive
(G:\transcripts\transcription_template.xls)

Use “Save as…” to create a new file for the video you are working on. Save the file as the same name as the video file, but add your initials to the end of the file name
  • E.g. if the video file is named 1099_2b_04142011_KD.wmv, name the transcription of the video as 1099_2b_04142011_KD_[your initials].xls. Save the file to the “transcripts” folder on the external hard drive.

When time stamping, using the following format: hh:mm:ss
  • E.g. if you want to timestamp an event that occurs 7 minutes and 42 seconds into the video, you would write 00:07:42 for the timestamp

Timestamp the following events:
  • When the instructions appear on the screen (label this “I” under Item)
    o Put [Start] in the utterance column
  • When the participant begins to study each problem (labeled 1, 2, 3, 4, 5, or 6 under the “Item” column)
  • When the participant finishes studying the last problem (i.e. when the next set of instructions appears on the screen. Label this “E” for End under the “Item” column)
    o Put [End] in the utterance column

The Speaker column refers to whether the Subject (“S”) or the Researcher (“R”) is speaking
Under the utterance column, transcribe everything that the speaker is saying.

- Put each new thought on a separate row. Essentially, a thought is usually what a person can say in one breath. Usually a new thought is denoted by a change in subject or a long pause.
- Sometimes you will hear the researcher say things like “mmhmm” or “OK”. These types of statements can be placed in parentheses within what the subject is saying. Utterances made by the researcher only need to be placed on a separate line if they are separate thoughts from what the subject is saying (e.g. if the researcher is answering a question).
  - Example: Subject: So the diagram is more confusing than the equation (OK) because...
- Include hesitations, such as “umm” and “like”
- Use two periods (..) to denote a pause that lasts between 1/2 and 1 second that occurs within the same thought
- Use an elapse (…) to denote a pause longer than 1 second that occurs in the same thought
  - Example: So this piece relates to these numbers because…umm…well I don’t really know.
- After listing to a segment of video several times, if you still cannot determine what the speaker is saying, write [inaudible] in place of the inaudible passage
  - E.g. I think the blue part is [inaudible] and the yellow part is the 2 in the equation
- The subject reads aloud (or at least are supposed to) the word problems before every worked example. If a subject reads the problem all the way through without saying anything but what is written in the problem (umms and other pause are OK), you can simply write [Reads problem] in the Utterance column. If the subject stops reading in the middle of the problem (e.g., to ask a question, see the problem relates to the diagram), write out all of what the subject says while reading the problem, including the problem itself. The same guidelines apply for gestures as well.

Use the gesture column to record gestures the subject makes while studying the problems

- Often subjects will say things like “this is the same as that.” Looking at just a transcript it is difficult, if not impossible to determine what they are referring to. This is why it is important to also record gestures made by the participants.
- Place the gesture made on the same row as the utterance referring to a gesture was made
  - E.g. Utterance: This is the same as that
  - E.g. Gesture: points to orange highlight “Megan” then 1/4
- Rarely the subject may make a gesture without saying anything. But if this occurs, record these in the gesture column and write [No utterance] in the Utterance column.
- Refer to the Item Sets packet to more easily see what problem the subject is working on

**Instructions Presented to Participant**

**General instructional presented at the beginning of the experiment**
In this study you will be asked to solve math problems that you may have encountered before in high school. Read each problem and type your answer into the computer. Scratch paper has been provided if you need to write out calculations, figure out steps, or take notes. For these problems, the answers are usually written in fractions, but you may also write the answer using
decimals. For example, if the answer is 1/4, you may also type in .25 or 0.25. If you have any questions about these tasks, please ask the researcher.

**Condition Specific Instructions Presented at the Start of the Study Phase**
Below are the instructions participants in each of the three learning conditions received before they studied the worked-examples. The instructions differ from each other slightly. Words shown in regular font denote instructions that are the same across all three conditions; italicized words are the part of the instructions that differ between the three conditions.

**Generate + Explain Instructions**
You will now be shown a series of different probability word problems. These problems have been solved for you in two different ways—by using a tree diagram and by using an equation. An example of a problem has been provided below.

You flip a coin twice. What is the probability that the coin lands on heads both times?

![Tree Diagram Example](image)

\[
P(\text{heads twice}) = \frac{\text{# of ways to get heads twice}}{\text{Total # of outcomes}} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}
\]

As you can see from the above example, a word problem will be presented at the top of the screen. Below the problem you will see a worked solution that uses a tree diagram. Below the tree diagram, you will see another worked solution that uses an equation to solve the problem.

Study these problems to understand how to find the answer because you will be asked to solve similar problems after you finish studying. As you read the problems and study the solutions, you should read and think aloud. Tell the experimenter what you are thinking about—say whatever comes to mind as you study these problems, even if it seems unimportant. Talk aloud about how information relates to what you have already seen, things you are confused about, insights you have, or anything else that comes to mind as you attempt to understand the problem.

*Important parts of the tree diagram have been highlighted for you. Your task is to make connections between the highlighted parts in the tree diagram and parts of the equation. That is, how do the highlighted parts in the tree diagram relate to parts in the equation? As you make these connections, explain why you think these connections are important. Making connections and explaining why they are important will help you better understand the problem and how to*
solve similar problems. Remember, make connections, explain the connections, and think aloud the entire time.

**Provided + Explain Instructions**

You will now be shown a series of different probability word problems. These problems have been solved for you in two different ways—by using a tree diagram and by using an equation. An example of a problem has been provided below.

**You flip a coin twice. What is the probability that the coin lands on heads both times?**

As you can see from the above example, a word problem will be presented at the top of the screen. Below the problem you will see a worked solution that uses a tree diagram. Below the tree diagram, you will see another worked solution that uses an equation to solve the problem.

Study these problems to understand how to find the answer because you will be asked to solve similar problems after you finish studying. As you read the problems and study the solutions, you should read and think aloud. Tell the experimenter what you are thinking about—say whatever comes to mind as you study these problems, even if it seems unimportant. Talk aloud about how information relates to what you have already seen, things you are confused about, insights you have, or anything else that comes to mind as you attempt to understand the problem.

*Important connections between the two different ways of solving the problem have been highlighted for you. Explain why you think these connections are important. Explaining why they are important will help you better understand the problem and how to solve similar problems. Remember, explain the connections and think aloud the entire time.*

**Explain Only Instructions**

You will now be shown a series of different probability word problems. These problems have been solved for you in two different ways—by using a tree diagram and by using an equation. An example of a problem has been provided below.
You flip a coin twice. What is the probability that the coin lands on heads both times?

\[ P(\text{heads twice}) = \frac{\text{# of ways to get heads twice}}{\text{Total # of outcomes}} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} \]

As you can see from the above example, a word problem will be presented at the top of the screen. Below the problem you will see a worked solution that uses a tree diagram. Below the tree diagram, you will see another worked solution that uses an equation to solve the problem.

Study these problems to understand how to find the answer because you will be asked to solve similar problems after you finish studying. As you read the problems and study the solutions, you should read and think aloud. Tell the experimenter what you are thinking about -- say whatever comes to mind as you study these problems, even if it seems unimportant. Talk aloud about how information relates to what you have already seen, things you are confused about, insights you have, or anything else that comes to mind as you attempt to understand the problem.

As you study these examples, explain why you must multiply to get the total number of acceptable outcomes. Explaining why you have to multiply will better help you understand the problem and also how to solve similar problems. Remember, explain and think aloud the entire time.

**Instructions That Appear Before Post-test**

You will now see problems similar to the ones you just studied. Use what you learned from studying the examples to solve these problems. You do not have to think aloud as you solve these problems.

**Instructions That Appear Before Procedural Far-transfer Items**

The final set of problems may look different from the examples you studied and the problems you solved previously. Nonetheless, use what you learned from studying the examples to solve these problems.
APPENDIX C

PILOT DATA

A plot study was run to determine where highlights may be most beneficial for the Generate + Explain condition. The research question of focus was whether highlighting important pieces of information in the diagram and requiring learners to relate those highlights to relevant pieces of information in the equation was more or less beneficial to procedural and conceptual than providing highlights in the equation and requiring learning to relate those highlights to relevant pieces of information in the diagram.

METHOD

Participants
Sixteen UIC undergraduates were recruited from the psychology subject pool to participate in this study, randomly assigned to one of the two conditions.

Design
The basic design is a repeated measures analysis of variance (ANOVA), with the learning condition (Highlighted Diagram, Highlighted Equation) as the between-subjects variable, post-test type (procedural isomorphic-transfer items, procedural near-transfer items, procedural far-transfer items, conceptual post-test items) as the within-subjects variable and test scores as the dependent measures.

Learning Conditions
As mentioned above, participants were placed in one of two learning conditions. In the “Diagram Highlights” learning condition, as participants studied the worked examples during the learning phase, important parts of the diagram were highlighted and participants were instructed to relate and explain those highlighted pieces of information to the relevant parts of the equation. In the “Equation Highlights” learning condition, as participants studied the worked examples during the learning phase, important parts of the equation were highlighted and participants were instructed to relate and explain those highlighted pieces of information to the relevant parts of the diagram.

Materials
The learning material came from the domain of probability theory, using the same questions participants were required to answer in the main study. The six worked examples presented during the study phase were the same as those used in the main study. Participants’ knowledge was assessed both before and after the study phase using a pre-test/post-test design. The pre-test consisted of procedural and conceptual knowledge questions. The post-test consisted of three types of procedural knowledge questions (isomorphic transfer, near transfer, and far transfer items) and conceptual knowledge questions. These were the same questions used in the main study.

Procedure
The procedure for the pilot study was identical to the procedure used in the main study. To briefly summarize: participants used a computer to first work through the basic probability pre-test questions, then answered procedural and conceptual knowledge pre-test questions, then studied the worked-examples as they thought aloud, and finally answered procedural and conceptual post-test items. All pre-test items, worked-examples, and post-test items were identical to those encountered by participants in the main study.

**Scoring**

**Procedural Knowledge Items**
Procedural pre-test and post-test knowledge items were scored as either correct (1 point) or incorrect (0 points). Equivalent probability fractions to the answer were scored as correct, as was the corresponding decimal value. For example, if the answer to an item was 2/10 then 1/5 and 0.20 would also be considered a correct response.

**Conceptual Knowledge Items**
Each conceptual knowledge item was scored on a three-point scale: no credit (0 points), partial credit (0.5 points), and full credit (1 point), as outlined in Appendix A.

**Justification Statements and Mappings**
Due to time constraints, participants’ think aloud data were not analyzed for the types of justifications made or the number of mappings produced between the representations for the pilot data. Therefore, results for these measures are not reported.

**RESULTS**

**Conceptual Knowledge**
Conceptual knowledge was assessed at pre-test and again at post-test. A 2(Learning Condition: Highlighted Diagram, Highlighted Equation) X 2 (Test: Pre-test, Post-test) mixed-model ANOVA with Learning Condition as the between-subjects variable and Test as the within-subjects variable was utilized to examine differences in conceptual knowledge understanding between the two learning conditions. The results revealed that participants in the Highlighted Diagram condition answered marginally more conceptual knowledge items correct than in the Highlighted Equation condition, $F(1, 14) = 4.18, p = .06$. Additionally, there was a main effect for test; more conceptual knowledge questions were answered correctly on the post-test compared to the pre-test, $F(1,14) = 12.22, p < .01$. These results were qualified by a Learning Condition by Test interaction, $F(1, 14) = 5.20, p < .05$ (see Figure 8).

Following up the interaction revealed the two conditions did not differ on pre-test conceptual scores, $F < 1$. However, by post-test the Highlighted Diagram condition correctly answered more conceptual knowledge items than the Highlighted Equation condition, $F(1,14) = 7.89, p < .05$.

**Procedural Knowledge**
Two analyses were conducted to assess understanding of procedural knowledge. First, an Analysis of Variance (ANOVA) examined changes in procedural knowledge understanding from pre-test to post-test across the three learning conditions. In this analysis, procedural pre-test scores and near transfer scores were the appropriate scores were used as the pre-test and post-test measures. That is, near transfer items were not encountered previously, either in the pre-test or
during the study phase. In the second analysis, two analyses of covariance (ANCOVAs) were used to assess the effects of learning condition on the other two types of post-test procedural knowledge.

A 2(Learning Condition: Highlighted Diagram, Highlighted Equation) X 2(Test: Pre-test, Near transfer Post-test) mixed-model ANOVA, with Learning Condition was the between subjects variable and Test as the within-subjects variable, revealed that participants demonstrated a higher level of procedural knowledge understanding on near transfer post-test compared to pre-test, \( F(1, 14) = 12.50, p < .01 \). However, there were no differences due to Learning Condition, or a Learning Condition X Test interaction, \( F^* < 1 \) (see Figure 9). Two ANCOVAs were utilized to assess the effects of learning condition on post-test procedural knowledge, with Learning Condition as the independent measure, scores on the remaining two post-test procedural item types (isomorphic transfer and far transfer items) as dependent measures, and procedural knowledge pre-test scores treated as a covariate. No differences emerged between the two learning conditions on isomorphic transfer or far transfer items, \( F < 1; F(1, 13) = 2.80, ns \), respectively. (see Figure 9).

**DISCUSSION**

In sum, the learning conditions did not differ from one another at post-test knowledge for any of the procedural items. However, at post-test the Highlighted Diagram condition demonstrated higher conceptual knowledge than the Highlighted Equation condition. Due to this advantage in conceptual knowledge performance, the decision was made to use highlights in the diagrams rather than in the equations for the Generate + Explain condition in the main study.
Figure 8. Mean percentage correct of conceptual knowledge items at pre-test and post-test by Learning Condition. Error bars represent standard error.

Figure 9. Mean percentage correct of procedural knowledge items by Learning Condition and Test Type. Error bars represent standard error.
APPENDIX D

**Think-Aloud Coding Rubric**

Transcribed think-alouds for each participant were coded with respect to how well the participant demonstrated understanding of the underlying probability principles and the degree to which participants integrated the diagrammatic- and text-based (i.e. algebraic equations) representations. In order to examine participants’ understanding of probability principles, their explanations for each of the six study items were coded for understanding of “procedural justifications” and “conceptual justifications.” Additionally, in order to examine how participants integrated the different representations (i.e. the word problem, the diagram, and the equation), the “mappings” (i.e. connections) participants made between the various representations were analyzed for what types of mappings were made (e.g., between the word problem and the diagram, between the diagram and the equation) and how often each type of mapping occurred.

**Understanding of Probability Principles**

**Procedural Justification Codes.**
In procedural justifications the participant relates a solution step to an underlying probability principle. (e.g. replacement does not occur). These types of explanations are considered indicators of procedural knowledge. These are statements identifying what principle is occurring and explaining how that principle manifests itself in the problem.

Participants’ think aloud data was coded using the codes found in Table VI. Each of the six study items was coded separately based on these codes. In the examples found in the table, words in *italics* indicate the key piece of information used to make the determination as to whether the participant understood each principle were used.

**Conceptual Justification Codes**
Conceptual justifications are deeper and more elaborated than procedural justifications because conceptual justifications require explaining why a particular probability concept is important and useful at a given solution step. These are statements about why a principle is important.

Participants’ think aloud data was coded using the codes found in Table VII. Each of the six study items were coded separately based on these codes. In the examples found in the table, words in *italics* indicate the key piece of information used to make the determination as to whether the participant understood why the principle was being used.

**Mapping coding**
All study items for each participant were also be analyzed for what representations were “mapped” onto one another, and how often these mappings occurred. In each study item there were three distinct elements present: the word problem, the tree diagram worked solution, and the equation-based worked solution. Think-alouds were examined for what types of mappings (see Table VIII) occurred and how often each mapping is occurred (i.e. tokens). Both utterances and gestures made by the participants were used to make determinations of mappings between representations. In the examples, words in *italics* indicate the key piece of information used to make the determination as to what mappings are occurring.
General points to keep in mind:

- Even if the participant incorrectly uses a principle when that principle is not appropriate (e.g. the participant states that order matters, when in fact it does not for that item), still mark the participant as having used a procedural justification. Incorrect explanations will be analyzed separately if needed.

- The participant does not have to use the exact terms shown in Tables VI and VII. Any term that sounds similar in your judgment is acceptable. In addition to the phrases found in the tables, some additional phrases you may encounter include: order matters, order doesn’t matter, the item is replaced, the item is not put back, the item can only be selected once, the same item can be picked again, each pick is separate because…, because the order doesn’t matter there are more acceptable outcomes, the fractions are multiplied because…”

- If a participant skipped over an item, mark that item with a “-9” in all fields on the scoring sheet
<table>
<thead>
<tr>
<th>Principle</th>
<th>How the Principle Manifests</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order Relevance</td>
<td>The order in which objects are selected matters. (i.e. one item must be picked first and one item must be picked second).</td>
<td>Item 1: “Hillary Picks first, so that’s why she shows up first on the tree diagram. And then Megan’s the second one.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Item 4: “Here the order matters. So you have to pick the green one first...and then the second pick has to be yellow.”</td>
</tr>
<tr>
<td>Order Irrelevance</td>
<td>The order in which items are selected does not matter.</td>
<td>Item 3: “It doesn’t matter which one happens first. You just need to win both times”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Item 6: “It doesn’t matter who gets what color as long as you get both of them”</td>
</tr>
<tr>
<td>With Replacement</td>
<td>After an object is selected it has the opportunity to be selected again for the next event.</td>
<td>Item 3: “You want to spin again. There are still six colors. It’s not going to remove a color from the wheel”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Item 4: “The [ball] can be put back in the bag”</td>
</tr>
<tr>
<td>Without Replacement</td>
<td>After an object is selected it cannot be selected again.</td>
<td>Item 2: “When you take one out it is not put back in for the second event”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Item 5: “The second time…it is not being replace. You only have five [marbles] the second time”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Item 6: “The next pick you’re going to take it away because you already got that color”</td>
</tr>
<tr>
<td>Multiplication Rule</td>
<td>The two probability fractions are multiplied to get the total number of desired outcomes over the total possible outcomes.</td>
<td>“So you multiply [the fractions] together to get 1/30”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“It’s 2/8 times 1/7 which equals 2/56”</td>
</tr>
</tbody>
</table>
### TABLE VII
CONCEPTUAL JUSTIFICATIONS AND EXAMPLES

<table>
<thead>
<tr>
<th>Principle</th>
<th>Justification for Why the Concept is Important</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Order Relevance</strong></td>
<td>Order relevance only allows a set of events to occur in one manner (e.g. AB and not BA). This results in a restricted set of desirable outcomes compared to order irrelevance. Participants show conceptual understanding of this principle through explaining how the ordered event is related to the underlying probability.</td>
<td>Item 4: “So there are five colors and you are going to focus on the green one the first time then the second time, you are going to want it to be yellow. So you can only pick green once and you can only pick yellow once. And each time it’s—you have a chance of one. The probability is 1/5 each time.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Item 2: “So you need to pick two or three first because those are the two possible outcomes you want over the total. So you put 2/8. And then pick seven. So that’s 1/7 because you already picked two or three, so there’s one less number to pick from.”</td>
</tr>
<tr>
<td><strong>Order Irrelevance</strong></td>
<td>Order irrelevance allows for a set of events to occur in two different manners (e.g. AB or BA). This results in a less restricted set of desirable outcomes compared to order relevance. Participants show conceptual understanding of this principle through explaining how the unordered event is related to the underlying probability.</td>
<td>Item 6: “Without having to worry about what the order is, you can just specifically look at the red for your first pick and then green. Or red for your second pick. It has to be the opposite for each one cause you want one of each color…red and green. That’s why there are 2/56, the probability of that actually happening”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Item 6: “At the first pick there are eight helmets to pick from. And it doesn’t matter which one is picked as long as we pick the two that we want. So that means there is a 2/8 chance or picking one of the helmets we want.”</td>
</tr>
<tr>
<td><strong>With Replacement</strong></td>
<td>Replacing a desired outcome back into the sample space for the second draw, you a increasing the probability of drawing a desired event on the second draw compared to not replacing the outcome; With replacement, the probability of Event 1 and Event 2 are the same -OR- Conversely, replacing a non-desired outcome back into the sample space for the second draw, you a decreasing the probability of drawing a desired event on the second draw compared to the first draw. In this instance, the probability of drawing a desired outcome is less than if the non-desired outcome was not replaced. Participants demonstrate conceptual understanding through explaining how replacement affects the underlying probability of the event occurring.</td>
<td>Item 3: “The total sample size remains the same. So you still have equal probability of getting gold or silver”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Item 4: “The probability of picking the green ball is 1/5. But since the green ball is replaced, the sample size remains the same and the probably of pick yellow second is 1/5”</td>
</tr>
<tr>
<td>Principle</td>
<td>Justification for Why the Concept is Important</td>
<td>Examples</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Without Replacement | Without replacing a desired outcome back into the sample space for the second draw you a decreasing the probability of drawing a desired event on the second draw compared to the first draw; Without replacement of a desired outcome results in a lower probability for a desired outcome compared to replacing the desired outcome. **-OR-** Conversely, without replacing a non-desired outcome back into the sample space for the second draw, you are increasing the probability of drawing a desired event on the second draw compared to the first draw. In this instance, the probability of second draw is higher than if the non-desired outcome was replaced into the sample space. Participants demonstrate conceptual understanding through explaining how replacement affects the underlying probability of the event occurring. | Item 2: “When you take out a slip it is not put back for the other [event]. The second [event] that you go to reach in would be different. Which is why it changes the total outcomes from seven to six. Because there were originally seven and you took one out and didn’t put it back.”  
Item 5: “You have to draw again, but the marble you picked before isn’t replaced so the sample size is reduced to five…but still have two green marbles to pick, so the probability of picking any of the two marbles is 2/5”  
Item 6: “If I pick the red helmet then that’s been removed from the sample size and that will leave 7 colors including the green helmet for my friend to pick. She needs to get the green helmet, so that’s a probability of 1/7” |
### TABLE VIII
#### TYPES OF MAPPINGS

<table>
<thead>
<tr>
<th>Representation Connected</th>
<th>Explanation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Problem and Diagram</td>
<td>Relates an element of from the word problem to the corresponding element in the tree diagram (or vise versa).</td>
<td>“[Reads Problem: A basket contains 1 red, 2 black, and 3 green marbles. Two marbles are drawn at once. Find the probability the sample contains 2 green marbles]. You start out with one red two back and three green marbles [points to first column of the diagram showing one red, two back and three green marbles] which is why they are divided that way.”</td>
</tr>
<tr>
<td>Word Problem and Equation</td>
<td>Relates an element in from the word problem to the corresponding element in the equation (or vise versa).</td>
<td>“[Reads Problem: At the carnival you spin a wheel of fortune twice. The wheel is divided into 5 colored segments. Of the 5 segments one is colored gold and one is colored silver. You win if your spin lands on the gold of silver segment. What is the probability that you win on each of your two spins?] So two ways to win, the gold or the silver. So two over five [points to equation] because there are five segments, so times it by two.”</td>
</tr>
<tr>
<td>Equation and Diagram</td>
<td>Relates an element from the equation to the corresponding element in the tree diagram (or vise versa).</td>
<td>Equation to Diagram: “In the equation the total number of outcomes is...okay...I see...42. That’s the same as the outcomes up here [points to the “outcomes” column of the diagram showing all possible outcomes].”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diagram to Equation: “Once the green [marble] has been picked, the highlighted ones show that there are two green left to be picked. Which is how they got 2/5 [pointing to 2/5 fraction in equation]”</td>
</tr>
</tbody>
</table>


APPENDIX E

Mapping Token Analysis

An analysis of the mapping tokens largely mirrors the results obtained from the mapping types analysis. A 3(Learning Condition: Generate + Explain, Provided + Explain, Explain Only) X 3 (Mapping Category: Problem-Diagram, Problem-Equation, Diagram-Equation) mixed-model ANOVA revealed a main effect for Learning Condition, $F(2, 72) = 5.88, p < .01$, $\eta^2_p = .14$. Additionally, a main effect for Mapping Category was also observed, $F(2, 144) = 18.02, p < .001$, $\eta^2_p = .20$. These main effects were qualified by a Learning Condition X Mapping Category interaction, $F(4, 144) = 3.78, p < .01$, $\eta^2_p = .10$ (see Figure 10).

Planned comparisons following up the main effect of Learning Condition, revealed no differences between the number of total mappings made by the Generate + Explain and Provided + Explain conditions, $F(1, 72) = 2.20, ns$. However, the Generate + Explain condition produced more total mappings than the Explain Only condition, $F(1, 72) = 11.17, p < .01$. Following up the main effect for mapping category revealed more Problem-Equation mappings were produced than Problem-Diagram mappings, $F(1, 144) = 8.49, p < .01$. Additionally, more Problem-Diagram mappings were produced than Diagram-Equation mappings, $F(1, 144) = 9.53, p < .01$.

The Diagram-Equation mapping follow up analysis revealed that participants in the Generate + Explain condition produced only marginally more diagram-equation mappings than participants in the Provided + Explain condition, $F(1, 144) = 2.99, p < .10$. However, both the Generate + Explain and Provided + Explain conditions produced more diagram-equation mappings than the Explain Only condition, $F(1, 144) = 11.29, p < .01; F(1, 144) = 4.45, p < .05$, respectively. Tukey HSD tests revealed that Generate + Explain and Provided + Explain did not differ in the number of problem-diagram mappings produced, however, both conditions produced more problem-diagram mappings than the Explain Only condition. Furthermore, all three learning conditions did not differ in the number of problem-equation mappings produced.

In sum, the results from this token analysis largely mirror the results from the mapping type analysis, specifically that participants in the Generate + Explain did not produce statistically more diagram-equation mappings than participants in the Provided + Explain condition.
Figure 10. Mean number of mappings made during the study phase by Learning Condition and Mapping Category. Error bars represent standard error.
APPENDIX F

Procedural Knowledge Meditational Relationships

Examining the correlations between performance post-test measures, mapping types, and the number of procedural justifications (see Table I in the main text) revealed several relationships that warranted further exploration. Specifically, there were positive correlations between problem-equation mapping types, procedural justifications, and post-test isomorphic transfer knowledge. There were also positive associations between problem-equation mapping type, procedural justifications, and post-test far transfer procedural knowledge. These relationships were investigated using the Preacher and Hayes (2008) bootstrapping method described previously. Results of both analyses are based on 1000 bootstrap samples.

First, we analyzed whether higher use of problem-equation mappings types was associated with higher levels of procedural justifications, which in turn, were associated with higher post-test knowledge on isomorphic transfer procedural items. The total mediation model accounted for significant variance in isomorphic transfer knowledge, $R^2 = .13$, $F(2, 72) = 5.56$, $p < .05$. As shown in Figure 1, higher use of the problem-equation mapping types were associated with higher use of procedural justifications. Additionally, higher use of procedural justifications were associated with higher post-test knowledge on isomorphic procedural transfer items. Finally, the analysis revealed, with 95% confidence, the direct effect of problem-equation mapping types on post-test isomorphic transfer knowledge was fully mediated by procedural justifications, with a BCA bootstrap confidence interval of .03 to 1.09.

Next, we analyzed whether greater use of problem-equation mapping types were associated with higher levels of procedural justifications, which in turn, were associated with higher post-test knowledge on far transfer procedural items. The total mediation model accounted for significant variance in far transfer knowledge, $R^2 = .12$, $F(2, 72) = 4.80$, $p < .05$. As shown in Figure 2, higher use of problem-equation mapping types was associated with higher use of procedural justifications. Additionally, higher use of procedural justifications was associated with higher post-test knowledge on far transfer procedural items. Finally, the analysis revealed, with 95% confidence, the direct effect of problem-equation mapping types on post-test far transfer knowledge was fully mediated by procedural justifications, with a BCA bootstrap confidence interval of .03 to .76.

In summary, the number of procedural justifications mediated the relationship between the problem-equation mapping types and isomorphic and far transfer procedural knowledge at post-test. Interestingly, post-test near transfer knowledge was not significantly associated with any type of mapping or statement.
Figure 11. Path diagram of mediation model for isomorphic transfer procedural items. * p < .05

Figure 12. Path diagram of mediation model for far transfer procedural items. * p < .05
IRB APPROVAL

UNIVERSITY OF ILLINOIS
AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Approval Notice
Continuing Review

July 22, 2011

Kevin Dietz, BA
Psychology
1007 W Harrison
Learning Sciences Research Institute, M/C 285
Chicago, IL 60607
Phone: (630) 965-0921 / Fax: (312) 413-4122

RE: Protocol # 2009-1043
“Moderating Effects of Mathematical Reasoning”

Dear Mr. Dietz:

Your Continuing Review was reviewed and approved by the Expedited review process on July 21, 2011. You may now continue your research.

Please note the following information about your approved research protocol:

**Protocol Approval Period:** July 21, 2011 - July 19, 2012
**Approved Subject Enrollment #:** 200 (150 subjects enrolled)
**Additional Determinations for Research Involving Minors:** The Board determined that this research satisfies 45CFR46.404, research not involving greater than minimal risk. Therefore, in accordance with 45CFR46.408, the IRB determined that only one parent's/legal guardian's permission/signature is needed. Wards of the State may not be enrolled unless the IRB grants specific approval and assures inclusion of additional protections in the research required under 45CFR46.409. If you wish to enroll Wards of the State contact OPRS and refer to the tip sheet.
**Performance Sites:** UIC
**Sponsor:** None
**Research Protocol:**
   a) Moderating Effects of Mathematical Reasoning; Version 5; 02/21/2011
   b) **Recruitment Material:**
      a) UIC Psychology Department Subject Pool recruitment procedures will be followed.
Informed Consents:

a) Math Problem Solving; Version 4.0, 02/21/2011
b) Debriefing Form; no version number, no date

Parental Permission:

a) A waiver of parental permission has been granted under 45 CFR 46.116(d) and 45 CFR 46.408(c); however, as per UIC Psychology Subject Pool policy, at least one parent must sign the Blanket Parental Permission document prior to the minor subject’s participation in the UIC Psychology Subject Pool.

Your research meets the criteria for expedited review as defined in 45 CFR 46.110(b)(1) under the following specific category:

(7) Research on individual or group characteristics or behavior (including but not limited to research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

Please note the Review History of this submission:

<table>
<thead>
<tr>
<th>Receipt Date</th>
<th>Submission Type</th>
<th>Review Process</th>
<th>Review Date</th>
<th>Review Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>07/18/2011</td>
<td>Continuing Review</td>
<td>Expedited</td>
<td>07/21/2011</td>
<td>Approved</td>
</tr>
</tbody>
</table>

Please remember to:

→ Use your research protocol number (2009-1043) on any documents or correspondence with the IRB concerning your research protocol.
→ Review and comply with all requirements on the enclosure, "UIC Investigator Responsibilities, Protection of Human Research Subjects"

Please note that the UIC IRB has the prerogative and authority to ask further questions, seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

Please be aware that if the scope of work in the grant/project changes, the protocol must be amended and approved by the UIC IRB before the initiation of the change.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact OPRS at (312) 996-1711 or me at (312) 355-1609. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Rahab Mwangi, B.S.
IRB Coordinator, IRB # 2
Office for the Protection of Research Subjects
Enclosures:

1. UIC Investigator Responsibilities, Protection of Human Research Subjects
2. Informed Consent Documents:
   a) Math Problem Solving; Version 4.0, 02/21/2011
   b) Debriefing Form; no version number, no date

cc: Gary E. Raney, Psychology, M/C 285
    Susan R. Goldman, Psychology, M/C 285
# VITA

## Kevin D. Dietz

### Education

<table>
<thead>
<tr>
<th>Degree</th>
<th>Institution</th>
<th>Field</th>
<th>Thesis Title</th>
</tr>
</thead>
</table>

Advisor: Susan R. Goldman, Ph.D.

<table>
<thead>
<tr>
<th>Degree</th>
<th>Institution</th>
<th>Major</th>
<th>Minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.A., 2007</td>
<td>University of Michigan, Ann Arbor</td>
<td>Psychology</td>
<td>Philosophy</td>
</tr>
</tbody>
</table>

### Publications


### Presentations


Honors and Awards

UIC Graduate Student Council Travel Award (2010; $300)
Psychology Department Travel Award (2010, 2011; $200)
UIC Graduate College Travel Award (2012, $200)
Psi Chi Membership (National Honors Society in Psychology)
University Honors, University of Michigan

Research Experience

National Mathematics Center Project, UIC (2010 - 2012)
   PIs: Steve Schneider (WestEd), James W. Pellegrino, Susan R. Goldman, Kenneth Koediger (CMU), Julie Booth (Temple), Neil Heffernan (WPI), Mitchell Nathan (UW-M), & Martha Alibali (UW-M)
   Graduate Research Assistant
The goal of the National Math Center is to apply research-based cognitive
principles to redesign a widely used middle school curriculum and to evaluate the
efficacy of these redesigned materials.

_Curriculum Embedded Assessment, Learning Sciences Research Institute, UIC (2009-2010)_
- PIs: James W. Pellegrino, Lou DiBello, & Susan R. Goldman
- Graduate Research Assistant
Evaluating the cognitive, psychometric, and instructional validity of assessments
present in elementary mathematics curriculum.

_Effects of Diagrammatic Representations on Algebraic Problem-solving (2009 - 2010)_
Investigated how diagrammatic representations (tables, graphs, etc.) impact
solution accuracy in solving algebra word problems.

_Spatial Intelligence & Learning Center, University of Chicago (2007 – 2009)_
- PIs: Susan C. Levine & Janellen Huttenlocher
- Research Assistant
This project focused on understanding the development of spatial reasoning and its
relationship to mathematical thinking and success in the sciences. Specifically, I
was responsible for creating and implementing a training program intended to
increase children’s comprehension of measurement units.

_Center for Early Childhood Research, University of Chicago (2007 – 2009)_
- PI: Janellen Huttenlocher
- Research Assistant
Designing a set of studies to evaluate infants’ basic knowledge of categorical fields,
such as number, color, and size

_Pathways to Literacy, University of Michigan (2006 – 2007)_
- PI: Frederick J. Morrison
- Undergraduate Research Assistant
Investigating the longitudinal development of factors that influence childhood
literacy.

**Mentoring Experience**

**Undergraduate Research Assistants**
Dylan Olson (Winter Semester 2010 – Spring Semester 2011)
- Projects: Effects of diagrammatic representations on algebraic problem-solving,
  Creating mappings when learning from multiple representations, National Math
  Center Project
Ankita Matharu (Winter Semester 2011)
Projects: Creating mappings when learning from multiple representations, National Math Center Project

Laura Whiteford (Fall Semester 2011)
Projects: Creating mappings when learning from multiple representations

Michelle Khatchadourian (Fall 2011 – Spring 2012)
Projects: Creating mappings when learning from multiple representations

Professional References

Susan R. Goldman, Ph.D.
Distinguished Professor of Psychology and Education,
Department of Psychology
University of Illinois at Chicago
1007 W. Harrison Street (M/C 285)
Chicago, Illinois 60607
Phone: (312) 996-4462
Email: sgoldman@uic.edu

Jennifer Wiley, Ph.D.
Associate Professor,
Department of Psychology
University of Illinois at Chicago
1007 W. Harrison Street (M/C 285)
Chicago, Illinois 60607
Phone: (312) 355-2501
Email: jwiley@uic.edu

James W. Pellegrino, Ph.D.
Distinguished Professor of Psychology and Education,
Department of Psychology
University of Illinois at Chicago
1007 W. Harrison Street (M/C 285)
Chicago, Illinois 60607
Phone: (312) 355-2493
Email: pellegjw@uic.edu

Kristin Ratliff, Ph.D.
Project Director,
Psychological and Educational Assessment Development
Western Psychological Services
625 Alaska Avenue,
Torrance, California 90503
Phone: (424) 201-8800
Email: kratliff@wpspublish.com