Effect on Superficial Variability of Examples on Learning Applied Probability

Tiantian Jin

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy under the Executive Committee of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018
ABSTRACT

Effect of Superficial Variability of Examples on Learning Applied Probability

Tiantian Jin

Learning through examples is a central and widely used instructional device for teaching mathematically-based subjects such as statistical probability. However, the applications of the superficial variability of examples remain controversial. This dissertation investigates how the superficial variability of multiple examples influences students’ learning and transfer of probability problem-solving. Moreover, the author discovers whether content difficulty affects the influence of examples’ superficial variability. Three conditions were developed and compared: consistent-surface condition (CS), varied-surface-within-rule condition (VSWR), and varied-surface-between-rule condition (VSBR). For the purpose of exploration and methodology improvement for the dissertation study, two pilot studies were conducted. However, conflicting results were shown in those two studies. In the first pilot study, students in CS condition performed the worst. In the second pilot study, students in VSBR condition performed the worst. These conflicting results encouraged the author even more to conduct the dissertation study with a larger sample size and improved methodology. In this dissertation study, the author found that students’ performance on the posttests in VSBR was significantly worse than in the other two conditions, which was consistent with the second pilot study, and that their performance in CS and VSWR condition was not different. Contrary to expectation, the strength of the pattern of the effect of the superficial variability of examples did not vary between the easy and difficult types of problems. Moreover, the pattern was the same when the difficulty variable was not
included. These results suggested that examples’ superficial consistency between different problem types promotes more effective learning than superficial variation between different problem types. The consistency can be one single cover story used multiple times for each type of problem or the same battery of varied cover stories used repeatedly for different types of problem. Moreover, the pattern of the influence of superficial variability of examples is robust among types of the problem at varying difficulty levels.
List of Tables ................................................................. ii
List of Figures ................................................................. iii
Acknowledgements .......................................................... v
1. Introduction (Research Goal and Scholarly Significance of Dissertation) ........ 1
2. Literature Review .................................................................. 3
   2.1. Schema Construction ......................................................... 2
   2.2. Analogical Transfer ............................................................ 4
   2.3. Learning through examples .................................................. 7
      2.3.1. Multiple Examples ......................................................... 7
      2.3.2. Varied Examples ............................................................ 8
      2.3.3. The Role of Prior Knowledge ........................................... 13
   2.4. Conceptual Framework of This Dissertation Project ....................... 14
3. Pilot Studies ............................................................................. 20
   3.1. Pilot One ........................................................................... 20
      3.1.1. Methodology ............................................................... 25
      3.1.2. Results and Discussion .................................................. 27
      3.1.3. Conclusions ............................................................... 34
      3.1.4. Limitations and Improvements of the Experiment Design ............ 34
   3.2. Pilot Two ............................................................................ 35
      3.2.1. Methodology ............................................................... 36
      3.2.2. Results and Discussion .................................................. 38
      3.2.3. Conclusions ............................................................... 44
      3.2.4. Limitations and Improvements of the Experiment Design ............ 45
4. Dissertation Study Overview ..................................................... 47
5. Research Questions and Hypotheses for the Dissertation Study ..................... 48
6. Methodology ............................................................................ 50
   6.1. Participants ....................................................................... 50
   6.2. Design and Manipulation ..................................................... 50
   6.3. Procedure ....................................................................... 51
   6.4. Materials and Measures ..................................................... 53
7. Results .................................................................................... 56
8. Discussion ................................................................................ 71
9. Limitations and Future Research Direction ................................................. 75
References ................................................................................ 77
Appendix A. Personal Information Survey ..................................................... 86
Appendix B. Comprehension Questions in the Pretest ........................................ 87
Appendix C. Application Questions in the Pretest .............................................. 88
Appendix D. Questions in the Transfer Test .................................................... 90
Appendix E. Feedback Survey .................................................................. 92
Appendix F. Sample Worked Problem ........................................................ 93
Appendix G. Screen Shots for the Instructional Slides ..................................... 94
Appendix H: Cover Stories Used in the Pretest and Posttests and Instruction .......... 100
List of Tables

Table 1 Comparison of Design and Definition in Ross (1989b) and Quilici & Mayer (1996) ..........................................................13
Table 2 Three Conditions..........................................................................................................................................................21
Table 3 Differences in Design for Two Pilot Studies ........................................................................................................38
Table 4 Comparisons of the Results in Two Pilot Studies ............................................................................................45
Table 7.1 Accuracies of the four tests of all conditions and each separate condition......56
Table 7.2.1 Values of cond_new ........................................................................................................................................60
Table 7.2.2 Helmert Contrast Results (K Matrix) of the Two Posttests .................................................60
Table 7.3 Helmert Contrast Results (K Matrix) for the Transfer Test.................................................62
Table 7.4.1 Difficulty Ratings of the Easy and Difficult Rules..................................................................................64
Table 7.4.2 Performance on Two Difficulty Levels of Rules in the Tests ........................................65
Table 7.4.3 Within-Subject Factors in the Three Way Mixed ANCOVA Test (BWW)...67
Table 7.4.4 Tests of Within-Subjects Effects ...............................................................................................67
Table 7.4.5 Helmert Contrast Results (K Matrix) for the Posttests with Difficulty as a Within-Subject Factor.................................................................................................................................68
Table 7.4.6 Helmert Contrast Results (K Matrix) for the Transfer Test with Difficulty as a Within-Subject Factor.................................................................................................................................70
List of Figures

Figure 1. Analogical Transfer & Schema Construction......................................................16
Figure 2. The Learning Process Through Examples..........................................................22
Figure 3. Different Processes of Schema Construction in Three Conditions .....................23
Figure 4. Participants' Performance Accuracy in Pretest and Posttest.............................28
Figure 5. Performance Accuracy in Pretest and Posttest of Three Groups.........................28
Figure 6. Equivalent Prior Knowledge Across Groups ....................................................29
Figure 7. Accuracy Improvements Across Groups..........................................................30
Figure 8. Accuracy Improvements for Permutation Questions across Groups ...................31
Figure 9. Comparisons of Helpfulness for Instructions of Each Rule ...............................32
Figure 10. Overall Performance Accuracy of Each rule...................................................34
Figure 11. The Task Flow in Two Pilot Studies ...............................................................37
Figure 12. Participants' Performance Accuracy in Pretest and Posttest in Pilot 2 ...............39
Figure 13. Performance Accuracy in Pretest and Posttest of Three Groups in Pilot 2 .......39
Figure 14. Inequivalent Prior Knowledge across Groups in Pilot 2 .................................40
Figure 15. The Performance Improvement in Three Groups in Pilot 2 ............................41
Figure 16. The Performance Improvement of the Difficult Rule in Three Groups in Pilot 2  ........................................................................................................................................43
Figure 17. The Performance in the Transfer Test Across Three Groups in Pilot 2 ..........44
Figure 18. The Performance in the Delayed Transfer Test Across Three Groups in Pilot 2 ........................................................................................................................................44
Figure 6.3. Task Flow in the Dissertation Study.............................................................52
Figure 7.1.1. Mean of Accuracies of the Pretest, Posttest, Transfer Test and Delayed
Posttest in Each Condition (Clustered by Condition) ..........................................................58

Figure 7.1. Mean of Accuracies of the Pretest, Posttest, Transfer Test and Delayed Posttest in Each Condition (Clustered by the Type of Tests) ..............................................58

Figure 7.2. Profile Plots of Students’ Performance in Two Posttests in Three Conditions .................................................................................................................................61

Figure 7.3. Profile Plots of Students’ Performance in the Transfer Test..............................63

Figure 7.4. Perceived Difficulty of the Easy and Difficult Rules...........................................65

Figure 7.4. Performance on the Easy and Difficult Rules in Four Separate and Overall Tests ........................................................................................................................................66

Figure 7.4. Profile Plots of Students’ Performance in Two Posttests at Two Difficulty Levels in Three Conditions ..................................................................................................68

Figure 7.4. Profile Plots of Students’ Performance in the Transfer Test at Two Difficulty Levels in Three Conditions ..................................................................................................70
Acknowledgement

After an intensive period, today is the day to reflect on the people who have supported and helped me so much throughout all difficulty and stress.

I would first like to express the deepest appreciation to my advisor Professor John Black, who has provided me with generous support and advice in designing and conducting this study. I am also grateful to Professor James Corter, who always directed me to the right path, and Professor Ryan Baker, who never failed to be supportive and encouraging. I also want to thank Professor Fran Blumberg for her continued support and care, and Professor Mitchell Rabinowitz for his valuable suggestions.

A special thank you to Kim Moy. I am so lucky to have met you. You have no idea how much your help has meant, not only for this dissertation but also for my career and life. Other many thanks go to Juan Wu. Thank you for taking care of Meatball and me with the warmest heart all the time.

Nobody has been more important to me in the pursuit of this project than the members of my family. I would like to thank my parents, who sacrificed all they had and more, so that I could pursue whatever I want. Finally, to my partner, best friend, and love of my life, Hao Qu: thank you for your love, support and understanding. Thank you for making me a better me. I believe that the future will be brighter and happier, with you.
1 Introduction

Research Goal and Scholarly Significance of this Dissertation Project

In the past, research on learning and cognition has shown that superficial variability of examples affects learning and transfer (Ross 1989b; Quilici & Mayer, 1996). However, the mechanism by which and the extent to which superficial variability of multiple examples exerts influence remain controversial. Without a clear understanding of the influence, it is difficult for instructors to find a more effective way to teach students. The purpose of this dissertation project is to study the effects of superficial variability of instructional examples on learning performance and transfer. Moreover, the author explores whether the strength of its effects is different for concepts and problem types at different difficulty levels. In this dissertation study, the learning domain is statistical probability, chosen because the author had observed that statistical probability was challenging for most college students. The information can be used to help refine instructional design for probability and to help students acquire a more effective learning outcome. Furthermore, the learning process and strategy of most mathematically-based curricula are very similar. If the instructional method has a positive effect on the learning of statistical probability, it is reasonable to suggest that the same approach may be useful in the learning of other mathematically-based subjects.
2 Literature Review

2.1 Schema Construction

Learning is a constant process of acquiring new knowledge, combining it with existing knowledge, and rebuilding the internal deep structures of all related knowledge points. The internal mental structures are referred to as "schemas" (Piaget, Elkind, & Tenzer, 1968). The appropriateness of schemas affects whether individuals are able to understand concepts successfully and solve related problems (Mayer, 1992). Word problems are a common question type for mathematics-related knowledge, including statistics. As individuals look at word problems, beyond linguistic and semantic knowledge, they also need knowledge of the problem type. To clarify, the problem type is not about the format of a problem, but rather the essential concepts involved in that problem. Therefore, concept and problem type have very similar definitions in this study. The knowledge of the problem type is schematic knowledge, which helps to distinguish relevant from irrelevant information (Mayer, 1991). As demonstrated in various studies, schematic knowledge is a fundamental component of mathematical problem-solving expertise (Chi, Feltovich, & Glaser, 1981; Cummins, 1992; Hinsley, Hayes, & Simon, 1977; Mayer, 1981, 1982; Riley, Greeno, & Heller, 1983; Schoenfeld & Herrmann, 1982; Silver, 1981). Research on expertise has suggested that experts are more likely to sort problems on the basis of deep structural features and less likely to sort on the basis of superficial features when compared to novices (Chi, Feltovich, & Glaser, 1981; Rabinowitz & Hogan, 2001; Silver, 1981). Superficial features depend on attributes of objects in the problem and are derived from aspects of cover stories, whereas deep structural features depend on relations among objects and determine aspects of the
required solution procedure (Vosniadou & Ortony, 1989). For instance, in this dissertation study, superficial features of permutations problem include story characters (drawing cards from a deck of poker cards, picking marbles from a jar of colorful marbles, etc.). The deep structural features of permutations problem include whether the order of things (a card of King and a card of A, and a blue marble and a red marble) is considered. If the order is not considered, the problem may be a combinations problem instead of a permutations problem. Obviously, successful problem solving is closely related to structure-based problem schemas, rather than surface-based ones. For that reason, learning and instructional materials, such as examples and practices, should promote students’ ability to recognize structural features rather than focus on surface features.

When individuals learn with examples, schemas play the role as abstractions from specific problem instances that can be used to make inferences about instances of the concepts they represent (Anderson, 2010). In other words, schemas are at the heart of successful problem representation, which is a major factor in solving a word problem. Hence, appropriate problem schemas can affect whether subjects are able to solve a word problem. For example, Riley, Greeno, and Heller (1983) have suggested that failure to solve word problems might be caused by a lack of appropriate schemas, rather than poor arithmetic or logical skills. Moreover, Hinsley, Hayes, and Simon (1977) have illustrated that wrong schemas can result in many difficulties in solving word problems.

Previous studies have shown that schema acquisition plays a critical role in reaching transfer of problem-solving skills (Pass & Van Merrienboer, 1994). Cognitive schemas are conceptualized as cognitive structures that enable problem solvers to
recognize problems as belonging to particular categories, requiring particular operations to reach a solution (Pass & Van Merrienboer, 1994). Similarly, Gick and Holyoak (1983) have defined a schema as the generalized description of two or more problems and their solutions. Additionally, their results indicated that when problem solvers generated a more effective schema, transfer was enhanced. Subsequently, Cooper and Sweller (1987) have examined the effect of schema acquisition on problem-solving transfer as well. According to their empirical results, schema acquisition precedes rule automation, and has a strong impact on problems similar to initial acquisition problems, which promotes problem-solving transfer.

Since schema construction is a critical component of successful learning and problem solving, it is necessary to probe how the process of schema construction can be promoted. According to research on learning from examples, experience with example problems has promoted the construction of problem schemas (Bransford, 1979; Reed & Bolstad, 1991; Sweller & Cooper, 1985; Ward & Sweller, 1990; Zhu & Simon, 1987). Moreover, Ranzijn (1991) and Shapiro and Schmidt (1982) have demonstrated that schema acquisition benefits from wide variability of practice along the task dimensions, and in turn transfer of acquired skills can be facilitated because varied practice increases the chances that similar features can be identified and that relevant features can be distinguished from irrelevant ones.

2.2 Analogical Transfer

Learning is never only for being able to solve any specific example question, but rather for being able to transfer the knowledge or skill acquired from the solved problems to new ones. Thinking by analogy is one critical strategy during this process. For
example, when a person is confronted with a new statistics word problem (which is the target problem), he/she can solve it by drawing an analogy to a similar problem (which is the source problem) that this person has successfully solved in the past, sometimes using a method abstracted from the source problem. This process is defined as analogical transfer (Gentner, 1989; Holyoak, 1985; Mayer, 1992; Ross, 1987; Vosniadou & Ortony, 1989). Based on the analogical transfer hypothesis, there are three essential components of successful analogical transfer from a known problem (the source problem) to a new one (the target problem): recognition, in which a problem solver identifies a potential source problem from which to reason; abstraction, in which a problem solver abstracts a general structure, principle or procedure from the source; and mapping, in which a problem solver applies that knowledge and solution to the target (Mayer, 1992).

Knowing a solution for an analogous problem (the source) is not necessarily useful, unless students realize that this problem is analogous to the one they are working on (the target). In other words, recognition is the necessary prerequisite to successful analogical reasoning. This process depends on whether the problem solver can recognize the similarity between the new problem that he/she is working on and related problems that he/she has solved. In order to recognize similarities between problems, two techniques are available: one is to focus on superficial similarity, and the other is to focus on structural similarity of problems (Vosniadou & Ortony, 1989). As mentioned before, the structural feature is the one that affects the solution procedure, thus for the purpose of solving new problems, successful recognition relies more on identifying structural similarity, rather than superficial similarity.
According to Mayer (1992), the second step is to abstract the general characteristics from the analogs (source questions) for use in solving other problems. Those general characteristics are formed by finding the commonalities in the structure of the analogs, which forms a schema (Gick & Holyoak, 1983). In other words, the process of abstraction corresponds to structure-based schema construction. Therefore, based on the definition of abstract process and schema construction, it is reasonable to argue that the abstraction process does not happen, at least not completely, after recognition. For instance, students are abstracting the characteristics and building schemas while learning how to solve the source problem that will become the analogous problems when they meet a new question. Holyoak (1985) has also claimed the feature of disorder in the process of analogical transfer. Holyoak (1985) said the following:

These steps need not be carried out in a strictly serial order, and they may interact in many ways. For example, during the selection of an appropriate source, a partial mapping with the target is typically required. Moreover, since mapping can be conducted in a hierarchical manner, the process might iterate at different levels of abstraction. (Holyoak)

The third step, mapping, involves finding an appropriate connection between the solutions for source and target problems (Mayer, 1992). In order to probe how to help individuals use the solution procedure from the source problem to solve the new one, Holyoak and Koh (1987) have conducted research to discover the role of structural and superficial similarity in analogical transfer. In their study, subjects were asked to read and summarize one of four versions of the light bulb story, and then to solve a series of problems (e.g., the tumor problem) in which they were asked to list as many solutions as
possible. Finally, those subjects were given a hint suggesting that the light bulb problem could be used to solve the tumor problem, and they were again asked to list possible solutions to the tumor problem. The results indicated that transfer by analogy from the solution procedure of the source problem to the target problem was better if subjects had read a structurally similar version of the source problem story, but superficial similarity did not affect transfer.

Successful schema construction and analogical transfer are both necessary components for the learning and the transfer of mathematics (Chi et al., 1981; R. E. Mayer, 1992; Polya, 1945; Silver, 1981) and statistics (Quilici & Mayer, 1996; Ross, 1987, 1989a). Moreover, these two components are not independent of each other. Schemas can mediate analogical transfer (Holyoak & Thagard, 1989). Based on previous research about each component, it is apparent that structural similarity among problems influences both schema construction and analogical transfer. Identifying the most effective approach for implementing these findings in instruction in order to improve students' learning becomes a very practical and meaningful problem. In the next section, the author discusses related studies focusing on this problem.

2.3 Learning through examples

Learning through examples has been considered an important instructional device, particularly for teaching students in fields such as mathematics (Anderson, 1993; Catrambone, 1994; Chi et al., 1989; Cooper & Sweller, 1987; S. K. Reed, Dempster, & Ettinger, 1985). Research on the use of examples in instructional design has focused on three major issues: the multiplicity of instruction examples, the variability of examples
present during instruction and the role of learners' prior knowledge in learning from various examples (Atkinson, Derry, Renkl, & Wortham, 2000; Guo & Pang, 2011)

2.3.1 Multiple examples

With respect to the quantity of examples, most researchers have agreed that multiple examples are necessary when students are learning complex concepts during instruction (Cooper & Sweller, 1987; Quilici & Mayer, 1996; Spiro, Feltovich, & Coulson, 1992; Sweller & Cooper, 1985). For example, Gick and Holyoak (1983) found that students receiving two examples of a concept exhibited more evidence of structural schema construction than did students who received only one example. Furthermore, Reed and Bolstad (1991) provided empirical evidence through the experiment. Their results concluded that students who were taught with two examples outperformed those who were shown only one example, which indicates that multiple examples can facilitate learning better than a single example.

2.3.2 Varied examples

Because the change of structural features affects a concept and a corresponding solution, how superficial features should be manipulated during instruction to facilitate recognition of structural features attracts researchers’ attention. Consequently, researchers have conducted many studies on the effect of variability of superficial features. However, in what way and to what extent varying multiple examples exerts influence on learning and transfer are still controversial questions. In other words, how similar or different should multiple examples be in terms of superficial and structural features (Guo, Pang, Yang, & Ding, 2012; Renkl, Stark, Gruber, & Mandl, 1998; Rittle-Johnson & Star, 2009)? Should examples be designed with cover stories that are similar
for the same concept or problem type? Alternatively, should examples rely on a varied cover story within a concept or a problem type?

One view has suggested that superficial similarity among examples should be used during instruction of both a concept and procedure (Gick & Holyoak, 1980; Namy & Gentner, 2002; Renkl et al., 1998; Ross, 1989a). Within a new domain, it is highly possible that novices will be overwhelmed by the great number of concepts and types of problems (e.g., basic probability rules include four concepts including additions, multiplications, permutations, and combinations\(^1\)) and the complicated solutions for each. The common solution is to segregate by concepts. In other words, recognizing the problem type is an important skill in the process of learning. Ross (1989b) has suggested one possible means to design a lesson would be to mix up the concepts but use superficial similarity to ensure that learners use appropriate methods. For example, in a lesson on probability, all problems about additions, multiplications, permutations and combinations could be presented with examples about poker cards. Ross (1989b) has presumed that this would force learners to concentrate first on the methods, but in a less segregated environment. In addition, he has demonstrated that superficial similarity helps to deal with the difficulty in determining the relevant distinctions between types and deciding how to categorize problems as a function of these distinctions. This difficulty is often aggravated by focusing learners' attention on the individual types rather than on their commonalities and differences. Still, take probability as an example.

---

\(^1\) The addition, multiplication, permutation and combination rules are fundamental in probability. These rules provide us ways to calculate the probability of a certain event happening. Addition rules include two variations depending on whether events are mutually exclusive. Multiplication rules also include two variations depending on whether events are independent. The number of permutations of \(n\) objects taken \(r\) at a time is denoted by \(nP\). The number of combinations of \(n\) objects taken \(r\) at a time is denoted by \(nC\).
Permutations and combinations are two concepts that are similar but have an important distinction (whether the order of arrangement is important). Having examples in each with similar cover stories might help learners to notice the similarities and distinctions. Ross (1989b) went on to say, "as learners become more capable and confident, they could be weaned away from their reliance on superficial similarities until they are able to categorize the problems by structural aspects only" (p. 464). Moreover, it is often harder to discover the underlying common structure when varied superficial features are presented (Gick & Holyoak, 1980, 1983; Ross & Kennedy, 1990).

Other researchers have also suggested not using examples with varied superficial features (Gentner & Namy, 1999; Namy & Clepper, 2010; Namy & Gentner, 2002). They agreed with Ross but contended that superficial similarity among examples helps learners to notice and align the structural features, and to construct the schema and improve structure mapping, which are important for both learning and transfer. Besides, Renkl, Stark, Gruber, and Mandl (1998) have also supported this suggestion and demonstrated that examples with similar superficial features do not overwhelm the students and can help them to focus on the structural features. Moreover, based on the theory that increased cognitive load may exert negative impact on schema acquisition (Sweller & Cooper, 1985), it would make sense that wide variability of examples, which requires more cognitive demand, would impede the learning progress (Paas & Van Merriënboer, 1994).

On the contrary, another group of researchers have held the opposite opinion; they have suggested that examples with varied superficial features should be used during instruction, because varied superficial features expedite capturing structural features of
different types of problems and concepts, and help induce a schema (Merrill & Tennyson, 1978; Paas & Van Merriënboer, 1994; Quilici & Mayer, 1996; Ranzijn, 1991; S. Reed, 1989). Though Paas and Van Merriënboer (1994) raised the concern that variability of superficial features (the value and problem format in their particular study) during instruction increases cognitive load, they still suggested that variability should be included as long as the extraneous cognitive load is reduced. Extraneous cognitive load pertains to the processes not directly relevant for learning (Sweller & Cooper, 1985). In Paas and Van Merriënboer's study (1994), worked examples were proven to save cognitive demand and exert the positive effects of the variability of superficial features during geometry instruction. The results showed that students who studied worked examples with high variability in superficial features outperformed students who were in the condition of low variability.

Quilici and Mayer (1996) have investigated this approach as well by manipulating the similarity/variability of cover stories of examples for teaching statistical concepts. They designed two distinct sets of examples. One set emphasized superficial features by using very similar/unvaried cover stories for each example of a given concept; the other set emphasized structural features by using different/varied cover stories for each example of a given concept. For example, suppose students are learning probability rules. Three different cover stories (e.g., poker cards, marbles, and dice) are used during instruction for learning an addition rule, and the same three cover stories are adopted for a multiplication rule. In this condition, students are expected to realize that the relevant commonalities of concepts are not superficial features but rather the structural features. This process can facilitate the structure-based schemas of statistical concepts and future
tasks, such as sorting problems into corresponding categories. Their results were consistent with their prediction: students who were given structure-emphasizing examples were more likely to sort problems based on structural features and were less likely to sort based on superficial features compared to students who received surface-emphasizing examples.

It seems that two groups of researchers, Ross (1989b) and Quilici and Mayer (1996) had two opposite opinions; however if we look at the experimental designs from another angle, we realize that this is not true. Quilici's structure-emphasized examples were varied within each concept, but were similar between concepts (Table 1). The surface-emphasized examples were varied between concepts, but similar within each concept. The similar-surface-design mentioned by Ross (1989b) was the instructional design composed of similar cover stories within and between concepts. Though he did not directly clarify whether the varied-surface-design was with examples that varied between or within concepts, it can be inferred through the example (p. 463) given in his book (Ross, 1989b) that the variability exists between concepts. Therefore, precisely speaking, the results from Quilici and Mayer do not completely contradict Ross's opinion. On the contrary, the results reached similar conclusions that examples with varied cover stories between concepts facilitate learning in a lesser way than the other two types of instructional designs. However, the impact on instruction of the other two types – cover stories varied only within concepts, and cover stories that were consistent within and between concepts – still needs to be explored.
Table 1

Comparison of Design and Definition in Ross (1989b) and Quilici & Mayer (1996)

<table>
<thead>
<tr>
<th></th>
<th>Superficial Similarity by Ross</th>
<th>Superficial Variability by Quilici and Mayer (Structure-Emphasized)</th>
<th>Superficial Similarity by Quilici and Mayer (Surface-Emphasized)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Addition Rule</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>example 1</td>
<td>Cards</td>
<td>Spinner</td>
<td>Dice</td>
</tr>
<tr>
<td>example 2</td>
<td>Cards</td>
<td>Cards</td>
<td>Dice</td>
</tr>
<tr>
<td>example 3</td>
<td>Cards</td>
<td>Marbles</td>
<td>Dice</td>
</tr>
<tr>
<td>example 4</td>
<td>Cards</td>
<td>Dice</td>
<td></td>
</tr>
<tr>
<td><strong>Multiplication Rule</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>example 5</td>
<td>Cards</td>
<td>Spinner</td>
<td>Cards</td>
</tr>
<tr>
<td>example 6</td>
<td>Cards</td>
<td>Cards</td>
<td>Cards</td>
</tr>
<tr>
<td>example 7</td>
<td>Cards</td>
<td>Marbles</td>
<td>Cards</td>
</tr>
<tr>
<td>example 8</td>
<td>Cards</td>
<td>Dice</td>
<td></td>
</tr>
<tr>
<td><strong>Permutation Rule</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>example 9</td>
<td>Cards</td>
<td>Spinner</td>
<td>Spinner</td>
</tr>
<tr>
<td>example 10</td>
<td>Cards</td>
<td>Cards</td>
<td></td>
</tr>
<tr>
<td>example 11</td>
<td>Cards</td>
<td>Marbles</td>
<td></td>
</tr>
<tr>
<td><strong>Combination Rule</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>example 12</td>
<td>Cards</td>
<td>Spinner</td>
<td>Marbles</td>
</tr>
<tr>
<td>example 13</td>
<td>Cards</td>
<td>Cards</td>
<td></td>
</tr>
<tr>
<td>example 14</td>
<td>Cards</td>
<td>Marbles</td>
<td></td>
</tr>
</tbody>
</table>

2.3.3 The role of prior knowledge

Although most researchers have believed that students’ prior knowledge in related fields impacts the effect of example variability, there has been no definitive conclusion about its influence on students' learning and transfer (Guo et al., 2012; Rittle-Johnson & Star, 2009). Some researchers (Quilici & Mayer, 1996) have argued that lower-performing students can benefit more from structural-emphasizing examples than higher-performing students, because lower-performing students have more difficulty in categorizing test problems, and examples with emphasized structural features help to reduce this kind of difficulty. Some researchers have found that learners with higher levels of prior knowledge benefit more from high-variability examples, but learners with lower levels of prior knowledge benefit more from low-variability examples (Große &
Moreover, some researchers have claimed that students with lower levels of prior knowledge would benefit more from less complex contrasting examples than from highly varied examples (Holmqvist, Gustavsson, & Wernberg, 2007; Schwartz & Bransford, 1998). Unlike other previous researchers, some researchers have not detected an interaction effect between students' prior knowledge and the variability of examples (Renkl et al., 1998). In a nutshell, it is still debatable how learners' varying levels of prior knowledge affect the usage of example variability. However, it raises an important concern for the author, that learners' prior knowledge should be controlled when conducting related experiments.

### 2.4 Conceptual framework of this dissertation project

Previous studies have concluded that schema construction and analogical transfer are two indispensable components of mastering knowledge that mutually promote each other (Donovan, Bransford, & Pellegrino, 1999). In order to foster these two components and successfully solve new problems, recognizing and abstracting structural features of both source and target problems, and mapping a solution procedure based on structural similarity of both, are necessary (Figure 1). Figure 1a illustrates how these two components and steps interact in the process of learning and solving problems. In order to successfully map the correct solution to a target problem, individuals need to precisely recognize the most appropriate source problem that is relevant to the target problem. Each type of problems can be presented in various versions with different superficial features but the same structural features (Figure 1b). By abstracting the shared structural features, structure-based schema is constructed. With a high-quality structure-based schema, individuals have a deep understanding of the structure of certain types of
problems, which ensures the precision of the recognition and in turn prompts the analogical transfer.

Learning by worked examples is a critical and effective technique for teaching mathematics and mathematically-based curriculum (Sweller & Cooper, 1985). But in what way and to what extent the variability of multiple examples should be utilized for instruction is still controversial. Some researchers have suggested that superficial similarity among examples should be used during instruction (Gick & Holyoak, 1980; Namy & Gentner, 2002; Renkl et al., 1998; Ross, 1989a), since controlling superficial similarity can be helpful for students to notice and align the structural features, as well as reduce cognitive load (Paas & Van Merriënboer, 1994). On the contrary, other researchers have believed that examples with varied superficial features can expedite identifying structural features of different types of problems and induce schemas (Merrill & Tennyson, 1978; Paas & Van Merriënboer, 1994; Quilici & Mayer, 1996; Ranzijn, 1991; Reed, 1989).
Figure 1. Analogical Transfer & Schema Construction
Although each aspect is supported by multiple studies, they are not entirely parallel. For example, the studies adopted different measures. Some studies examined the influence of the variability of superficial features on sorting problem types (Namy & Gentner, 2002; Quilici & Mayer, 1996; Rabinowitz & Hogan, 2001; Ranzijn, 1991) and some focused on retrieval (Holyoak & Koh, 1987) or access (Ross, 1987). Measuring the influence on successfully solving a problem is imperative, but it was employed in a limited way. This type of influence is what the author measures in her research. In addition, researchers have conducted studies in different domains; most researchers have focused on mathematical or mathematically-based subjects (Paas & Van Merriënboer, 1994; Reed, 1989; Renkl, Stark, Gruber, & Mandl, 1998), while some chose non-mathematically-based subjects (Holyoak & Koh, 1987; Ranzijn, 1991). Of studies about mathematically-based curriculum, few of them have concentrated on statistics (Quilici & Mayer, 1996; Ross, 1987, 1989a), in which the author is interested.

Furthermore, in carefully comparing studies of Quilici and Mayer (1996) and Ross (1987), the author notices that the definition of the variability of examples' superficial features is not exactly same. As Ross described, superficial similarity meant the cover story was unvaried within and between concepts (Table 1). The superficial variability was between concepts rather than within. However, in Quilici and Mayer's study (1996), superficial variability was addressed in the structural-emphasizing techniques where each concept was exemplified by a battery of cover stories that differed from one another and the same battery of cover stories was used across concepts. And the superficial similarity was within concepts, which corresponded to the superficial variability defined by Ross. Distinct practical definition has affected the experimental
design. Ross (1989b) compared unvaried cover stories among all examples of mixture concepts with varied cover stories between different concepts. Quilici and Mayer (1996) compared the condition under which cover stories did not vary between concepts but varied within concepts with the condition under which cover stories varied between concepts, but did not vary within a concept. In other words, these two groups of researchers used two different conditions to compare the same condition (Table 1). Therefore, it is not warranted to claim that these two studies fully contradicted each other. In fact, the instructional suggestion by Ross (1989b) does not have direct empirical evidence and was based on his previous studies (1987, 1989a) in which he manipulated the superficial similarity of test problems (target problems) instead of instructional examples (source problems), to investigate how similarity between a new and an earlier problem affects the access and use of the earlier problem. Hence, in the dissertation study, the author fills in the gaps in the literature by: 1) verifying Ross’s suggestion on instruction and 2) comparing three instructional methods that are discussed above, including the two (learning with examples with cover stories not varied between concepts but varied within concepts, and examples with cover stories not varied among all mixture concepts) that were not compared in previous studies, and exploring which one better improves students’ learning of probability.

Previous research has probed the effect of students’ prior knowledge (Große & Renkl, 2007; Holmqvist, Gustavsson, & Wernberg, 2007; Quilici & Mayer, 1996; Renkl, Stark, Gruber, & Mandl, 1998), but there was no certain conclusion of the role it played (Guo et al., 2012; Rittle-Johnson & Star, 2009). However, the influence of the difficulty of learning content on the effect of examples’ variability has been overlooked. Previous
studies show that students who were given difficult source problems had better analogical transfer (Didierjean & Nogry, 2004; Gick & McGarry, 1992). Moreover, Didierjean and Nogry (2004) have proven that difficulty induces students to build a more abstract representation of the solution. Therefore, it is natural to reason that for easy content involving no abstract concepts or complex procedure, students may gain the same benefit from well-structured and not well-structured instruction; however, for difficult content, students may benefit more from well-structured instruction. In the dissertation study, the author explores the role that content difficulty plays in the influence of examples’ superficial variability and hypothesizes that the pattern of influence is stronger for difficult content than for easy content.

The author believes that although the research focuses on the learning of probability, the findings can be applied to the general design of instruction for various domains, particularly for mathematically-based ones. Better-structured instructional design and learning strategy can bring about more effective and efficient learning.
3 Pilot Studies (Design and Preliminary Findings)

3.1 Pilot one

In the first pilot study, the author explored three major questions:

1) whether learning through multiple examples with a similar cover story among all mixtures of problem types improves students’ performance in probability more than learning through multiple examples with varied cover stories between problem types; 2) whether learning through examples with cover stories not varied between concepts, but varied within problem types improves students’ performance in probability more than learning through multiple examples with a similar cover story among all mixtures of problem types; 3) whether the superficial variability of examples exerts different patterns of effect on problem types at varying difficulty levels. The dependent variable of this experiment was the learning performance of probability rules. The independent variables were the variability of cover stories for examples and the difficulty level of learning content.

In order to probe the above research questions, the author developed three conditions (Table 2). In the first condition (named as the consistent-surface condition, CS), four probability theories—addition, multiplication, permutation, and combination rules—were taught through multiple examples with similar poker-cards-related cover stories. In the second condition, superficial variability existed within rules, so this condition was named varied-surface-within-rule condition (VSWR). In this condition, cover stories for each example of a given rule only varied within that rule. On the contrary, in the third condition—varied-surface-between-rules (VSBR) condition, cover
stories for each example of a given rule were not changed. However, cover stories varied for different rules.

Table 2

*Three Conditions*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Consistent Surface Features</th>
<th>Varied Surface Features Within Rules</th>
<th>Varied Surface Features Between Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition Rule</td>
<td>Example 1: Cards</td>
<td>Spinner</td>
<td>Dice</td>
</tr>
<tr>
<td></td>
<td>Example 2: Cards</td>
<td>Cards</td>
<td>Dice</td>
</tr>
<tr>
<td></td>
<td>Example 3: Cards</td>
<td>Marbles</td>
<td>Dice</td>
</tr>
<tr>
<td></td>
<td>Example 4: Cards</td>
<td>Dice</td>
<td>Dice</td>
</tr>
<tr>
<td>Multiplication Rule</td>
<td>Example 5: Cards</td>
<td>Spinner</td>
<td>Cards</td>
</tr>
<tr>
<td></td>
<td>Example 6: Cards</td>
<td>Cards</td>
<td>Cards</td>
</tr>
<tr>
<td></td>
<td>Example 7: Cards</td>
<td>Marbles</td>
<td>Cards</td>
</tr>
<tr>
<td></td>
<td>Example 8: Cards</td>
<td>Dice</td>
<td>Cards</td>
</tr>
<tr>
<td>Permutation Rule</td>
<td>Example 9: Cards</td>
<td>Spinner</td>
<td>Spinner</td>
</tr>
<tr>
<td></td>
<td>Example 10: Cards</td>
<td>Cards</td>
<td>Spinner</td>
</tr>
<tr>
<td></td>
<td>Example 11: Cards</td>
<td>Marbles</td>
<td>Spinner</td>
</tr>
<tr>
<td>Combination Rule</td>
<td>Example 12: Cards</td>
<td>Spinner</td>
<td>Marbles</td>
</tr>
<tr>
<td></td>
<td>Example 13: Cards</td>
<td>Cards</td>
<td>Marbles</td>
</tr>
<tr>
<td></td>
<td>Example 14: Cards</td>
<td>Marbles</td>
<td>Marbles</td>
</tr>
</tbody>
</table>

Since each of the addition and multiplication rules has two formulas according to different types of events (e.g., addition rule includes mutually exclusive and not mutually exclusive events; multiplication rule includes dependent and independent events), each of these two rules was taught through four examples, two examples for each type of event. The four rules had varying difficulty levels: additions and multiplications were considered more fundamental than permutations and combinations, as the understanding of the later two rules had to be based on the understanding of the former two rules; permutations was considered easier and more foundational than combination, so
permutations was always taught before combination in class. Because of this, the difficulty level of learning content was naturally manipulated.

As discussed before, structure-based schema construction plays an essential role in learning and solving new problems. While learning through examples, each type of problem is taught with source problems, which are composed of both superficial features exhibited by a diamond frame and structural features exhibited by a star frame in Figure 2. To better understand a given type of problem, individuals need to abstract the structural features from the source problem so that they can construct appropriate structure-based schemas. Subsequently, appropriate structure-based schemas will be used to sort more problems in correct categories. In that way, they can access and remember proper solutions corresponding to different problem types. In this particular study, the author intervened in the process of structure-based schema construction by manipulating the superficial similarity—the cover stories of examples.

Figure 2. The Learning Process Through Examples
Figure 3. Different Processes of Schema Construction in Three Groups
For illustration purposes, Figure 3 only sketches the instructional design of the addition rule and multiplication rule. In the VSWR condition, each type of problem was exemplified by the same battery of different cover stories. It was expected that, individuals are able to realize that the superficial features, like the cover stories, do not matter and what they need to focus on to solve this type of problem is the structure of the problem. Another way of saying this is that the deep structural features of the problems are less likely confounded with the problems' surface features. Moreover, the varied cover stories provide individuals more chances to practice "ignoring" the cover story altogether. Once individuals understand the structural features and constructed schemas based on those features, the boundary between different types of problems would become clear and corresponding schemas would become appropriate. Therefore, when these individuals encountered problems, they can differentiate them more skillfully and find the correct solutions more precisely.

In the CS condition, four rules were taught with multiple examples that were embedded into the same cover story. Based on Ross's theory (1989b), by providing the same cover stories, students could conserve attention for determining relevant distinctions between problem types and for focusing on the structural features. The structural features in this condition are also not likely be confounded with the problems' surface features. However, participants in this condition do not get a chance to practice how to distinguish structural features from superficial features. As a result, their schemas and understanding of each type of problem are not as thorough as in VSWR condition. When coming across new problems, they could not recognize the corresponding problem types and solutions as quickly and precisely as in VSWR condition.
By the same token, in the VSBR condition, the same cover story used four times within each rule hinders the participants from determining structural features. Moreover, it was possible that the students were misled to associate the cover story with certain types of problems by being given various cover stories for different rules. In other words, the structural features of problems were more easily confused with superficial features.

Based on previous research and theories, the author has developed four hypotheses.

Hypothesis 1: Students in the CS condition, learning through examples with similar cover stories among different probability rules, outperform students in the VSBR condition, learning through examples with cover stories varied between rules.

Hypothesis 2: Students in the VSWR condition, learning through examples with cover stories varied within rules, outperform students in the CS condition, learning through examples with similar cover stories.

Hypothesis 3: The pattern of influence of superficial variability of examples is stronger for difficult probability rules than easy ones.

Hypothesis 4: Students in the VSWR condition, learning through examples with cover stories varied within rules, have a better transfer rate than students in the other two groups.

The first three hypotheses were tested in the first pilot study and the fourth hypothesis was tested in the second pilot study.

3.1.1 Methodology
Participants. 39 graduate students (13 for each group), including eight males and 31 females, were recruited from Teachers College, Columbia University and randomly assigned to three groups/conditions.

Procedure. All participants sequentially completed the personal information survey, pretest, instructional intervention, posttest, and feedback of the instruction and learning, requiring approximately one hour in total. Except for the instruction section, materials for the other sections were the same across conditions. During the instruction, participants read worked examples to learn the probability rules.

Material. The personal information survey contained questions used to collect participants’ gender, major, and familiarity with probability theory and with the contexts of cover stories presented in this study. Before exposure to instructional worked examples (the intervention), the author tested participants’ ability to solve probabilistic problems. The pretest aimed to examine participants’ ability to solve probabilistic problems. The test was composed of seven multiple-choice questions with multiple cover stories. Take this question for the addition rule as an example: Peter is randomly picking a card from a standard deck of 52 playing cards. What is the probability of his choosing a queen or a heart? The answer options included \( \frac{15}{52}, \frac{16}{52} \) (the correct answer), \( \frac{17}{52} \), and \( \frac{18}{52} \). Questions on permutation and combination had two correct answers. One was in the form of a formula \( \frac{1}{52} P_3 \) and the other one was in the form of a fraction \( \frac{1}{52} \frac{1}{51} \frac{1}{50} \).

In the instructional phase, worked examples were presented on PowerPoint slides (Appendix G) with well-designed animations and images corresponding to the cover
story of each example. According to Paas and Van Merriënboer (1994), using worked examples can balance the cognitive load caused by problem variability, thus, they provide more possibility to add variability and, in turn, to enhance schema acquisition. The learning process was entirely self-paced. By tapping the right- and left-arrow-button on the keyboard, participants could view the next or previous solution steps. In order to improve engagement, researchers suggested participants take notes during their study. After the instructional phase, participants were required to finish a posttest to see which condition improved their ability to solve problems.

The posttest was identical to the pretest, containing seven questions as well. However, the order of answer options for a given question changed. The last section consisted of several Likert-scale questions to examine clarity, helpfulness, and effectiveness of the instruction through participants’ self-report.

3.1.2 Results and Discussion

Does participants’ performance improve? The basic idea of this study was to see the score difference between pretest and posttest to discover whether and how much participants’ performance on probabilistic questions could be improved by intervention. The paired-samples t-test was conducted to test the effects of instruction on participants’ ability to solve probabilistic problems. The results suggested that students’ performance after learning significantly increased, \( t(38) = 6.401, p < 0.001 \), which means the instruction effectively improved participants’ performance on such problems (Figure 4 & 5).
Figure 4. Participants' Performance Accuracy in Pretest and Posttest

Figure 5. Performance Accuracy in Pretest and Posttest of Three Groups

Are the groups equivalent in prior ability? An analysis of variance (ANOVA) conducted on the pretest scores of each participant indicated that the three groups did not differ significantly on prior knowledge of probability (Mean of accuracy = .59, .46, and .48 for similar cover stories (CS condition), varied cover stories within rules (VSWR
condition), and varied cover stories between rules (VSBR condition) respectively, $F (2, 36) = 2.062, p = .142$ (Figure 6).

![Figure 6. Equivalent Prior Knowledge Across Groups](image)

**Does participants’ improvement differ by condition?** The accuracy improvement was obtained by deducting the pretest accuracy from the posttest accuracy. Results from one-way ANOVA illustrated that the difference in accuracy improvement between groups was marginally significant, $F (2, 36) = 3.092, p = .058$ (Figure 7). Considering that small sample size might limit significance value and the effect size was large ($\eta^2 = .147$), the author still did a post-hoc comparison. The finding showed that the accuracy improvement in the VSWR condition ($M = .27, SD = 15.5$) was significantly higher ($F (1, 36) = 6.088, p = .018$) than improvement in CS condition ($M = .12, SD = 17.3$), which was consistent with the second hypothesis that VSWR condition would outperform CS condition. However, there was no significant difference between VSWR and VSBR conditions, which was not consistent with the findings from the work of Quilici and Mayer (1996). In addition, the researcher did not find the difference between
the CS condition and the VSBR condition, which was not consistent with the author’s expectations based on Ross’s suggestion (1989b) (Hypothesis 1).

**Figure 7. Accuracy Improvements Across Groups**

The results confirmed Hypothesis 2. However, the results did not exhibit the expected disadvantage of the VSBR condition compared to the others. The author's interpretation of such results was likely due to the small sample size, which was not sufficient enough to show a significant difference. Moreover, the author noticed that the questions for multiplication in the test were a bit of confusing, which might also have affected the general results.

*Does the pattern differ by content difficulty?* Since the researcher expected that varying difficulty levels naturally exist among the four probability rules, she examined the group difference in performance in each rule. The significant difference in accuracy improvement was only found in permutations (Figure 8), $F (2, 36) = 4.486, p = .018$. The improvement in varied cover stories within rules (the VSWR condition) ($M = .35$,
SD = 21.7, \( p = .007 \)) and between rules (the VSBR condition) (M = .29, SD = 26.7, \( p = .032 \)) was significantly higher than in similar cover stories (the CS condition) (M = .08, SD = 23.7). However, there was no significant difference between the VSWR condition and the VSBR condition. In other words, the intervention was only evident in the permutation rule that was considered a relatively difficult concept. The resulting pattern involving the VSWR condition and the CS condition fully coincided with the one in the general comparison. Moreover, the difference tendency between the VSBR condition and the CS condition became significant here. In order to draw the conclusion that the findings were consistent with the third hypothesis that the difficulty level of content would affect the strength of the influence from superficial variability, more analyses were conducted.

*Figure 8. Accuracy Improvements for Permutation Questions across Groups*

It is possible that a difference in the quality of instruction caused VSWR and VSBR to perform better than CS on permutation. For example, instructions for
permutations in VSWR and VSBR were better than instructions for the other rules in these two groups. Or instructions for permutations in CS was worse than instructions for the other rules in CS. To test this possibility, the author analyzed participants’ feedback. As mentioned before, participants rated the helpfulness of instruction for each rule in the last experimental section. A Kruskal-Wallis H test was used to examine whether there was a difference in helpfulness of instruction for rules. The results did not display significant differences across rules in CS \( (\chi^2(3) = 1.908, p = .592) \) and VSWR \( (\chi^2(3) = 6.918, p = .075) \), which means participants in these two groups considered instruction for each rule equally helpful. In VSBR condition, significant difference in helpfulness across rules was found \( (\chi^2(3) = 23.636, p < .001) \) (Figure 9): ratings for additions \( (Mdn = 5) \) and multiplications \( (Mdn = 5) \) were both significantly higher than permutations \( (Mdn = 4) \) and combinations \( (Mdn = 4) \). Accordingly, instruction for permutation in the VSBR condition was not viewed as more helpful than other rules either. Therefore, the findings disproved the likelihood of better instruction for permutations in VSWR and VSBR condition and worse instruction for permutations in CS condition.

**Figure 9. Comparisons of Helpfulness for Instructions of Each Rule**
Next, the author compared the performance accuracy of the four rules across groups (Figure 10). This accuracy was no longer learning gain (pretest minus posttest), but the average performance accuracy of pretest and posttest. Significant differences across rules were found (Welch's $F (3, 83.949) = 11.862, p < .001$): performance accuracy of combinations ($M = .465, SD = 24.1$) was lower than that of additions ($M = .679, SD = 28.1$), multiplications ($M = .808, SD = 27.2$), and permutations ($M = .644, SD = 21.2$), while the performance of permutations was lower than that of multiplications. It is reasonable to believe that the performance accuracy was related to the difficulty of the concepts. Naturally, when difficulty increases, accuracy drops. Therefore, the varied performance on corresponding rules supported the assumption that varying difficulty level existed among the four rules. Accordingly, the combination rule was the most difficult among the four rules. As the addition rule and multiplication rule were conceptually similar to each other and no significant difference in accuracy was found between them, these two rules were considered easy concepts among the four rules. Also given that conceptually the permutation rule substantially resembled the combination rule, but was distinct from addition and multiplication, and its accuracy was significantly higher than the combination rule, the permutation rule was identified as a moderately difficult concept among the four rules.

In a nutshell, the third hypothesis that the influential pattern of superficial variability was stronger for difficult concepts than for easier ones was confirmed. Moreover, the effectiveness of the intervention was best with the moderately difficult concept (e.g., permutations).
Figure 10. Overall Performance Accuracy of Each rule

3.1.3 Conclusions

Based on all of the above statistical evidence, the hypotheses were partially
confirmed. Learning through examples with consistent cover stories among different
probability rules and learning through examples with cover stories varied within rules
exerts influence at disparate levels. The latter approach is found to better facilitate
learning. However, Ross’s suggestion (1989b) was not validated in this study, as there
was no difference between learning through examples with similar cover stories and with
cover stories varied between rules. Moreover, the difference between examples with
cover stories varied within and cover stories varied between rules was not found as
significant as in Quilici and Mayer’s results (1996). Such patterns also proved stronger
for permutations, which was moderately difficult to learn.

3.1.4 Limitations and Improvements of the Experiment Design

Although part of the results was not consistent with what was expected, the author
could not assert that the hypotheses based on previous studies had been proven wrong, as
the first pilot study had limitations, which might affect the significance level of results. First of all, the sample size was small; thus, some trends might not be detected at a statistically significant level. Therefore, in the second pilot study, the author recruited a larger number of participants to increase the sample size. Moreover, some experimental materials needed to be refined. For example, the wording of one multiplication question was determined to be confusing to participants; therefore, in the second pilot study the author refined that question. In order to guarantee the clarity of all refined materials, the author asked both experts and peer students to approve them before conducting additional experiments. Thirdly, although it is reasonable to believe that varying difficulty levels naturally exist among four rules, there was no direct examination of participants’ opinion around that. Thus, in the next study, a relevant survey was added. Last but not least, in the first pilot study, the author did not test the influence from superficial variability on transfer, which is also important and might be affected by superficial variability. Therefore, the group difference in learning transfer was an additional research question in the next study.

3.2 Pilot two

There were three major purposes for the second pilot study. First of all, the author aimed to verify the findings in the first pilot study. Secondly, in the first pilot study, the author did not directly test the difficulty level of four rules. Though varying difficulty level exists among the four rules conceptually and practically – addition and multiplication rules are fundamental for permutation and combination rules, and the combination rule is more advanced than the permutation rule – there was no evidence that students considered the difficulty level among the four rules in the same way. Therefore,
in the second study, the author aimed to directly test students' understanding of the difficulty level among the four rules. Thirdly, the author aimed to examine the difference in transfer that might be caused by the superficial variability of learning examples.

Based on previous research and theories that have been discussed in the previous sections, the author kept the same four hypotheses raised in the first study.

3.2.1 Methodology

Participants. 45 graduate students (13 for each condition) were recruited from Teachers College, Columbia University and Fordham University. They were randomly assigned to three conditions.

Procedure. Unlike in the first pilot study, there were two sessions separated by one week in the second pilot study. In the first session, the whole procedure was nearly the same as in the first pilot study. In the instruction, participants still read the worked examples to learn the probability rules. But the task sequence changed for the feedback of the instruction and learning. In the first pilot study, the self-reported feedback was asked to complete after the posttest. However, considering possible bias caused by delay, the author asked participants to provide feedback on the instruction and learning for each rule right after they finished the corresponding section (Figure 11). In addition, to directly testing students’ understanding of the difficulty level of the four rules, the author added one more 5-point Likert scale in the feedback survey. All participants were asked to rate how difficult the rule that they learned was, with that scale. While they were learning with examples, participants could take notes. However, they could not keep the notes when they worked on exams, including the posttest, transfer test, and delayed transfer test. After finishing the posttest, participants were required to complete a transfer test.
The whole session took about 90 minutes. One week after they finished the first session, participants needed to come back for the second session for a delayed transfer test, which required no more than 30 minutes to complete.

Figure 11. The Task Flow in Two Pilot Studies

**Materials.** In the second pilot study, the same pretest, posttest, and instructional materials as in the first pilot study were used. The author refined the confusing multiplication questions and added one conceptual question for each rule in both pretest and posttest (e.g., please select the correct statement(s) about the Addition Rule). The transfer test was also composed of multiple-choice questions. The cover stories for questions in the transfer test were never used in the pretest, posttest, or instruction.
Moreover, unlike in previous tests or instruction, some questions in the transfer test were decimal instead of fractional (Table 3). The materials for the delayed transfer test were identical to the ones used in the transfer test.

Table 3

* Differences in Design for Two Pilot Studies *

<table>
<thead>
<tr>
<th></th>
<th>Pilot 1</th>
<th>Pilot 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants</strong></td>
<td>TC only (39)</td>
<td>TC(34) &amp; Fordham (11)</td>
</tr>
<tr>
<td><strong>Feedback</strong></td>
<td>After posttest</td>
<td>During learning</td>
</tr>
<tr>
<td><strong>Difficulty of Concept (self-report)</strong></td>
<td>✔</td>
<td>✗</td>
</tr>
<tr>
<td><strong>Learning Content</strong></td>
<td>Same</td>
<td>Same</td>
</tr>
<tr>
<td><strong>Instructional Method</strong></td>
<td>Same</td>
<td>Same</td>
</tr>
<tr>
<td><strong>Pretest</strong></td>
<td>✗ (N=7)</td>
<td>• Refined the wording of Multiplication question</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Added 4 more conceptual questions (N=8+4)</td>
</tr>
<tr>
<td><strong>Posttest</strong></td>
<td>✗</td>
<td>• Refined the wording of Multiplication question</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Added 4 more conceptual questions (N=8+4)</td>
</tr>
<tr>
<td><strong>Transfer Test</strong></td>
<td></td>
<td>✗</td>
</tr>
<tr>
<td><strong>Delayed Transfer</strong></td>
<td></td>
<td>✗</td>
</tr>
</tbody>
</table>

**3.2.2 Results and Discussion**

*Does participants’ performance improve?* Similarly, the paired-samples t-test was used to check whether there were performance differences between the pretest and posttest. The result indicated that students’ performance was significantly improved in the posttest rather than in the pretest, \( t(44) = 10.439, p < 0.001 \), which means the instruction was effective (Figure 12 & 13).
Figure 12. Participants' Performance Accuracy in Pretest and Posttest in Pilot 2

Figure 13. Performance Accuracy in Pretest and Posttest of Three Groups in Pilot 2
Are the groups equivalent in prior ability? An analysis of variance (ANOVA) conducted on the pretest scores of each participant indicated that the prior knowledge of probability (Mean of accuracy = .59, .44, and .65 for similar cover stories (the CS condition), varied cover stories within rules (the VSWR condition), and varied cover stories between rules (the VSBR condition) respectively was significantly inequivalent, $F(2, 42) = 24.448, p < 0.001$ (Figure 14). Therefore, the unadjusted performance difference between the posttest and pretest could not be used to indicate the effect of the instructional method in three groups. In other words, the same approach as in the first pilot study—comparing learning gains of each group, was not appropriate in this pilot study.

![Pilot 2: Pretest across Groups](image)

**Figure 14. Inequivalent Prior Knowledge across Groups in Pilot 2**

Does participants’ improvement differ by condition? In order to control the undesired effect from nonequivalent prior knowledge, the one-way analysis of covariance (ANCOVA) was employed (accuracies of the pretest was the covariate).
To test the first hypothesis, the author compared the posttest accuracies of the CS and VSBR conditions while controlling both groups’ pretest accuracies. After adjustment for pretest, there was a marginally significant difference in posttest between the CS and VSBR condition, $F(1, 27) = 3.491, p = .073$, partial $\eta^2 = .114$. This finding supported the first hypothesis that students learning through examples with similar cover stories among different probability rules, outperform students learning through examples with cover stories varied between rules, which was consistent with Ross’s suggestion (1989b). However, the expected difference in posttest between CS and VSWR condition was not found, $F(1, 27) = 1.17, p = .735$, partial $\eta^2 = .004$. That means the second hypothesis was not confirmed, which was not consistent with the finding in the first pilot. The small sample size in both studies was the most likely reason for such erratic findings.

![Figure 15. The Performance Improvement in Three Groups in Pilot 2](image)
Do students think the difficulty of the four rules is different? In the feedback survey, participants were asked to rate the difficulty level of each rule. Epsilon (ε) was 0.756, as calculated according to Greenhouse and Geisser (1959), and was used to correct the one-way repeated measures ANOVA. The rating for the difficulty level was significantly different for different rules, $F(3, 132) = 60.582, p < 0.001$. The addition rule was the easiest ($M=1.87, SD = .919$), compared with the multiplication rule ($M=2.47, SD = 1.079, p = 0.028$), permutation rule ($M=3.44, SD = 1.324, p < 0.001$), and combination rule ($M=3.64, SD = 1.358, p < 0.001$). Moreover, the multiplication rule was significantly easier than the permutation ($p < 0.001$) and combination rule ($p < 0.001$). However, there was no significant difference between the permutation rule and combination rule ($p = .358$). To sum up, the addition rule was considered the easiest; the multiplication rule was less easy; the permutation and combination rules were both the most difficult.

Does the pattern differ by content difficulty? According to the difficulty ratings, the performance of permutation and combination rules was combined as the difficult rules. Within each level of rules, the author applied the same methods for general comparison to probe whether the influential pattern worked differently for different rules. Unfortunately, the author did not find significant differences in the posttest performance among groups for any level of the rules. This means the third hypothesis – that the pattern of influence from the superficial variability of examples will be stronger for difficult probability rules than easy ones – was not confirmed, which was not consistent with the finding in the first pilot. However, the trends within the difficult rules (permutations and
combinations) were more strongly consistent with the pattern discovered in general comparison than the trends within the easier ones (Figure 16).

Figure 16. The Performance Improvement of the Difficult Rule in Three Groups in Pilot 2

*Does learning transfer differ by condition?* To answer this question, the one-way ANCOVA was conducted to test participants’ scores on the transfer test and the delayed transfer test across groups, while controlling their pretest scores. Unfortunately, the participants’ performance was not different in transfer ($F(2, 41) = .432, p = .652$) and delayed transfer ($F(2, 41) = 2.241, p = .119$). Therefore, the fourth hypothesis was not proved. After looking further at the scores, the author realized that the test for transfer (Figure 17) and delayed transfer (Figure 18) was too easy, which might have been the main reason for the similar performance across groups. Moreover, the author found that students’ performance in the transfer test and the delayed transfer test was almost the same. Thus, in future studies, it is reasonable to consider only testing the delayed transfer.
3.2.3 Conclusion

In the second pilot study, the first hypothesis was marginally proven; it showed that examples with similar cover stories among different probability rules could improve
students’ learning performance, more than examples with cover stories varied between rules. This finding was not discovered in the first pilot study. However, in the second pilot study, the author did not get significant evidence that was found in the first pilot study for supporting the second and third hypothesis (Table 4). The author suggests that the primary reason for such inconsistent results was the small sample size, which limited the significance of difference and affected the equivalence of participants. However, one conclusion could be drawn based on those findings that students always learn better in VSWR condition, where examples were presented with cover stories that varied only within each rule. For the learning transfer, although there were no significant differences found in the study, the author would not draw a strong conclusion that the variability of superficial similarity of examples does not influence learning transfer because of limitations with the design of the study.

Table 4

Comparisons of the Results in Two Pilot Studies

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Pilot 1</th>
<th>Pilot 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Learning gain, G1 &gt; G3</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>H2: Learning gain, G2 &gt; G1</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>H3: The pattern works stronger for difficult rules</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>H4: Transfer, G2 &gt; G1 &amp; G2 &gt; G3</td>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

3.2.4 Limitations and Improvements of the Experiment Design

In two previous pilot studies, the author discovered several interesting findings but also recognized limitations in the design. Those limitations have been improved in the dissertation study.
The first issue was the small sample size that could affect the representativeness and equivalence of participants and statistical results. In the dissertation study, the author enlarged the sample size and strictly randomized group assignment.

In addition, there were limitations in exam materials. To test students’ learning improvement, the author asked them to complete a pretest and a posttest. The exam questions for these two tests were identical, which might result in practice effect. In order to avoid such effect in the dissertation study, the author changed the posttest and made it a test parallel to the pretest. For example, the order of questions and answers in the posttest would be different from the order in the pretest. The numbers in the questions would be different from the corresponding ones in the pretest. Besides, the transfer test was too easy, which might affect the reflection of differences across groups. To improve that, the author enhanced the difficulty of the exam in several ways. For example, the author would increase the number of answers and complexity of problem contexts.
4 Dissertation Study Overview

In the two pilot studies, two conflicting results were discovered. In the first pilot study, students in the CS condition performed the worst and VSBR condition performed in the middle. However, in the second pilot study, the performance of students in these two groups completely reversed. Therefore, in the dissertation study, the author intended to find which of the results were correct, with a larger participant pool. Another unexpected result in the two pilot studies was the similar performance on transfer tests across conditions. The inappropriate difficulty level of the exam might have been the problem. Thus, in the dissertation study, the author increased the difficulty of the transfer tests to get more precise results. As students’ performance of the transfer test and delayed transfer test were very similar, the author excluded the delayed transfer test. Instead, a delayed posttest was included to test how the intervention would affect students learning over a long-term period. Last but not least, the author realized that reading only the solution procedure of worked-examples without practice was not enough for learning. Therefore, in the dissertation study, a practice section was added during the instruction, which was expected to emphasize the effects of the independent variable — superficial variability of examples — on learning. In the following sections, the author elaborates on the final hypotheses, revised methodology, results analyses and corresponding discussion.
5 Research Questions and Hypotheses for the Dissertation Study

As similar to the previous pilot studies, the author intended to discover the relationship between superficial variability of examples and students’ learning and transfer performance of probability rules as well as the possible different influential patterns for problem types of varying difficulty levels. The dependent variables were students’ learning performance in the two posttests (the posttest and delayed posttest), and transfer performance. The independent variables were variability of examples’ cover stories (CS, VSWR, and VSBR) and the difficulty level of problems (easy rules and difficult rules).

Research Question 1: Will the three types of the superficial variability of examples result in different learning performance in the short term and in the long term and in different learning transfer? Which type of variability will lead to a better outcome and which type will lead to a worse outcome?

Research Question 2: How difficult do students consider the rules? Does their perceived difficulty of each rule coincide with the actual difficulty of the rules?

Research Question 3: Will the pattern of the influence of the superficial variability of examples be different for problem types at varying difficulty levels? Will that influence be stronger for more difficult problem types (rules)?

Based on previous research (Gick & Holyoak, 1980, 1983; Paas & Van Merriënboer, 1994; Ross 1989b; Quilici & Mayer, 1996) and the pilot studies, three hypotheses have been raised.

Hypothesis 1: In the posttest, the delayed posttest, and the transfer test, students in the VSWR condition, learning through examples with cover stories varied within rules,
and in the CS condition, learning through examples with consistent cover stories, outperform students in the VSBR condition, learning through examples with cover stories varied between rules.

Hypothesis 2: In the posttest, the delayed posttest, and the transfer test, students in the VSWR condition, learning through examples with cover stories varied within rules, outperform students in the CS condition, learning through examples with consistent cover stories.

Hypothesis 3: The pattern of the effect of the superficial variability of examples is stronger for difficult types of problems than for easy ones.
6  Methodology

6.1 Participants

100 college students in the Teachers College, Columbia University, participated in this study. Limited or no prior knowledge of probability was necessary. Three cases were dropped from the sample because of an unstandardized experimental operation. So the final sample included 97 participants.

6.2 Design and Manipulation

There were three conditions in this study (Table 2). In the first condition (named as the consistent-surface condition), four probability theories — addition, multiplication, permutation, and combination rules — were taught through multiple examples with similar poker-cards-related cover stories. In the second condition, superficial variability existed within rules, so this condition was named varied-surface-within-rule condition. In this condition, cover stories for each example of a given rule only varied within that rule. On the contrary, in the third condition—varied-surface-between-rules condition, cover stories for each example of a given rule did not change. However, cover stories varied for different rules.

The difficulty level of problem type was manipulated by using four probability rules: additions, multiplications, permutations, and combinations. Based on their conceptual and practical definitions, the addition and multiplication rules were more fundamental than the permutation and combination rules. Moreover, the multiplication rule was more complex than the addition rule. Though it was reasonable to draw assumptions that the addition rule was easier than the multiplication rule and both of them were easier than the permutation and combination rules, the manipulation was
checked through a self-reported survey where participants were asked to rate the difficulty of each rule (Appendix E).

6.3 Procedure

Data were collected in two sessions separated by one week. In total, two sessions took about 80-100 minutes. (Figure 6.3)

Session 1 (70-85 minutes):

1. Participants’ personal information was surveyed (Appendix A).
2. Pretest: Participants answered 12 multiple-choice questions (one comprehension question and two application questions for each rule). These two types of questions were separately provided to participants (Appendix B & C).
3. Learning (worked examples and practice questions) and feedback: Participants were randomly assigned to one of the three conditions and learned four probability rules with worked examples. Those examples were presented through PowerPoint slides. The whole learning process was self-paced. During learning, participants could take notes. However, they were allowed to keep their notes after the learning section. After participants finished reading worked examples of each rule, they practiced the knowledge by solving a problem similar to the worked examples. After the practice and before learning the next rule, participants were asked to provide feedback for the content just learned (Appendix E).
4. Posttest: Participants completed a posttest that is a parallel version to the pretest. The comprehension and application questions were provided separately to participants also.

5. Transfer test: This was the last part of the first session. The purpose of the transfer test was to probe how the superficial variability of examples affected students’ problem-solving transfer for probabilistic problems. This test was composed of 16 multiple-choice questions (four application questions for each rule) (Appendix D).

Session 2 (10-15 minutes)

One week after participants finished the first session, they were required to come back to complete a delayed posttest.

*Figure 6.3. Task Flow in the Dissertation Study*
6.4 Materials and Measures

**Personal information survey.** This survey contained questions to collect participants’ gender, major, educational level, expertise in mathematics and familiarity with probability theory and with the contexts of cover stories used in instructional examples (Appendix A).

**Pretest, posttest, and delayed posttest.** There were two versions (A and B version) of the test to examine students’ understanding of probability rules. These two versions were parallel to each other, which meant the context and structure of questions always stayed the same; however, the numbers in questions and the order of questions and answers changed (Appendix B & C). Each participant was randomly assigned a version for the pretest, and the other version was used as his/her posttest. His/her delayed posttest was the same as the posttest. For example, if a participant took Version A for his/her pretest, then he/she would have Version B for his/her posttest and delayed posttest. Each version was composed of 12 multiple-choice questions including two types of question — the comprehension questions and application questions. The comprehension questions were designed to test participants understanding of the definition (Appendix B). There were two correct answers for each comprehension question. The purpose of application questions was to measure participants’ ability to solve probabilistic questions (Appendix C). Questions for permutations and combinations had two types of correct answers. One was in the form of a formula (e.g., $\frac{1}{52} P_3$) and the other one was in the form of a fractions (e.g., $\frac{1}{52} \frac{1}{51} \frac{1}{50}$). For questions that had two correct answers, the full score was two. Participants received one point if they choose one
correctly and were deducted one point if they chose one incorrectly. The lowest score was zero. For example, if a participant chose two correct answers and one incorrect answer for one multiple-choice question, then his/her score was one.

**Worked examples and practice.** In the instructional phase, worked examples (Appendix F) were presented on PowerPoint slides with well-designed animations and images (Appendix G) corresponding to the cover story of each example. According to Paas & Van Merriënboer (1994), using worked examples can balance the cognitive load caused by problem variability. Thus, they provided more possibility to add variability and, in turn, to enhance schema acquisition. The learning process was entirely self-paced. By tapping the right- and left-arrow-button on the keyboard, participants could view next or previous solution steps. In order to improve engagement and reinforce learning, the author suggested that participants take notes and asked them to solve practice questions. The practice questions were very similar to the example questions. Participants were allowed to review examples while they worked on the practice questions.

Cover stories that used in the pretest, the two posttests, and the instruction for each of the four rules are shown in the Appendix H. To control participants’ exposure to varying cover stories, only four cover stories (poker cards, dice, marbles, and spinner) were used in these four sections. As the two posttests were parallel tests for the pretest and all conditions had the same set of parallel tests, all participants were considered familiar with their posttests. In other words, for all conditions, the questions in the posttests were not considered transfer questions. Moreover, to counterbalance participants’ familiarity of each cover story for each rule in each condition, the cover
stories used for each rule in the pretest and posttests overlapped with the cover stories used for that rule’s instructional examples.

**Transfer test.** The transfer test was composed of 16 multiple-choice questions including four application questions for each rule and no comprehension questions. The cover stories were never used in the pretest, the posttest, and instructional examples. The format and structure used to present some of the questions in the transfer test were different from the previous tests and instructional examples (Appendix D). Also, for questions that had two correct answers, the full score was two. Participants received one point if they chose one correctly and were deducted one point if they chose one incorrectly. The lowest score was zero.
7 Results

This section reports the major results of the dissertation study and is composed of four sub-sections. The first sub-section presents the descriptive statistical data. These descriptive statistical data include participants’ mean accuracy for the pretest, posttest, transfer test and delayed posttest. As the score ranges of the tests are different, the author used accuracies instead of raw scores to analyze participants’ performance. The second and third sub-section reports the test results for the first two hypotheses of the two posttests and the transfer test respectively. The last sub-section reports the test of the third hypothesis. Several statistical analyses are presented in this whole section to validate the hypotheses.

7.1 Performance Overview

Does participants’ performance improve after the intervention? Participants’ mean accuracies for the pretest, posttest, delayed posttest, and transfer test for all conditions are reported in Table 7.1. In order to discover whether the learning materials improved students learning, one-tailed paired-samples t-tests were conducted to check whether the accuracies in the posttest, transfer test, and delayed posttest were higher than the accuracy in the pretest. The results (Figure 7.1.1 & 7.1.2 & Table 7.1) indicated that among three conditions all participants’ performance in the pretest was significantly lower than in the posttest, $t(96) = -10.881, p < .001$, $d = 1.106$, delayed posttest, $t(96) = -9.848, p < .001$, $d = 1.000$, and transfer test, $t(96) = -3.604, p < .001$, $d = 0.366$. The same analysis was conducted within each condition. In the CS condition, participants’ performance in the pretest was also significantly lower than in the posttest, $t(29) = -6.167, p < .001$, $d = 1.124$, delayed posttest, $t(29) = -5.351, p < .001$, $d = 0.979$, and
transfer test, $t(29) = -1.963, p = 0.03, d = 0.360$. In the VSWR condition, participants’ performance in the pretest was also significantly lower than in the posttest, $t(33) = -9.344, p < .001, d = 1.604$, delayed posttest, $t(33) = -7.288, p < .001, d = 1.248$, and transfer test, $t(33) = -3.216, p = .003, d = 0.550$. In the VSBR condition, participants’ performance in the pretest was also significantly lower than in the posttest, $t(32) = -4.294, p < .001, d = 0.747$, and delayed posttest, $t(32) = -4.607, p < .001, d = 0.801$. However, there was no difference between pretest and transfer test, $t(32) = -1.064, p = .074, d = 0.188$. Based on the above results, the learning materials were proven effective in improving participants’ learning of the four probability rules.

Table 7.1

<table>
<thead>
<tr>
<th></th>
<th>All Conditions</th>
<th>Condition 1 CS</th>
<th>Condition 2 VSWR</th>
<th>Condition 3 VSBR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>N</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Pretest Accuracy</td>
<td>.531</td>
<td>97</td>
<td>.162</td>
<td>.533</td>
</tr>
<tr>
<td>Posttest Accuracy</td>
<td>.719</td>
<td>97</td>
<td>.159</td>
<td>.733</td>
</tr>
<tr>
<td>Transfer Test Accuracy</td>
<td>.595</td>
<td>97</td>
<td>.209</td>
<td>.600</td>
</tr>
<tr>
<td>Delayed Posttest Accuracy</td>
<td>.704</td>
<td>97</td>
<td>.161</td>
<td>.723</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 7.1.1. Mean of accuracies of the pretest, posttest, transfer test and delayed posttest in each condition (clustered by condition). AccuracyPre represents the accuracy of the pretest. AccuracyPost represents the accuracy of the posttest. AccuracyT represents the accuracy of the transfer test. AccuracyD represents the accuracy of the delayed posttest.

Figure 7.1.2. Mean of accuracies of the pretest, posttest, transfer test and delayed posttest in each condition (clustered by the type of tests).
7.2 Effect of Superficial Variability of Examples on Learning Performance in the Posttest and Delayed Posttest

*Does participants’ improvement in two posttests differ by condition?* In order to examine the first and second hypothesis regarding the posttest and delayed posttest, the author tested the differences in the posttest and delayed posttest scores across conditions, while controlling participants’ pretest scores. A two-way mixed ANCOVA with Helmert contrast was conducted to determine (a) whether participants’ performance in the posttest and delayed posttest under the CS condition and the VSWR condition was statistically significantly better than in the VSBR condition; and (b) whether participants’ performance in the posttest and delayed posttest in the VSWR condition was statistically significantly better than in the CS condition, controlling their performance in the pretest.

In this analysis, the dependent variable was participants’ learning performance in two types of posttest. The covariate was participants’ performance in the pretest. The within-subject factor was the time of test, which had two levels: posttest and delayed posttest. The between-subject factor was the condition, which had three categories: Condition 1 (CS), 2 (VSWR), and 3 (VSBR). In order to employ the appropriate Helmert contrast, the author recoded the data and generated a new variable *cond_new*. The value of *cond_new* equaled one if the participant was in Condition 3 (VSBR); *cond_new* equaled two if the participant was in Condition 2 (VSWR); *cond_new* equaled three if the participant was in Condition 1 (CS) (Table 7.2.1).
According to the results, there was neither a statistically significant interaction between the time of test and the pretest performance, $F(1, 93) = .101, p = .751$, partial $\eta^2 = .001$, nor between the time of test and the condition, $F(2, 93) = .437, p = .647$, partial $\eta^2 = .009$. In the Helmert contrast test (Table 7.2.2), there was a statistically significant difference between the VSBR condition versus later, $M_{\text{diff}} = -.055, p = .049$, which means participants’ performance in the CS condition and the VSWR condition were statistically significantly better than the VSBR condition. However, there was no statistically significant difference between the CS condition and the VSWR condition. Therefore, Hypothesis 1 was confirmed by the performance in the two posttests; however, Hypothesis 2 was not.

Table 7.2.2

<table>
<thead>
<tr>
<th>Helmert Contrast Results (K Matrix) of the Two Posttests</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3 (VSBR) vs. C1 (CS) &amp; C2 (VSWR)</td>
</tr>
<tr>
<td><strong>Contrast Estimate</strong></td>
</tr>
<tr>
<td><strong>Sig.</strong></td>
</tr>
<tr>
<td><strong>95% Confidence Interval</strong></td>
</tr>
<tr>
<td><strong>Lower Bound</strong></td>
</tr>
<tr>
<td><strong>Upper Bound</strong></td>
</tr>
<tr>
<td>C1 (CS) vs. C2 (VSWR)</td>
</tr>
<tr>
<td><strong>Contrast Estimate</strong></td>
</tr>
<tr>
<td><strong>Sig.</strong></td>
</tr>
<tr>
<td><strong>95% Confidence Interval</strong></td>
</tr>
<tr>
<td><strong>Lower Bound</strong></td>
</tr>
<tr>
<td><strong>Upper Bound</strong></td>
</tr>
</tbody>
</table>
Figure 7.2.1. Profile Plots of Students’ Performance in Two Posttests in Three Conditions

7.3 Effect of Superficial Variability of Examples on Learning Transfer

Does learning transfer differ by condition? Since the transfer test had a different number of items from the other three tests, the transfer test was not considered a repeated measure for the other tests. Therefore, a one-way analysis of covariance (ANCOVA) with Helmert contrast was applied to determine (a) whether participants’ performance in the transfer test under the CS condition and the VSWR condition was statistically significantly better than in the VSBR condition; and (b) whether participants’ performance in the transfer test in the VSWR condition was statistically significantly better than in the CS condition, while participants’ performance in the pretest was controlled. The dependent variable was participants’ performance in the transfer test. The
independent variable was the condition (\textit{cond\_new}). The covariate was the performance in the pretest. Based on the Helmert contrast test (Table 7.3), there was neither a statistically significant difference between the VSBR condition versus the other two conditions (\(M_{\text{diff}} = -0.041, p > .05\)) nor between the CS condition and VSWR condition (\(M_{\text{diff}} = 0.020, p > .05\)) (Figure 7.3). Therefore, the hypotheses of students’ performance in the transfer test were not confirmed.

\textit{Table 7.3}

\textit{Helmert Contrast Results (K Matrix) for the Transfer Test}

<table>
<thead>
<tr>
<th>Contrast Estimate</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3 (VSBR) vs. C1 (CS) &amp; C2 (VSWR)</td>
<td>-0.041</td>
<td>0.273</td>
<td>-0.114</td>
</tr>
<tr>
<td>C1 (CS) vs. C2 (VSWR)</td>
<td>0.002</td>
<td>0.637</td>
<td>-0.065</td>
</tr>
</tbody>
</table>
7.4  Strength of the Pattern of the Effect of the Superficial Variability of Examples

*Does the strength of the influence of superficial variability of examples differ by problem types' difficulty?* It was expected that the pattern of the effect of the superficial variability of examples varies in strength among rules that are at different difficulty levels (Hypothesis 3). In order to test this hypothesis, it was necessary to prove that the difficulty levels of the four rules was different. Based on the findings in the pilot studies and conceptual complexity of each rule, those four rules were categorized into two level groups — easy rules and difficult rules. The addition and multiplication rules were considered easy rules. The permutation and combination rules were considered difficult rules. To validate this categorization, the comparison between students’ perceived difficulty of the four rules and the actual difficulty of the four rules was conducted.
The perceived difficulty of the rules was measured by participants’ rating of the
difficulty of the rules. Immediately after finishing the learning process of each rule,
participants were required to provide feedback of the learning content, which included
the ratings of the difficulty level of each rule. The ratings ranged from 1, *Not (Difficult)*
*At All*, to 5, *Very Much (Difficult)* (Appendix E). The mean rating of the easy rules
(*M*=1.990, *SD*= 0.794) was significantly lower than the difficulty rules (*M*=2.899,
*SD*=0.876), *F* (1, 96) = 254.105, *p* < .001, partial η² = .726 (Figure 7.4.1). The actual
difficulty was measured by participants’ actual performance on the easy rules and
difficult rules. The mean of accuracies of participants’ performance on the easy rules was
significantly higher than on difficult rules in each of four tests and overall tests (Table
7.4.2 & Figure 7.4.2). Therefore, participants’ performance on the two difficulty levels of
rules was consistent with their ratings of the difficulty level of rules. In other words, the
actual difficulty of rules coincided with the perceived difficulty of rules, which also
supported the categorization of these four rules.

Table 7.4.1

*Difficulty Ratings of the Easy and Difficult Rules*

<table>
<thead>
<tr>
<th></th>
<th>Easy</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.990</td>
<td>2.899</td>
</tr>
<tr>
<td>SD</td>
<td>0.794</td>
<td>0.876</td>
</tr>
</tbody>
</table>
Figure 7.4.1. Perceived difficulty of the easy and difficult rules

Table 7.4.2

Performance on Two Difficulty Levels of Rules in the Tests

<table>
<thead>
<tr>
<th></th>
<th>Easy Mean</th>
<th>Diff Mean</th>
<th>Easy Mean</th>
<th>Diff Mean</th>
<th>Easy Mean</th>
<th>Diff Mean</th>
<th>Easy Mean</th>
<th>Diff Mean</th>
<th>Easy Mean</th>
<th>Diff Mean</th>
<th>Easy Mean</th>
<th>Diff Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>0.664</td>
<td>0.442</td>
<td>0.800</td>
<td>0.664</td>
<td>0.821</td>
<td>0.628</td>
<td>0.777</td>
<td>0.504</td>
<td>0.765</td>
<td>0.560</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest</td>
<td>0.196</td>
<td>0.177</td>
<td>0.197</td>
<td>0.184</td>
<td>0.200</td>
<td>0.210</td>
<td>0.260</td>
<td>0.145</td>
<td>0.178</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delayed Posttest</td>
<td>0.204</td>
<td>0.196</td>
<td>0.177</td>
<td>0.197</td>
<td>0.200</td>
<td>0.210</td>
<td>0.260</td>
<td>0.145</td>
<td>0.178</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To test whether the strength of the pattern of the effect of the superficial variability of examples varies by a rule’s difficulty, two separate analyses were applied for the performance on the two posttests and the transfer test. For the two posttests, the same analysis method as in Section 7.2 was applied with difficulty level as an additional within-subject factor (Table 7.4.3). According to the results, there was no three-way or two-way interactions between variables (Table 7.4.4). The difficulty level of rules had a significant effect on participants’ performance, $F(2,93) = 18.701, p < .001$. In the Helmert contrast test, the same results were found as in Section 7.2: participants’ performance in the CS condition and the VSWR condition were statistically significantly better than the VSBR condition, $M_{diff} = -.060, p = .028$ (Table 7.4.5); however, there was no statistically significant difference between the CS condition and VSWR condition.
Table 7.4.3

*Within-Subject Factors in the Three Way Mixed ANCOVA Test (BWW)*

<table>
<thead>
<tr>
<th>time</th>
<th>difficulty</th>
<th>Dependent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Posttest)</td>
<td>1 (Easy)</td>
<td>AccuracyPostEasy</td>
</tr>
<tr>
<td>2 (Difficult)</td>
<td></td>
<td>AccuracyPostDiff</td>
</tr>
<tr>
<td>2 (Delayed Posttest)</td>
<td>1 (Easy)</td>
<td>AccuracyDEasy</td>
</tr>
<tr>
<td></td>
<td>2 (Difficult)</td>
<td>AccuracyDDiff</td>
</tr>
</tbody>
</table>

Table 7.4.4

*Tests of Within-Subjects Effects*

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Observed Power*</th>
</tr>
</thead>
<tbody>
<tr>
<td>time</td>
<td>1</td>
<td>.000</td>
<td>.990</td>
<td>.000</td>
<td>.050</td>
</tr>
<tr>
<td>time * AccuracyPre</td>
<td>1</td>
<td>.038</td>
<td>.846</td>
<td>.000</td>
<td>.054</td>
</tr>
<tr>
<td>time * cond_new</td>
<td>2</td>
<td>.402</td>
<td>.670</td>
<td>.009</td>
<td>.113</td>
</tr>
<tr>
<td>Error(time)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>difficulty</td>
<td>1</td>
<td>18.701</td>
<td>.000</td>
<td>.167</td>
<td>.990</td>
</tr>
<tr>
<td>difficulty * AccuracyPre</td>
<td>1</td>
<td>3.716</td>
<td>.057</td>
<td>.038</td>
<td>.479</td>
</tr>
<tr>
<td>difficulty * cond_new</td>
<td>2</td>
<td>.678</td>
<td>.510</td>
<td>.014</td>
<td>.161</td>
</tr>
<tr>
<td>Error(difficulty)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>time * difficulty</td>
<td>1</td>
<td>.004</td>
<td>.953</td>
<td>.000</td>
<td>.050</td>
</tr>
<tr>
<td>time * difficulty * AccuracyPre</td>
<td>1</td>
<td>.454</td>
<td>.502</td>
<td>.005</td>
<td>.102</td>
</tr>
<tr>
<td>time * difficulty * cond_new</td>
<td>2</td>
<td>.515</td>
<td>.599</td>
<td>.011</td>
<td>.133</td>
</tr>
<tr>
<td>Error(time*difficulty)</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7.4.5

*Helmert Contrast Results (K Matrix) for the Posttests with Difficulty as a Within-Subject Factor*

<table>
<thead>
<tr>
<th>Contrast Estimate</th>
<th>Sig.</th>
<th>95% Confidence Interval for Difference</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3 (VSBR) vs. C1 (CS) &amp; C2 (VSWR)</td>
<td>-0.060</td>
<td>0.028</td>
<td>-0.113</td>
<td>0.006</td>
</tr>
<tr>
<td>C1 (CS) vs. C2 (VSWR)</td>
<td>1.969E-5</td>
<td>0.999</td>
<td>-0.062</td>
<td>0.062</td>
</tr>
</tbody>
</table>

Figure 7.4.3: Profile Plots of Students’ Performance in Two Posttests at Two Difficulty Levels in Three Conditions

As the questions for the easy rules and difficult rules were separated, examination within the transfer test was considered a repeated measure. Therefore, the same analysis method (the two-way mixed ANCOVA with Helmert contrast) as in Section 7.2 was
conducted. In this analysis, the dependent variable was participants’ learning performance in the transfer test. The covariate was participants’ performance in the pretest. The within-subject factor was the difficulty level, which had two levels: easy and difficult. The between-subject factor was the condition \( (\text{cond}\_\text{new}) \). Based on the results, there was no significant interaction between the condition and the difficulty level, \( F(2, 93) = .902, p = .409, \) partial \( \eta^2 = .019 \). In the Helmert contrast, there was no statistically significant difference between conditions (Table 7.4.6). This finding was consistent with the one in Section 7.3, where the effect of the superficial variability of examples on the transfer test was tested regardless of difficulty level.

Table 7.4.6

*Helmert Contrast Results (K Matrix) for the Transfer Test with Difficulty as a Within-Subject Factor*

<table>
<thead>
<tr>
<th>Contrast Estimate</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3 (VSBR) vs. C1 (CS) &amp; C2 (VSWR)</td>
<td>-0.041</td>
</tr>
<tr>
<td>Sig.</td>
<td>0.241</td>
</tr>
<tr>
<td>95% Confidence Interval</td>
<td>Lower Bound</td>
</tr>
<tr>
<td>for Difference</td>
<td>Upper Bound</td>
</tr>
<tr>
<td>C1 (CS) vs. C2 (VSWR)</td>
<td>0.006</td>
</tr>
<tr>
<td>Sig.</td>
<td>0.889</td>
</tr>
<tr>
<td>95% Confidence Interval</td>
<td>Lower Bound</td>
</tr>
<tr>
<td>for Difference</td>
<td>Upper Bound</td>
</tr>
</tbody>
</table>

In the above analyses, the pattern of the effect of the superficial variability of examples on both the transfer test and the two posttests did not show a statistically significant difference at two difficulty level. Moreover, the pattern at two difficulty levels was visually alike in the two posttests and the transfer test (Figure 7.4.3 &7.4.4). Therefore, Hypothesis 3 was not supported.
Figure 7.4.4. Profile Plots of Students’ Performance in the Transfer Test at Two Difficulty Levels in Three Conditions

Covariates appearing in the model are evaluated at the following values: AccuracyPre = .5309

* difficulty 1 represents the easy rules, difficulty 2 represents the difficulty rules.
8 Discussion

In this dissertation, the author has investigated the effect of superficial variability of examples on learning applied probability rules. The possibility of a varying level of strength of that effect at different difficulty levels of rules has been examined as well. Consolidating findings from the analyses, three major conclusions have been drawn.

First of all, students learning through examples with consistent cover stories and learning with cover stories varied within rules performed statistically significantly better than students learning through examples with cover stories varied between rules did.

This finding coincides with the finding in the second pilot study, the suggestion by Ross (1989b) and the findings in Quilici and Mayer’s study (1996). Compared to the other two conditions, students in the varied-surface-between-rule condition had the least exposure to structure-feature-emphasizing examples. Moreover, this group of students had less opportunity to practice how to distinguish structural features from superficial features. Therefore, their learning outcome was worse than students in the other two conditions. On the contrary, in both of the varied-surface-within-rule condition and the consistent-surface condition, the structural differences between different problem types were better emphasized. Moreover, the students in the varied-surface-within-rule condition were expected to have even better learning outcome than the ones in the consistent-surface condition, as the former condition provided students more opportunities to learn to abstract structural differences. However, the analysis results showed that students’ learning outcome in these two conditions was the same. Looking back at the design of learning examples in these two conditions (Table 2), the same characteristic between them was the between-rules consistency of superficial features: in
the consistent-surface-feature condition, the same cover story was used in multiple examples for each rule; in the varied-surface-feature-within-rule condition, the same battery of cover stories was used across the four rules, even though this battery was composed of varied cover stories. Therefore, the author concludes that instructional examples’ superficial consistency existing between different problem types (between-problem-types superficial consistency) promotes better learning performance than superficial variation between different problem types (between-problem-types superficial variation). Moreover, the between-problem-types superficial consistency includes two conditions. One condition is that the superficial consistency of examples exists within each type of problem as well as between different types of problems at the same time. Another condition is that examples’ surface features vary within each type of problem; however, the pattern of the variation of surface features is consistent between types of problems.

In addition to the learning performance in the two posttests, the same pattern of the difference caused by superficial variability was expected in the learning transfer. However, that expected difference has not been found statistically significant in this dissertation study, even though the trend (the learning transfer in the varied-surface-between-rule condition was worse than in the other two conditions) has been visually observed (Figure 7.3). Based on previous research on analogical transfer, sufficient practice and time of learning are necessary to enable the adequate encoding of learning, which results in successful transfer (Bereiter, 1995; Hammond, Seifert, & Gray, 1991; Haskell & Haskell, 2001). In this dissertation study, participants probably did not conduct enough practice to enable the condition effect to exert on transfer: participants were only
asked to complete one practice for the easy rules and two practice questions for the difficult rules. Therefore, it is reasonable to infer that if participants had had more time and opportunities to practice, the degree of difference between conditions in transfer outcome could be enlarged. In other words, students would have obtained better learning transfer in the condition where the structural features are the least confounded with the superficial features—the varied-surface-within-rule condition.

Last but not least, the strength of the pattern of the effect of the superficial variability of examples does not vary between problem types at different difficulty levels. The effect of problem types’ difficulty level was investigated on the two posttests and the transfer test. The pattern was the same as when the difficulty variable was not included: in the two posttests, statistically significantly worse performance was merely shown in the varied-surface-between-rule condition, and there was no statistically significant difference between the consistent-surface condition and the varied-surface-within-rule condition; in the transfer test, there was no statistically significant difference across conditions. Therefore, the author concludes that the pattern of the influence of superficial variability of examples is robust among types of the problem at varying difficulty levels.

Based on the findings from this dissertation, important implications for instructional activities have been discovered. This study provides evidence that example-based instruction in acquiring the structural schemas of probabilistic problems can be instrumental. To have more effective learning effects, examples of various types of the problem should be presented with either one same cover story or the same battery of cover stories.
Although there were no statistically significant interaction effects between difficulty and condition, intersections between these two factors were visually observed in Figure 7.4.3 and Figure 7.4.4: for less difficult problems, students’ performance in the consistent-surface condition was better than those in the varied-surface-within-rule condition and vice versa. Therefore, the author suggests using a consistent cover story to teach easy problems, but using a same battery of various cover stories to teach difficult problems. Moreover, the author suggests applying these instructional strategies to all mathematically-based curriculum instruction.
9 Limitations and Future Research Direction

There are limitations to the methodology of this dissertation study. First, as mentioned above, the practice session was not enough to enable participants to encode learning, which is probably the reason for the insignificant difference in participants’ transfer across the three conditions. Follow-up studies are needed to examine this alternative before drawing the final conclusion on the effect of superficial variability of examples on transfer.

Second, even though both participants’ perception and actual performance support the idea that the difficulty level varies among the probability rules, the origin of the difficulty was not controlled. By that, the author means that it was uncertain whether the difficulty stemmed from the problem type or the cover story. In future studies, researchers need to control for the difficulty of the context of cover stories, if they are interested in how difficulty level for problem types interacts with the effect of superficial variability of examples.

Third, the representativeness of the learner population was not diverse. All participants in this dissertation study were from a selective school. To enhance the generality of the findings, more diverse population samples from various schools should be drawn in future studies.

Finally, the applicability of the findings to other subject domains is worthwhile to investigate. Both this dissertation study and most of previous research focused on mathematically-based subjects. As schema construction is a generic process (Brewer & Nakamura, 1984), the author suggests similar studies on learning other STEM subjects.
and arts subjects to discover how well the instructional strategy fits on a larger scale in future studies.
References


http://doi.org/10.4324/9781315806938


Ortony (Eds.), Similarity and analogical reasoning (pp. 197-229). New York, NY:


solution failures to a source problem produces later successes in analogical transfer.
Journal of Experimental Psychology: Learning, Memory, and Cognition, 18(3),

learning. Learning and Instruction, 16(2 SPEC. ISS.), 122–138.
http://doi.org/10.1016/j.learninstruc.2006.02.001

foster learning outcomes? Learning and Instruction, 17(6), 612–634.
http://doi.org/10.1016/j.learninstruc.2007.09.008

examples: The importance of variation and prior knowledge. European Journal of


http://www.tcrecord.org/library/Abstract.asp?ContentId=1815

http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&amp;path=ASIN/0691023565


http://doi.org/10.1207/s1532690xci0403_1
Appendix A: Personal Information Survey

You are a: Female or Male?

What is your major? ______________________________

Which school are you from? _____________________________

Have you taken any probability class within past 3 years?
Yes No

How would you describe your familiarity with the statistical probability?

1 2 3 4 5
Not at all Not Really Neutral Somewhat Very Much

How would you describe your mathematical skills?

1 2 3 4 5
Not at all Not Really Neutral Somewhat Very Much

How would you describe your familiarity with poker games?

1 2 3 4 5
Not at all Not Really Neutral Somewhat Very Much

How would you describe your familiarity with the activity of spinning a spinner?

1 2 3 4 5
Not at all Not Really Neutral Somewhat Very Much

How would you describe your familiarity with the activity of rolling a dice?

1 2 3 4 5
Not at all Not Really Neutral Somewhat Very Much

How would you describe your familiarity with colorful marbles?

1 2 3 4 5
Not at all Not Really Neutral Somewhat Very Much
Appendix B: Sample Comprehension Questions in the Pretest

1. Please select the correct statement(s) about the Addition Rule:

A. When two events, A and B, are mutually exclusive, the probability that A or B will occur is: \( P(A \text{ or } B) = P(A) + P(B) \) [correct]

B. When two events, A and B, are independent, the probability of both occurring is: \( P(A \text{ and } B) = P(A) \times P(B) \)

C. When two events, A and B, are not mutually exclusive, the probability that A or B will occur is: \( P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \) [correct]

D. When two events, A and B, are dependent, the probability of both occurring is: \( P(A \text{ and } B) = P(A) \times P(B|A) \)

2. Please select the correct statement(s) and formula(s) about the Permutation Rule:

A. A permutation is a rearrangement of the elements of an ordered list into a one-to-one correspondence with itself. [correct]

B. A permutation is a way of selecting items from a collection, such that the order of selection does not matter.

\[
^n P_r = \frac{n!}{(n-r)!}
\]

[correct]

\[
^n C_r = \frac{n!}{r!(n-r)!} = \frac{n P_r}{r!}
\]

D. \( P(C) \)
Appendix C: Sample Application Questions in the Pretest

Q1: Peter is randomly picking a card from a standard deck of 52 playing cards. What is the probability of his choosing a queen or a heart?

A. \[ \frac{15}{52} \]

B. \[ \frac{16}{52} \] [correct]

C. \[ \frac{17}{52} \]

D. \[ \frac{18}{52} \]

Q3: Terry is randomly picking two cards, one by one, from a standard deck of 52 playing cards. What is the probability that the first card chosen is a club and the second card chosen is a diamond?

A. \[ \frac{13}{52} \frac{12}{52} \]

B. \[ \frac{13}{52} \frac{13}{51} \] [correct]

C. \[ \frac{13}{52} \frac{13}{52} \]

D. \[ \frac{13}{52} \frac{12}{51} \]

Q5: Mike is drawing three cards from the pile of 52, one by one, what’s the probability of those 3 cards are exactly in the order of King of Club, 2 of Spade, 6 of Spade?

A. \[ \frac{1}{52 \binom{3}{3}} \]

B. \[ \frac{1}{13} \frac{1}{12} \frac{1}{11} \]

C. \[ \frac{1}{52 \binom{3}{3}} \] [correct]

D. \[ \frac{1}{52} \frac{2}{50} \]
Q8: A glass jar contains 3 blue, 2 green, and 2 yellow balls. Choose 3 balls from the jar one by one. What is the probability that those 3 chosen balls include 1 blue, 1 green, and 1 yellow?

A. \( \frac{1}{7} \cdot \frac{1}{6} \cdot \frac{1}{5} \) (3 2 2)

B. \( \frac{1}{\binom{7}{3}} \) 3 2 2 [correct]

C. \( \frac{3}{7} \cdot \frac{2}{6} \cdot \frac{2}{5} \) (3 2 1) [correct]

D. \( \frac{1}{\binom{7}{3}} \) 3 2 2

E. \( \frac{1}{\binom{7}{3}} \)
Appendix D: Sample Questions in the Transfer Test

1. On New Year's Eve, the probability of a person having a car accident is 0.09. The probability of a person driving while intoxicated is 0.32 and probability of a person having a car accident while intoxicated is 0.15. What is the probability of a person driving while intoxicated or having a car accident?
   A. 0.32 + 0.09
   B. 0.32 - 0.09
   C. 0.32 + 0.09 – 0.15 [correct]
   D. 0.32 + 0.09 + 0.15
   E. (0.32 + 0.09) x 0.15

2. What is the probability that the word of “SHIFT” is made from distinguishable arrangements of F, H, I, S, and T? These five letters cannot be used repeatedly.
   \[
   \frac{1}{5^5}
   \]
   A. \(\frac{1}{5^5}\)
   B. \(\frac{1}{5^5}\) \(\text{P}_5\) [correct]
   C. \(\frac{1}{5} \frac{1}{5} \frac{1}{5} \frac{1}{5} \frac{1}{5}\)
   D. \(\frac{1}{5} \frac{1}{5} \frac{1}{3} \frac{1}{2}\) [correct]
   E. \(\left(\frac{1}{5} \frac{1}{4} \frac{1}{3} \frac{1}{2}\right) (5 \ 4 \ 3 \ 2)\)
   F. \(\left(\frac{1}{5} \frac{1}{4} \frac{1}{3} \frac{1}{2}\right) (5+4+3+2)\)

3. A nationwide survey showed that 65% of all children in the United States dislike eating vegetables. If 4 children are chosen at random, what is the probability that all 4 dislike eating vegetables?
   A. 0.35 x 0.35 x 0.35 x 0.35
   B. 0.65 x 0.35 x 0.65 x 0.35
C. 0.65 x 0.65 x 0.65 x 0.65 [correct]
D. 0.65 + 0.65 + 0.65 + 0.65
E. 0.36 x 0.65 + 0.35 x 0.65
### Appendix E: Feedback Survey

#### For Addition Rule

1. **Do you think Addition Rule is difficult for you?**
   
<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>Not Really</td>
<td>Neutral</td>
<td>Somewhat</td>
<td>Very Much</td>
</tr>
</tbody>
</table>

2. **Do you think the instruction (PowerPoint slides) of Addition Rule is clear?**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>Not Really</td>
<td>Neutral</td>
<td>Somewhat</td>
<td>Very Much</td>
</tr>
</tbody>
</table>

3. **Do you think the instruction is helpful for your understanding of Addition Rule?**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>Not Really</td>
<td>Neutral</td>
<td>Somewhat</td>
<td>Very Much</td>
</tr>
</tbody>
</table>

4. **How would you rate your understanding of Addition Rule after going through the slides compared to before you going through the slides?**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worse</td>
<td>As well as before</td>
<td>Better</td>
</tr>
</tbody>
</table>

#### For Multiplication Rule

5. **Do you think Multiplication Rule is difficult for you?**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>Not Really</td>
<td>Neutral</td>
<td>Somewhat</td>
<td>Very Much</td>
</tr>
</tbody>
</table>

6. **Do you think the instruction (PowerPoint slides) of Multiplication Rule is clear?**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>Not Really</td>
<td>Neutral</td>
<td>Somewhat</td>
<td>Very Much</td>
</tr>
</tbody>
</table>

7. **Do you think the instruction is helpful for your understanding of Multiplication Rule?**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td>Not Really</td>
<td>Neutral</td>
<td>Somewhat</td>
<td>Very Much</td>
</tr>
</tbody>
</table>

8. **How would you rate your understanding of Multiplication Rule after going through the slides compared to before you going through the slides?**

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worse</td>
<td>As well as before</td>
<td>Better</td>
</tr>
</tbody>
</table>
Appendix F: Sample Worked Problem

Q2:
Draw three cards from the pile of 13, one by one, what’s the probability of those three cards are A,2,3, in that exact order?

\[
\begin{align*}
\text{You want the first card to be an A. } P(\text{the first card chosen is A}) &= \frac{1}{13} \\
\text{You want the second card to be an 2. } P(\text{the second card chosen is 2}) &= \frac{1}{12} \\
\text{You want the third card to be an 3. } P(\text{the third card chosen is 3}) &= \frac{1}{11} \\
\end{align*}
\]

\[
P(\text{choosing A, then 2, then 3}) = P(\text{first card is A}) \times P(\text{second card is 2}) \times P(\text{third card is 3}) \\
= \frac{1}{13} \times \frac{1}{12} \times \frac{1}{11} \\
= \frac{1}{13 \times 12 \times 11} \times \frac{10!}{10!} = \frac{1 \times (10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)}{(13 \times 12 \times 11) \times (10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)} \\
= \frac{10!}{13!} \times \frac{(13-3)!}{13!} = \frac{1}{(13-3)!} \times P_3
\]
Appendix G: Screen Shots for the Instructional Slides

The following images are screen shots for a worked example of the permutation rule. Each image is for one step of the solution. Each step is triggered by tapping the right- and left-arrow-button on the keyboard.

**Step 1**

Q2: Draw three cards from the pile of 13, one by one, what’s the probability of those three cards are A, 2, 3, in that exact order?

**Step 2**

You want the first card to be an A. \( P(\text{the first card chosen is } A) = \frac{1}{13} \)

**Step 3**

You want the second card to be an 2. \( P(\text{the second card chosen is } 2) = \frac{1}{12} \)
Q2:
Draw three cards from the pile of 13, one by one, what’s the probability of those three cards are A, 2, 3, in that exact order?

You want the first card to be an A. $P($the first card chosen is A$) = \frac{1}{13}$

You want the second card to be an 2. $P($the second card chosen is 2$) = \frac{1}{12}$

You want the third card to be an 3. $P($the third card chosen is 3$) = \frac{1}{11}$

$P($choosing A, then 2, then 3$) = P($first card is A$) \times P($second card is 2$) \times P($third card is 3$)$

Step 4

Step 5
Q2:
Draw three cards from the pile of 13, one by one, what’s the probability of those three cards are A,2,3, in that exact order?

You want the first card to be an A. \( P(\text{the first card chosen is } A) = \)

You want the second card to be an 2. \( P(\text{the second card chosen is } 2) = \)

You want the third card to be an 3. \( P(\text{the third card chosen is } 3) = \)

\[
P(\text{choosing } A, \text{ then } 2, \text{ then } 3) = P(\text{first card is } A) \times P(\text{second card is } 2) \times P(\text{third card is } 3)
\]

\[
= \frac{1}{13} \times \frac{1}{12} \times \frac{1}{11} = \frac{1}{13 \times 12 \times 11}
\]

Step 6

Q2:
Draw three cards from the pile of 13, one by one, what’s the probability of those three cards are A,2,3, in that exact order?

You want the first card to be an A. \( P(\text{the first card chosen is } A) = \)

You want the second card to be an 2. \( P(\text{the second card chosen is } 2) = \)

You want the third card to be an 3. \( P(\text{the third card chosen is } 3) = \)

\[
P(\text{choosing } A, \text{ then } 2, \text{ then } 3) = P(\text{first card is } A) \times P(\text{second card is } 2) \times P(\text{third card is } 3)
\]

\[
= \frac{1}{13} \times \frac{1}{12} \times \frac{1}{11} = \frac{1}{13 \times 12 \times 11}
\]

\[
= \frac{1}{13 \times 12 \times 11} \times \frac{10!}{10!}
\]

Step 7
Q2:
Draw three cards from the pile of 13, one by one, what’s the probability of those three cards are A,2,3, in that exact order?

You want the first card to be an A. \( P(\text{the first card chosen is A}) = \)

You want the second card to be an 2. \( P(\text{the second card chosen is 2}) = \)

You want the third card to be an 3. \( P(\text{the third card chosen is 3}) = \)

\[
P(\text{choosing A, then 2, then 3}) = P(\text{first card is A}) \times P(\text{second card is 2}) \times P(\text{third card is 3})
\]
\[
= \frac{1}{13} \times \frac{1}{12} \times \frac{1}{11}
\]
\[
= \frac{1}{13 \times 12 \times 11} \times \frac{10!}{(13 \times 12 \times 11)} = \frac{10!}{(13 \times 12 \times 11) 	imes (10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)}
\]

Step 8

Q2:
Draw three cards from the pile of 13, one by one, what’s the probability of those three cards are A,2,3, in that exact order?

You want the first card to be an A. \( P(\text{the first card chosen is A}) = \)

You want the second card to be an 2. \( P(\text{the second card chosen is 2}) = \)

You want the third card to be an 3. \( P(\text{the third card chosen is 3}) = \)

\[
P(\text{choosing A, then 2, then 3}) = P(\text{first card is A}) \times P(\text{second card is 2}) \times P(\text{third card is 3})
\]
\[
= \frac{1}{13} \times \frac{1}{12} \times \frac{1}{11}
\]
\[
= \frac{1}{13 \times 12 \times 11} \times \frac{10!}{(13 \times 12 \times 11)} = \frac{10!}{(13 \times 12 \times 11) 	imes (10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)}
\]

Step 9
Q2:
Draw three cards from the pile of 13, one by one, what’s the probability of those three cards are A,2,3, in that exact order?

You want the first card to be an A. \( P(\text{the first card chosen is A}) = \)

You want the second card to be an 2. \( P(\text{the second card chosen is 2}) = \)

You want the third card to be an 3. \( P(\text{the third card chosen is 3}) = \)

\[
P(\text{choosing A, then 2, then 3}) = P(\text{first card is A}) \times P(\text{second card is 2}) \times P(\text{third card is 3})
\]

\[
= \frac{1}{13} \times \frac{1}{12} \times \frac{1}{11} = \frac{1}{13 \times 12 \times 11}
\]

\[
= \frac{10!}{13\times12\times11} = \frac{10!}{(13\times12\times11) \times 10!} = \frac{10!}{(13\times12\times11) \times (10\times9\times8\times7\times6\times5\times4\times3\times2\times1)}
\]

\[
= \frac{10!}{13!} = \frac{(13-3)!}{13!}
\]
Q2:
Draw three cards from the pile of 13, one by one, what’s the probability of those three cards are A, 2, 3, in that exact order?

You want the first card to be an A. \( P(\text{the first card chosen is A}) = \)

You want the second card to be an 2. \( P(\text{the second card chosen is 2}) = \)

You want the third card to be an 3. \( P(\text{the third card chosen is 3}) = \)

\[
P(\text{choosing A, then 2, then 3}) = P(\text{first card is A}) \times P(\text{second card is 2}) \times P(\text{third card is 3})
\]

\[
= \frac{1}{13} \times \frac{1}{12} \times \frac{1}{11} = \frac{1}{13 \times 12 \times 11}
\]

\[
= \frac{1}{13 \times 12 \times 11} \times \frac{10!}{10!} = \frac{1 \times (10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)}{(13 \times 12 \times 11) \times (10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)}
\]

\[
= \frac{10!}{13!} \times \frac{(13 - 3)!}{13!} = \frac{1}{13!} \times \frac{1}{13!} = \frac{1}{13!} \times \frac{1}{13!}
\]

Step 12

Q2:
Draw three cards from the pile of 13, one by one, what’s the probability of those three cards are A, 2, 3, in that exact order?

You want the first card to be an A. \( P(\text{the first card chosen is A}) = \)

You want the second card to be an 2. \( P(\text{the second card chosen is 2}) = \)

You want the third card to be an 3. \( P(\text{the third card chosen is 3}) = \)

\[
P(\text{choosing A, then 2, then 3}) = P(\text{first card is A}) \times P(\text{second card is 2}) \times P(\text{third card is 3})
\]

\[
= \frac{1}{13} \times \frac{1}{12} \times \frac{1}{11} = \frac{1}{13 \times 12 \times 11}
\]

\[
= \frac{1}{13 \times 12 \times 11} \times \frac{10!}{10!} = \frac{1 \times (10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)}{(13 \times 12 \times 11) \times (10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1)}
\]

\[
= \frac{10!}{13!} \times \frac{(13 - 3)!}{13!} = \frac{1}{13!} \times \frac{1}{13!} = \frac{1}{13!} \times \frac{1}{13!}
\]

Step 13
Appendix H: Cover Stories Used in the Pretest and Posttests and Instruction

<table>
<thead>
<tr>
<th></th>
<th>Pretest/Posttests</th>
<th>Instructional Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Condition 1 CS</td>
</tr>
<tr>
<td>Additions</td>
<td>Cards+Dice</td>
<td>Cards</td>
</tr>
<tr>
<td>Multiplications</td>
<td>Cards+Marbles</td>
<td>Cards</td>
</tr>
<tr>
<td>Permutations</td>
<td>Cards+Spinner</td>
<td>Cards</td>
</tr>
<tr>
<td>Combinations</td>
<td>Cards+Marbles</td>
<td>Cards</td>
</tr>
</tbody>
</table>