INFORMATION TO USERS

While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.

- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.

- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or as a 17”x 23” black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6”x 9” black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.
Developmental trends in understanding an illusion based on weight adaptation: The effect of cueing questions

Shing, Marn-Ling, Ph.D.
The Ohio State University, 1987

Copyright ©1987 by Shing, Marn-Ling. All rights reserved.
PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark √.

1. Glossy photographs or pages
2. Colored illustrations, paper or print
3. Photographs with dark background
4. Illustrations are poor copy
5. Pages with black marks, not original copy
6. Print shows through as there is text on both sides of page
7. Indistinct, broken or small print on several pages
8. Print exceeds margin requirements
9. Tightly bound copy with print lost in spine
10. Computer printout pages with indistinct print
11. Page(s) lacking when material received, and not available from school or author.
12. Page(s) seem to be missing in numbering only as text follows.
13. Two pages numbered. Text follows.
14. Curling and wrinkled pages
15. Dissertation contains pages with print at a slant, filmed as received
16. Other

University Microfilms International
DEVELOPMENTAL TRENDS IN UNDERSTANDING AN ILLUSION BASED ON WEIGHT ADAPTATION: THE EFFECT OF CUEING QUESTIONS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of the Ohio State University

By

Marn-Ling Shing, B. S., M. A.

* * * * *

The Ohio State University

1987

Dissertation committee:

Gerald A. Winer, Ph.D.  
Philip M. Clark, Ph.D.  
John C. Gibbs, Ph.D.

Approved by

Gerald A. Winer  
Advisor  
Department of Psychology
To My Parents

For Their Love, Trust, and Confidence in Me
ACKNOWLEDGEMENTS

I express sincere appreciation to Dr. Gerald Winer, my adviser, for his ideas and guidance throughout the research. Thanks go to other members of my advisory committee, Dr. Gibbs and Dr. Clark, for their time and comments on the dissertation.

I would like to thank the principals, the teachers and the students in Avery Elementary School, Beacon Elementary School, Hilliard Elementary School, J.C. Sommer Elementary School, and Brookpark Middle School, for their enthusiasm and help in completing the study.

Special thanks to Julie Ricker, Linda Malloy, Julie Woodruff, Rhonda Bichard, Brent Neubauer, Laurie L. Martin, and Bonnie Buehner, for their endless help in testing the children. Without them, the study would have been impossible.

Thanks to all my friends, who supported and encouraged me all these years.

Special gratitude to Lee, for his understanding and consideration throughout the period of my study.
VITA

June 28, 1956  ...................... Born in Taipei, Taiwan, R. O. C.

1978  .............................. B. S.
Department of Psychology
National Taiwan University

1978-1979  .......................... Teaching Assistant
Department of Psychology
National Taiwan University

1985  .............................. M. A.
Department of Psychology
The Ohio State University

1987  .............................. Graduate Research Associate
Department of Psychology
The Ohio State University
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDICATION</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>VITA</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>CHAPTER ONE: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER TWO: LITERATURE REVIEW</td>
<td>9</td>
</tr>
<tr>
<td>CHAPTER THREE: STUDY ONE</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>41</td>
</tr>
<tr>
<td>Subjects</td>
<td>44</td>
</tr>
<tr>
<td>Materials</td>
<td>45</td>
</tr>
<tr>
<td>Procedures</td>
<td>46</td>
</tr>
<tr>
<td>Treatments</td>
<td>51</td>
</tr>
<tr>
<td>Results</td>
<td>55</td>
</tr>
<tr>
<td>Summary and Discussion</td>
<td>65</td>
</tr>
<tr>
<td>CHAPTER FOUR: STUDY TWO</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>68</td>
</tr>
<tr>
<td>Subjects</td>
<td>70</td>
</tr>
<tr>
<td>Methods</td>
<td>71</td>
</tr>
<tr>
<td>Results</td>
<td>72</td>
</tr>
<tr>
<td>Overall Summary</td>
<td>118</td>
</tr>
<tr>
<td>CHAPTER FIVE: GENERAL DISCUSSION</td>
<td>120</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>130</td>
</tr>
</tbody>
</table>
Appendix A:
  Questionnaires Used in Weight-cueing Condition .......................... 136

Appendix B:
  Questionnaires Used in Nonweight-cueing Condition .......................... 142

Appendix C:
  Questionnaires Used in Weight+nonweight Condition .......................... 148

Appendix D:
  Questionnaires Used in Appearance-Reality Condition .......................... 154

Appendix E:
  Questionnaires Used in Control Condition .................................. 160

Appendix F:
  ANOVA Summary Tables of Practice Effects in Study One and Study Two ............ 162
<table>
<thead>
<tr>
<th>TABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The ANOVA Table on Rotation Test in Study One</td>
</tr>
<tr>
<td>2. Frequency Tables of Subjects' Responses to Conflict Test in Study One</td>
</tr>
<tr>
<td>3. Subjects' Responses to Conflict Test in Each Condition in Study One</td>
</tr>
<tr>
<td>4. ANOVA Table for Subjects' Responses to Cueing Questions in Study One</td>
</tr>
<tr>
<td>5. ANOVA Table of Feel-Really Test Using Sex, Grade, and Condition as Independent Variables and Number of Errors Subjects Responded to Feel-Really Test as Dependent Variable in Study Two</td>
</tr>
<tr>
<td>6. ANOVA Table of Rotation Test When Responses Were Scored According to the Illusion They Experienced in Study Two</td>
</tr>
<tr>
<td>7. ANOVA Table of Rotation Test When Responses Were Scored in Terms of Either Illusion or in Terms of Positions of the Weights in Study Two</td>
</tr>
<tr>
<td>8. ANOVA Table of Rotation Test When Only 2 Trials with Corresponding Successful Illusion Were Analyzed in Study Two</td>
</tr>
<tr>
<td>9. Grade Differences in Response to the Conflict Test in Study Two</td>
</tr>
<tr>
<td>10. Subjects' Responses to Conflict Test in Each Condition in Study Two</td>
</tr>
</tbody>
</table>
11. ANOVA Table for Number of Errors That Subjects Made on Experiencing the Illusion ........................................109

12. The Percentage Rate of the Error Patterns Subjects Made at the Period of Experiencing the Illusion among all the Age Groups in Study Two ........................................113

13. ANOVA Table for the Errors Subjects Made in Response to the Cueing Questions ........114

14. ANOVA Table for Rotation Test in Study One When Trials as Another Independent Variable.163

15. ANOVA Table for the Results of Feel-Really Test When Trials as One of the Independent Variables in Study Two ..................164

16. ANOVA Table for the Results of Rotation Test When Responses Scored According to the Illusion They Experienced and Trials as One of the Independent Variables in Study Two ..165

17. ANOVA Table for the Results of Rotation Test When Responses Were Scored Either According to the Illusion or According to Positions of the Weights and Trials .........................166
LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Grade Effects on Rotation Test When Responses Were Scored in Terms of Illusions Only</td>
<td>81</td>
</tr>
<tr>
<td>2. Condition Effects on Rotation Test When Responses Were Scored in Terms of Illusions Only</td>
<td>83</td>
</tr>
<tr>
<td>3. Interaction Effects on Rotation Test When Responses Were Scored in Terms of Illusions Only</td>
<td>84</td>
</tr>
<tr>
<td>4. Grade Effects on Rotation Test When Responses Were Scored in Terms of Either Illusion or the Positions of the Weights</td>
<td>90</td>
</tr>
<tr>
<td>5. Condition Effects on Rotation Test When Responses Were Scored in Terms of Either Illusion or the Positions of the Weights</td>
<td>92</td>
</tr>
<tr>
<td>6. Interaction Effects on Rotation Test When Responses Were Scored in Terms of Either Illusion or the Positions of the Weights</td>
<td>94</td>
</tr>
<tr>
<td>7. Grade Effects on Rotation Test When Only 2 Successful Illusion Trials Were Scored</td>
<td>98</td>
</tr>
<tr>
<td>8. Condition Effects on Rotation Test When Only 2 Successful Illusion Trials Were Scored</td>
<td>100</td>
</tr>
<tr>
<td>9. Interaction Effects on Rotation Test When Only 2 Successful Illusion Trials Were Scored</td>
<td>101</td>
</tr>
<tr>
<td>10. Number of Errors Subjects Made on Experiencing the Illusion</td>
<td>110</td>
</tr>
<tr>
<td>11. Number of Errors Subjects Made in Response to Cueing Questions among Age Groups</td>
<td>116</td>
</tr>
<tr>
<td>12. Number of Errors Subjects Made in Response to Cueing Questions in Each Condition</td>
<td>117</td>
</tr>
</tbody>
</table>
CHAPTER ONE

INTRODUCTION

From Piaget's (1929) early attention to "realism" to recent research (Arnold, Moye, & Winer, 1986; Flavell, Green, & Flavell, 1986), psychologists have found it valuable to know whether and how children understand the differences between outside appearance and inner reality. As Flavell, Flavell, and Green (1983) emphasized, the distinction between appearance and reality has its own ecological significance. In everyday life, we often have many chances to experience play, fantasy and the creation of imagination. Under all of these circumstances, the information available to us is not real or maybe misleading. Therefore, we need the ability to recognize what is real and what is not.

Several studies have examined children's understanding of appearance and reality. These studies have taken different routes, including the study of the developmental steps of realism (Piaget, 1929, 1972), the processes in which children attribute physical properties to mental phenomena (Piaget, 1929; Wellman & Estes, 1986), children's differentiation between
apparent and real emotion (Harris, Donnelly, Guz, & Pitt-Watson, 1986), children's understanding of "mental verbs" (e.g. remember, forget, guess, & know) and mental processes (Johnson & Wellman, 1980; Wellman, 1985a, 1985b; Wellman & Johnson, 1979), children's understanding (Flavell et al., 1983) and knowledge about the appearance-reality distinction (Flavell et al., 1986), and children's understanding of perceptual adaptation (Arnold et al., 1986). Other researchers have focused upon the conservation of illusion-distorted tasks and its relation to reality-appearance distinction (Braine & Shanks, 1965a, 1965b; Elkind, 1966; King, 1977; Langer & Strauss, 1972; Murray, 1965, 1968). While taking different approaches, all of this research has valued the importance of understanding children's minds from the perspective of the appearance-reality distinction.

In all of these studies, age trends are important. Flavell (1986) concluded that children at three or four years of age can differentiate between what objects are "really like" and what they "feel like". Confirming Flavell's result, Harris et al. (1986) found that six year olds have the ability to make justifications of facial expressions as apparent or real. Wellman and Estes (1986) figured that the age of development started earlier. They thought that children can
differentiate between mental products (for example: dog, tree, etc.), as opposed to mental activities, (for example: dreaming, thinking, etc.) as early as three years of age. In another article, Wellman (1985b) also insisted that the understanding of the distinction between a real object and the mental image of that object occurs earlier than six or seven years of age. Arnold et al. (1986) found that not until nine years of age did children have the ability to differentiate reality from illusion. Piaget (1969) suggested that children begin to differentiate between what was a dream and what was real around seven to eight years old. If the suggestion proposed by Braine and Shanks (1965b), that conservation is correlated to appearance-reality distinction, is true, then the age at which conservation develops (usually around 6 or so) probably is related to the age of the developing reality-appearance distinction.

One reason for the apparent discrepancies in the age of the development of the distinction between reality and illusion is the fact that the different researchers have employed different measures. Flavell and his colleagues (1983, 1986), for example, have asked children to distinguish between whether an object or its properties are real or not. Braine and Shanks (1965a, 1965b) employed a similar type of task. Arnold
et al. (1986) tested children's understanding of a particular type of process, namely perceptual adaptation. And Wellman (1985a, 1985b) and Wellman and Estes (1986) asked children about their mental processes. Indeed, there are even other approaches to the study of the distinction between reality and appearance. Piaget (1929), Markman (1976), and Scarlett and Press (1975) studied children's ability to differentiate words from their connotative meanings. And Keil (1979) focused on the relationship between metacognition and the appearance-reality distinction.

In short, as is the case in other areas of research different tasks yield different age trends. Another factor which might affect performance is the possibility that the children's performance on the various tasks reflect "performance" rather than "competence" differences. That is, it is possible that children often know (have a competence for) more than they actually show (performance).

The purpose of the present study is to address the ability of children to distinguish between appearance and reality on a specific task involving the understanding of perceptual adaptation, namely weight adaptation, and to examine the types of cues used in enhancing performance. The analysis of weight adaptation for studying the appearance-reality
distinction is basically an extension of earlier work on temperature adaptation and the Locke illusion. The work on the Locke illusion stemmed from writing by the philosopher John Locke (1690) who reported that water of one temperature might simultaneously feel cold to one hand and warm to the other. In some recent research (Arnold, Winer, Wickens, 1983; Arnold et al., 1986), this illusion was created in children by first having them stick one hand in warm water and the second in cold water. They then placed both hands in a single container of water at one temperature, and experienced the water as being of two temperatures.

While both children and adults experienced the illusion, there was a difference in how they interpreted it. Children thought there was a real difference in temperature. That is, they misidentified an illusion for an objective difference, and thus showed an inability to distinguish appearance from reality. Adults, on the other hand, could identify their experience as due to an illusion.

The work on the Locke illusion is important in different respects. For one, purely on an empirical level, success on this task appears much later in development than success on other more standard tests of appearance vs. reality. On a theoretical level, the task is of interest in that the source of the illusion
clearly resides in the child's own body and perceptual process. Thus, understanding this illusion shows that children realize that their perceptual process can distort reality. This contrasts with other work on appearance and reality in which the distortions are caused by changes in objects external to the subject. That is, the source of these illusions cannot be as readily identified by the children as due to their own perceptual processes.

An effect similar to the Locke illusion can be created in a different manner, namely using weight perception instead of water. Here, instead of judging the temperatures of water, the subject judges weights of bottles after lifting objects of different weights. A central question to this dissertation is whether similar age trends occur on this illusion as on the Locke effect. Do young children misunderstand the illusion and confuse reality and appearance, and is there a change with age? Answering this question positively would lend confirmation to the earlier reported findings.

A second major goal of the study is to examine the impact of cueing on performance on the weight test. Assuming that children have difficulty with the weight task we can ask whether and how cueing them by reminding them of their previous experiences, affects
their performance. The main reason for using cues is to determine how performance on the illusion task will vary under different circumstances. That is the main goal of the study was not to analyze the way in which cues work.

Nevertheless there are two roles that cues might play in functioning to enhance performance. One is to act as "organizing enzymes". Enzymes, as they were used in biology, are defined as the substances that act as organic catalysts in chemical reaction. It can be hypothesized that the information provided by cues can be a force to help to organize and connect the existing structures to form a higher level conceptual structure. And the conceptual structure can be used in any generalized situation (Gick & Holyoak, 1983). The other is to function as "reminding factors" that simply initiate children to find existing structures that are otherwise not used. It can be hypothesized that the younger child's failure on tasks involving reality and appearance might be due to the fact that the child doesn't recall and use other types of knowledge at the child's disposal. In this situation, cues play the role of reminding the child of the place where the knowledge is stored. In other words, the child might have the underlying knowledge of, and principles pertaining to, the distinction between reality and appearance, but
might merely fail to apply them. Under each of these
two hypotheses, children should perform well after
being introduced these cues. Therefore, cues should
function as "performance facilitators". (Cues, cueing
questions or performance facilitators will be used in
this dissertation interchangeably.)
CHAPTER TWO

LITERATURE REVIEW

As was previously stated there are many ways of examining the distinction between appearance and reality. In the review that follows we will first examine Wellman's work on distinctions between images and reality. We will then examine the phenomenon of "word realism". In the next section of the paper we will turn to an examination of Flavell's work on the appearance-reality distinction and related work on the conservations. The work on perceptual adaptation, the most critical for this dissertation, will be the last work reviewed on the section on the reality-appearance distinction.

The last section of the literature review will shift focus. Here, instead of examining the reality and appearance distinction we will examine the role of cueing. More specifically we will explore how cues can be "reminding factors" or "organizing enzymes" to induce schemas and improve performance.
MENTAL IMAGE AS REAL

Mental Verbs: guess, know, forget and remember

Wellman (1985a, 1985b, 1986), basing his work on some of Piaget's (1929) investigations, examined the ability of the child to distinguish between real and mental entities. He was concerned, for instance, with whether children could distinguish between thinking, remembering, pretending and dreaming about events (having a cookie) vs. the event itself (1985a, 1986). In several early studies, Wellman (1985b) found that children did not understand the distinctions among "guess", "know", "remember", and "forget". For example, in one of his studies, Wellman and his colleagues (1979, 1980) asked children to judge if the character of a scenario was remembering or forgetting under different situations. A toy character was depicted who saw his coat being hung in one of two closets. Then, later, the character came back and chose the correct closet as the one containing his coat. The subject was asked, did the character remember? Children always answered "yes" to this question. Did the character forget? Children always answered "no" to this question.
Results showed that children had different answers under different conditions. If the character saw his coat was hung in one of the closets (previous knowledge +), and he chose the right closet afterwards (performance +), children would think the character remembered. However, if the character did not see which closet his coat was hung in (previous knowledge -), and he chose the right closet afterwards (performance +), children still thought the character "remembered" or "knew" (as opposed to "forgot" or "guessed") where his coat was. This illustrated that the children did not know that "remember" was different from "know" and "forget" was different from "guess". In short, the results demonstrated that children's meaning of various words was affected by the consequences of actions. If the character found the coat, he was said to have remembered or knew (whether or not he knew about where he originally placed it), while if the coat was not found, the character was said to have forgotten or guessed (Wellman & Johnson, 1979; Johnson & Wellman, 1980). Children perceived things which they had not experienced but which they knew are the basis of logical inference. They differentiated all the mental processes by the observable behavioral outcomes. Piaget had reported similar finding on children's moral development (Piaget, 1965). Children thought that a
character in a story should be punished by his/her behavior outcomes rather than by his/her intentions. This issue is examined more fully in works by Wimmer and Perner (1983).

At this point, Wellman suggested that children of four years of age need to use external, observable, objective objects to externalize and represent their inner worlds and these objective objects were indistinguishable from objects in their inner worlds. This externalization of mental world and diffusion of the objects in the outside and inner worlds were some of the characteristics of children's cognitive development at this period. Children at this period had no idea that guessing, forgetting, remembering, and knowing were separate cognitive mental processes which could not be judged by observable outcomes only. Piaget (1929), incidentally, reported similar findings when asking children about dreams. Children believed that dreams existed in front of them and did not exist inside of them. Apparently, children had no idea that there existed an inner self and an outside world until up to four or five years of age. In more recent works, Wellman (1985b) suggested that young children can distinguish internal, mental worlds from external events, but they did not necessarily indicate the awareness of the distinctions among different inner
mental processes. For example, children would not understand that guessing, forgetting, remembering and knowing were four distinctive mental processes until they were in late childhood.

Finally, Wellman and his colleague listed the criteria for judging objects as real as: 1. Behavioral-sensory contact. Only objects which can be sensed, touched by the subjects are real. Mental images are not real because they cannot be touched. 2. Public existence. Not only can one person sense the object, but also the object can be sensed by others. 3. Consistent existence. Not only can everyone sense the real object, but also the object continues to exist whether it serves to be the object of cognition; that is, non-real objects exist only as one is thinking about them. However, real objects exist no matter one is thinking or not. A desk is real because it is there all the time. A dream is not real because it does not exist unless I imaginally generate it (Wellman & Estes, 1986).

Wellman (1985a) did not exclude the possibility that some other criteria existed for judging the reality of objects although Wellman defined three criteria of judging the real objects from the mental objects. However, he thought that the three main criteria have to exist simultaneously in order to make
correct judgments. Children who failed to differentiate mental entities from real objects did not use these criteria fluently or/and inclusively. They might only pick up one or two standards to judge mental objects as real.

There is every reason to believe that in spite of an early distinction between real and non-real objects, which occurs about four to five years of age according to Wellman (1985b), much remains to be acquired. Specifically, to say that the child has an early understanding of reality is not to say that he/she has a full understanding of the reality-non-reality distinction or its place in a full classification of ontological categories (Wellman, 1985b). The development of distinction between real objects and their mental images may begin at three years of age (Wellman, 1985b). However, it will not finish until children are about twelve years old (Arnold, Moye, & Winer, 1986).

As children grow up, different experiences might make them realize that there is a distinction between reality and appearance for each object. The child might dress like a honey bee but is not a honey bee really. However, there is no study which fully examines the role of experiences in the understanding of appearance-reality distinction.
Word-realism

Although Piaget (1929) only mentioned realism, which was defined as attribution of physical properties to mental phenomena, different psychologists had used the term broadly (Markman, 1976; Scarlett & Press, 1975). Markman (1976) defined "Word realism" as the lack of differentiation of an attributed empirical meaning from its linguistic form. For example, the word "rain" is the linguistic form of an object. The object "rain" is wet but word "rain" is not. When she asked children whether the word "rain" is wet she found kindergartners had difficulty answering the question correctly. But the first graders could solve the test.

Scarlett and Press (1975) studied word realism somewhat differently. They analyzed word-realism through children's modes of reconstructing a story. They hypothesized that children would become more involved in a story if the story characters used the same names as the children had rather than other names. Moreover children would emphasize that part of the story with the characters with their own names or even restructure the story to emphasize the character when they were asked to recall the story. The phenomena of the restructuring and egocentric distortion of stories were expected to decrease with age. The results
supported their hypothesis. First and third graders did have significantly higher recall on the part of the story with characters sharing their own names. A problem with this study is whether it is legitimate to assume that word-realism can be measured by children's restructuring stories.

The studies we have examined have used a number of different techniques and have had a number of different goals. The Wellman research (1985a, 1985b, 1986) examined the extent to which one could distinguish between processes such as knowing, remembering, and forgetting, and they employed a sorting procedure. They asked subjects to judge acts such as thinking, under the ontology-behavioral criteria: sensory contact, public existence, and consistent existence. The work on word realism (Markman, 1976; Scarlett & Press, 1975) examined the extent to which children saw a word as having the properties of its referent or the extent to which children fuse characters in a story with themselves. We now turn to studies that are more directly relevant to this dissertation and that involve a somewhat different procedure. In these studies subjects are confronted with an illusion or a typical representation of an object, and then they are asked directly to distinguish between the reality and appearance.
REALITY AND APPEARANCE

Flavell's Works

Flavell and his colleagues recently have turned to the distinction between reality and appearance as a developmental phenomenon worthy of study in its own right (Flavell, Flavell, & Green, 1983; Flavell, Green, & Flavell, 1985; Flavell, Zhang, Zou, Dong, & Qi, 1983; Pillow & Flavell, 1985). In one task they (Flavell et al., 1983; Flavell et al., 1985; Flavell, 1986) used a sponge-made rock as a test material to ask children what the object really was. One question posed was: "Is it 'really really' a rock or is it 'really really' a sponge?". Young children erred and said it was really a rock. This type of error was termed an error of phenomenism, because the child's thought was dominated by the appearance of the object.

These investigators also identified another type of error. If they asked children about a sponge-made rock, "Does this just look like a sponge or does it just look like a rock?", the child would sometimes say it looked like a sponge. Here the subject's knowledge of what the object really was interfered with their judgments. This was thus termed a realism error and is similar to realism errors shown by other investigators.
Thus in a study by Liben and Belknap (1981), children were asked to select a picture from an array to match their view of a display of three blocks. One large block occluded the other two blocks from the children's perspective. However, children who were aware of the three blocks tended to select a picture of three blocks despite the request to report appearance only. Flavell and his colleagues identified the conditions under which each type of error occurred (Flavell et al., 1983; Taylor & Flavell, 1984).

Both of the error patterns, phenomenism and realism, were discovered when children were asked to differentiate between reality and appearance (Flavell et al., 1983). The occurrence of each error pattern was dependent on the type of materials used and the types of questions asked in the experiment. Flavell, for instance, claimed that when the investigator asked about properties of objects, such as their color or shape, children were more likely to commit the phenomenism error. On the other hand when they were asked questions about the object's identity, the children were more likely to commit the realism error. However, it is important to realize that the type of errors that can be shown are not independent of the type of questions. Realism errors can only appear on questions in which the subjects are asked on questions
of the form: "Does it look like an 'A' or a 'B'?" while phenomenism errors only occur on questions of the form: "Is this really really an 'A' or really really a 'B'?". Thus it is not clear whether the type of error is a product of the form of the question or of the content being asked about.

Why did young children even make errors involving "intellectual realism"? If children were, as Piaget (1929) suggested, directing their attention to external, objective, and perceptually salient objects only, younger children should have had more phenomenism errors than intellectual realism errors. However, both error patterns were found in all the studies related to the development of appearance-reality distinction (Flavell et al., 1983; Liben & Belknap, 1981). Flavell et al. (1983) explained that children with intellectual realism named those objects which had most "cognitively salient" attributes while they were asked to differentiate appearance from reality. When children were asked to report their knowledge of reality of an object, they tended to report what was most "up front" in consciousness at that given moment. Consequently, Flavell et al. (1983) suggested intellectual realism errors occurred when children were asked to report the appearance of an object. They tended to report the reality of the object as they looked right past the
requested visual field to the unrequested and nonvisible but cognitively salient visual world. The process was like thinking about the complex layer of a problem when only the surface problem was asked. Maybe the hypothesis did explain why children had committed intellectual realism errors. However, whether they were really conscious of appearance and responded it to reality was a different matter. In the study done by Flavell et al. (1983), no age differences were found in making these two error patterns, namely phenomenism and realism errors, but the types of questions or tasks mattered. Perhaps the types of questions and the materials of the tasks produced the salient information related to the tasks in subjects' minds.

In other words, the salient information affected the children's mind and made them exhibit one kind of error or another. The hypothesis also echoed Flavell's hypothesis on the reality-appearance distinction (1986). Flavell (1986) assumed that the reason why children could not make the appearance-reality distinction was that children could not represent two images—reality and appearance—simultaneously. When one kind of information dominated another, the reality-appearance distinction failed and different error patterns showed.
Another possible solution to this problem was suggested by Pillow and Flavell (1985). They thought that children probably were confused by the meaning of "looks like" while receiving instructions from the experimenter. "Looks like", in one hand, can refer to perceptual appearance, but it can also refer to the state of things or to one's beliefs or expectations about them in everyday language. For example, "it looks like it is going to rain" means "it is going to rain". Children who did not understand what the experimenter asked about would display the intellectual realism error pattern. Consequently, they reported the things they knew in their mind.

Age differences were another important finding in their works (Flavell et al., 1983). They suggested that three-year-olds or four-year-olds appeared to have some ability to make correct appearance-reality discriminations and that this ability increased with age.

The Appearance-Reality Distinction and Conservation

Flavell's work was not the first to examine the reality-appearance distinction. Actually a considerable number of studies examined this phenomenon (or one
similar to it). Most of these studies can be thought of as offshoots of research on conservation.

Some of the earliest work in this area was conducted by Braine and Shanks (1965a, 1965b) who attempted to relate the reality-appearance distinction to conservation as studied by Piaget. In the typical Piagetian conservation task children are shown objects or sets of objects, which initially are quantitatively identical and identical in appearance, such as two rows of seven poker chips lined up in one-for-one correspondence. At this phase the child asserts that both have the same amount. The experimenter then changes the appearance of one set, e.g., by spreading out one row of chips, and asks the child whether the identity between the items has remained. For example, the experimenter might ask the child whether the longer row of chips has fewer, more or the same number as the shorter row. Children who conserve, those around six years of age or older, assert that the initial identity is intact, since the experimenter didn't actually alter the number and only changed the appearance. Younger children, non-conservers, maintain that the difference in appearance indicates a real difference, e.g., the longer row has more. Braine and Shanks (1965a, 1965b) modified this typical procedure by employing illusions. In one study (1965b), the experimenters examined
changes due to the refraction of light. Subjects were shown a white rod that was straight and another rod that was bent. The subjects identified each rod as bent or straight. They then inserted the rods in water. The refraction of light created an illusion so that the bent rod looked straight and the straight rod looked bent. Subjects were then asked to distinguish between which rod just looked bent and which was really bent. Young preschoolers were unable to overcome the effects of appearance. They thought that the rod that appeared bent was bent. By five years of age most children solved this test. In other research, Braine and Shanks (1965a), employing a slightly different illusion, the illusion of size, reported similar findings. They assumed that the ability to distinguish real effects from illusory ones formed the beginning of the origins of the conservations (Braine & Shanks, 1965b).

A number of other studies, possibly influenced by the works of Braine and Shanks, also examined the relationship between appearance-reality distinction and conservation (Elkind, 1966; King, 1971; Langer & Strauss, 1972; Murray, 1965, 1968). Some of these studies concluded that there was some kind of relationship between appearance-reality distinction and conservation (Elkind, 1966; Murray, 1965, 1968). Elkind (1966) found that children have no difficulty attaining
conservation across an illusory transformation. This firmly did not support Piaget's theory since there were no compensatory changes in an illusory transformation. The ability to make compensatory transformation underlies conservation according to Piaget (1970).

In Elkind's study (1966), the transformation for the tasks of conservation of size and form were based upon putting a spoon in a drinking glass with half-filled water. For example, children were asked to respond whether the spoon "looked" bigger or was "really" bigger or remained the same. In testing conservation of form, children were asked whether the spoon was "really" broken or only "looked" broken or remained the same. The Müller-Lyer illusion was administered to study conservation of length. Children were asked whether the two lines which were distorted by the arms of the figure had equal lengths. In this study, Elkind (1966) did not score children's responses on the appearance-reality distinction; rather, he used these responses as indicies of children's conservation ability. His idea, that appearance-reality and conservation were related to each other, was obvious.

Murray (1968) had a similar idea to Elkind's (1966). He also studied the conservation of length and area by using illusions as stimuli. He concluded that the transition from nonconservation to conservation
occurred between the age of six and seven which was one year earlier than the age of conservation development on typical Piagetian tasks. Whether the conservation of illusion-distorted length is the same as the conservation of length as Piaget studied needs to be further explored.

Murray and Elkind both proposed that the idea of compensation that Piaget used to explain conservation did not account for the conservation of illusion-distorted length (Murray, 1965; Elkind, 1966). Instead, Murray proposed another mechanism to account for the type of conservation he studied (Murray, 1968). He assumed that to attend to the discrepancy between appearance and reality of an illusion-distorted object was a way of acquiring conservation.

In his study, he tested kindergartners and first graders on conservation of length. The Müller-Lyer illusion was used to study conservation of length. After subjects responded to the Müller-Lyer illusion, questions about the appearance-reality distinction were asked. The relation between conservation of length and the appearance-reality distinction was analyzed. Results indicated that conservers tended to respond to the appearance-reality questions correctly and nonconservers tended to make errors on these questions. However, children who understood the appearance-reality
distinction were not necessarily conservers. Consequently, Murray (1968) concluded that the ability to make the appearance-reality distinction was a necessary but not sufficient condition for illusion-distorted conservation.

In discussing whether there was any relation between conservation and appearance-reality distinction, King (1971) and Langer and Strauss (1972) reached a negative conclusion. That is they concluded that the appearance-reality distinction had no relation to conservation.

King (1971) used a nonarbitrary behavioral criterion instead of using questions of linguistic forms to test conservation of length. In this study, he asked children to pick up one out of two sticks (which had different lengths physically), to build up a bridge across two blocks. However, the lengths of these sticks were distorted by the arms of the Müller-Lyer figure to look like they were the same. He found that this method could be used to detect children's behaviors on conservation reliably. Children of five years and three months of age could do well on conservation of length while measured by the behavioral criterion. In contrast to the findings suggested by Braine and Shanks (1965b) and Murray (1968), King (1971) found that although children can pick up the right stick to build up the
bridge, they did not know whether the stick they picked was really longer or just looked longer than the other one. In other words, they had no idea of the appearance-reality distinction. Therefore, there should be no relation between conservation and the appearance-reality distinction. However, he did not consider whether children understood the linguistic forms of the questions in examining the distinction.

Langer and Strauss (1972) reached a similar conclusion while studying the relations among appearance, reality, and identity. In their study, identity was defined as the conservation on typical Piagetian tasks. They concluded that appearance-reality judgments and conservation concepts were not "coextensive" cognitive phenomena. Also, neither training in appearance-reality discrimination on illusory tasks nor on regular conservation tasks had anything to do with subject's performances on conservation concepts afterwards. Therefore, they concluded that "little coordination was found within young child's cognitions of appearance, reality, and identity".

In general, the relationship between conservation and the appearance-reality distinction is still unsettled. In those studies which found a relation between conservation and the appearance-reality
distinction, only illusion-distorted conservation tasks were used (Elkind, 1966; Murray, 1965, 1968). On the other hand, when Piaget's conservation tasks were used to study the correlation, no firm results were found (Langer & Strauss, 1972). Even though King (1971) used illusory tasks, he could not find positive correlations between the two.

**Reality vs. illusion in the understanding of adaptation**

Arnold, Moye, and Winer (1986) addressed the appearance and illusion distinction by using a different task. They used the illusion of temperature adaptation to test the knowledge of reality and appearance subjects possessed. The illusion stemmed from an observation of the philosopher John Locke (1690) that water might simultaneously feel cool to one hand and warm to the other. This illusion can be created by first having subjects simultaneously stick one hand in warm water and the other in cold water for a brief adaptation period. The subject then takes both hands and simultaneously places them in a single container of water that is at a neutral temperature. The illusion: The water feels warmer to the hand that was in the cooler water. That is the water feels like it has two temperatures. Arnold et al. (1986) tested
the children's understanding of this phenomenon by asking two questions. In one the child was asked whether the water just felt warmer or whether it was really warmer. This was labeled the "feel-really" test. In the second, the child was asked what would happen if the pan in which the illusion were felt were rotated 180 degrees and the hands reinserted in it. This is called the "rotation test". Obviously, since the illusion is due to sensory adaptation, changing the orientation of the pan will have no effect on whether the water feels warmer to one hand or the other. But if the child believes that there is a real difference in the water temperature he or she should say that rotating the pan should reverse the experience of warmth and coldness. In other words, the child is acting as if there is a real difference in temperature in the water.

The results were similar to the findings of other studies in that children's understanding of the illusion-reality distinction or reality-appearance distinction increased with age. However, Arnold et al. (1986) reported the ability to distinguish reality and appearance developed later than was the case in the research by others (Braine & Shanks, 1965a, 1965b; Flavell et al., 1983; Wellman, 1985a, 1985b; Wellman & Estes, 1986). It is worth examining how much of the path
of development depended upon the tasks used in their studies. For example, it might be that the perceptual tasks used by Arnold et al. (1986) were much more difficult than those used by Flavell et al. (1983). That is, the study that reported earlier development in the appearance-reality distinction asked subjects to identify the color of a car, or to recognize the shape of a stick, or even to report the identity of an object. The questions used in that study required subjects to distinguish between "really really" and "looks like". However, Arnold et al. (1986) used tasks which were related to subjects' ideas of illusions. Subjects had to experience the illusions to respond correctly on both the feel-really and rotation tests. Sometimes their knowledge about the phenomenon might override the subject's perception while he or she was experiencing the illusion (realism errors). Although the questions asked by Flavell et al. (1983) were similar to the feel-really test used by Arnold et al. (1986), they did not employ a behavioral measure, e.g. rotation test, as Arnold et al. did (1986).

In some unpublished research Winer (1987) extended some of these findings. Two types of test question were used. One made reference to early experiences of the child. Children were asked about some common experiences that they have with perceptual adaptation.
For instance they might be asked, "When does a bath feel warmer, if you take it after playing outside in the snow, or after playing inside in the kitchen?". The question was asked to determine their idea about temperature adaptation. Questions were also asked about weight, odor, brightness and taste. Results indicated that the recognition of these perceptual phenomena increased with age. By the third grade many children were achieving success.

However, it is important to consider that being able to recognize the adaptation phenomena in different domains does not mean that the child can differentiate the appearance and reality. For example, although children responded correctly that the bath water would feel warmer after playing outside in the snow for a long time, than after playing inside in the kitchen, they might have thought that it was really warmer. Recall King's study (1971). He suggested that even though children can pick up the stick with the right length to build a bridge they cannot make a distinction between appearance and reality. In his study, Winer did not address this question (Winer, 1987).

In this same report Winer asked children to predict consequences of their perceptual experiences. For example, with one finger in cold water and another in warm water, children were asked to predict which
finger would experience warmth if both were inserted in water of equal temperatures. Or while holding a heavy weight in one hand, and a light one in the other, children were asked which of two equal weights would feel heavier if the bottles were picked up, one in each hand. Results of these studies showed that with developing age there is an increase in the understanding of perceptual adaptation. Notice, though, that in these studies, children were only asked to predict the results of perceptual adaptation. Children did not have the chance to experience the illusion.

Finally, Winer assumed that the children were not receiving feedback from these tasks. However, when the prediction test was used to test the understanding of the phenomenon, it may have created a cognitive demand that made children think of the solution to the problem. Therefore, even though they did not know the right answer spontaneously, they might have achieved some cognitive restructuring. Whether this hypothesis is true or not needs to be studied.

Finally in some pilot research, Winer and Shing discovered that children had difficulty with the appearance-reality distinction when they were tested with weights. The paradigm followed the procedures in the study of Arnold et al. (1986) fairly closely, except children were asked to judge weights instead of
water temperatures. Thus, after children identified two containers, one held in each hand as being equal in weight, they were asked to pick up a heavy container in one hand and a light one in the other. They then lifted the two original containers, the ones that weighed the same. At this point they experienced the illusion. The weight, previously held in the hand holding the lighter of the two weights was experienced as the heavier of the two. This illusion is similar to the Locke illusion created with water (Locke, 1690).

To test the child's understanding of the illusion two types of question were asked, namely the feels-really question (Does it just feel heavier or is it really heavier?), and the rotation question (what would happen if the equal weights were reversed?). Results indicated that many sixth graders were having trouble making correct judgments on both types of questions pertaining to the weight illusion.

In sum, there is evidence that with increases in age children understand the appearance-reality distinction better. Arnold et al. (1986) found that the sixth graders reached a peak in understanding temperature adaptation. However, Flavell et al. (1983) found that children of four or five years of age understood the distinction while using another task. Winer (1987) used prediction, instead of the
feel-really and rotation tests, to examine children's understanding of weight and temperature adaptation, and also found that there was an increase in understanding of both types of illusion.

The Roles of Cues

In justifying the role of cues in children's cognitive development, researchers have had different approaches (Flavell, 1970; Gordon & Flavell, 1977; Gick & Holyoak, 1980, 1983; Reed, Ernst, & Banerji, 1974). Flavell (1970) suggested the concept of "performance deficiency" in work of memory. The idea was that the poor performance of children on various tasks did not necessarily reflect an absence of an ability to memorize, or even an absence of knowledge, but instead reflected an inability to generate the correct response. Cues might overcome the deficiency. Recently, Paris and Lindauer (1982) suggested that knowledge and awareness are the two factors which can remediate a production deficiency in memory. Frequently, children's increasingly sophisticated knowledge base permits more concepts, rules, and schemata to mediate recall. Furthermore, children's awareness of all the parameters related to metamemory might help them to recall more too. For example, the awareness of using memory strategy (rehearsal, making notes, etc.) might increase
the number of recall. As suggested above, to make children aware of some important features of a task may help them to perform better on the task. In this case, cues play the role of "attending factors" or "reminding factors" to make subjects aware of the situation.

Gordon and Flavell (1977), on the other hand, studied the role of cognitive cueing on development of intuitions in different age groups. In their study, they defined cognitive cueing as "a fact in which the occurrence of one cognitive event may instigate the occurrence of another cognitive event". However, they did not clarify the meaning of the word "instigate". The word "instigate" might indicate "association" which might be the lowest level of cognitive transfer. For instance, one cognitive event is associated with, or reminds one of, another already existing cognitive event in the mind. On the contrary, "instigate" could also mean a cognitive transfer to a higher level. That is, the cueing event might reorganize, reform or transform the existing cognitive structures to a brand-new construct.

Gordon and Flavell (1977) found that there were three cognitive developmental levels in which cues were used to find a target. Children of three years of age were apparently at the lowest level of cognitive development. They could only use well-associated cues
to find a target. For example, a hose was used to find a fireman, a band-aid was used to find a doctor. Five-year-olds were able to perform well not only using the well-associated cues but also using concrete but irrelevant cues to find the target. However, both three and five-year-olds could not use arbitrary, abstract cues to associate to a target. To use arbitrary and abstract cues to associate to a target was defined by Gordon and Flavell (1977) as the highest level of development.

However, Gordon and Flavell (1977) did not seek to find the function of cues in a higher level as we had pointed out earlier. The function of cues used in their study was to associate one cognitive event to another that required no transformation at all. In other words, association was the only function of cues in their study. Therefore, when abstract cues were used to access the target, which probably needed to be transformed into or generated to concrete examples, children of three and five years of age had trouble.

In contrast to the findings of Gordon and Flavell (1977), Gick and Holyoak (1980, 1983) had researched schema induction and analogical transfer. The materials they used in the study were two stories containing the same analogy. Subjects first received a cueing-story which contained the solution to the problem which
followed. The purpose of the study was to examine whether subjects would be aware of the fact that the cueing-story was useful and whether they could use the underlying solution to solve the following problem. In part of their study, Gick and Holyoak (1980) found that the ability to notice the first story as having cues to solve the problem and the ability to apply the cues to solve the problem were different matters. Most of the 16-18-year-olds had to have an explicit hint in order to know both of the stories were related. However, after children had the idea that the first story was presented to solve a following problem, they could easily solve the problem.

In the study published in 1983, Gick and Holyoak tried to examine the factors which influenced subjects to solve the analogical problems. First, they assumed that a schema was responsible for the transfer between problems with the same analogy. Subjects, while receiving the cueing-story and the problem needing to be solved, received the semantic cues in the story and the problem. And then the inference process underlying the mapping process used the semantic cues to form the useful schema. For instance, in the example: water to plants is like food to human beings, subjects who received the semantic cues of "water to plants" needed to abstract the relation between water and plants. The
Inference process in everyday life helped subjects to find the meaningfulness of the relation. Once the relation was found, the formation of schema was possible. The mapping process was divided into "horizontal mapping" and "vertical mapping" according to Hesse (1966). A "horizontal mapping process" was defined as the comparison process between the similar aspects of the two analogies. A "vertical mapping process" was defined as the comparison between the two parts of a single analogy (Hesse, 1966). For instance, in the problem: water to plants is like food to human beings, the horizontal mapping was to compare water to food and plants to human beings. The vertical mapping was to compare water to plants and compare food to human beings.

In the same series of studies (1983), they found that the use of only one cueing-story was not sufficient to form a schema which could be used as a mediator to transfer the solution in the story to another problem. However, two cueing-stories could provide materials for the mapping process to form the schema. It was impossible to form a schema when only diagrams or verbal statements were used to present the cueing analogy. Rather, a cueing-analogy presented in the form of stories was found to be useful.
From the above findings, it can be seen that cues can function in at least two ways. First, cues might be used to associate with the existing structures or remind one of the existence of the structures. Such cues probably are just like the cues used in Gordon and Flavell's study (1977). Another type of cue might include more information in its own right, like the story used in Gick and Holyoak's studies, so that it can help to reform a schema to solve related problems. In this dissertation, either of these means might be operating. That is, cues might function as "reminding factors" or as "organizing enzymes".

SUMMARY OF LITERATURE AND STATEMENT OF THE PROBLEM

In the review of the research we have shown that there were several means of studying the appearance-reality distinction. Analyses were made of word-realism, of children's knowledge of mental states such as forgetting and remembering, and a series of research that more directly studies the child's ability to distinguish reality from illusion or distortion.

Two distinct developmental trends emerged from this review. The work by Flavell and earlier work on conservation suggested that development occurs very early. Work involving illusions caused by perceptual
adaptation or similar effects shows that the reality-appearance distinction appears later.

One goal of the present study is to determine whether the late development, shown on the tests involving temperature adaptation occurs when a different sensation is involved, namely, the experience of weight. The purpose here is not to directly compare appearance and reality on these two tasks, but to examine a task in which an illusion is caused by prior experience with weight and to determine whether the same general age trends appear. If this trend should appear then it would show the generalizability of earlier findings.

The review of the earlier research also showed that at a very early age children possess knowledge of perceptual adaptation-like effects. This raises the possibility that cueing them of their earlier experiences will facilitate performance. Studies were reviewed showing the effects of cueing. It is assumed that such cuing will enhance children's performance on the weight tests.
CHAPTER THREE

STUDY ONE

Introduction

There were two purposes of the present study. The first was to examine age trends in the understanding of the illusion produced by weight adaptation or by the weight contrast effect. Here one of the main goals was to determine whether there were age differences that were generally parallel to those reported on the Locke illusion. A second goal was to examine the role of cueing. The cueing strategies selected were based upon the notion that reminding the child of types of knowledge that she had previously acquired might improve performance on the weight illusion test. For example, previous research has established that at a fairly young age children can correctly describe what has happened to them in the past regarding a number of types of adaptation or contrast-like experiences. If they are asked, for instance, when does a ball feel heavier, if you lift up a feather first (and put it down) or lift up a brick first, the children will correctly answer that the ball will feel heavier after
having lifted up the feather. Likewise, they can answer other questions on adaptation correctly (Winer, 1987). For example, if asked when does it seem brighter outside at noon, if you come out of your house or out of a movie, children will correctly answer that it will seem brighter after coming out of a movie. Such findings raised the prospect that asking children questions of this sort, before they experience the weight illusion, will cue them into the phenomenon they are experiencing. They will be able to relate their current experience to their previous ones.

Four types of cueing were employed. In each the subject received five simple questions on perceptual adaptation or appearance-reality distinctions, prior to the main test trials on the weight illusion. That is the subjects received the block of five cueing questions prior to the test. In one condition, the five cueing questions made reference to past experiences with weight (e.g., When does a ball feel heavier? When does a carton of milk feel heavier? etc.). The key here was that all questions referred to weight. This condition was expected to be the most effective as the relation between the cueing items and the test questions was most direct. A second condition involved five questions on adaptation or contrast effects, but here the questions made no reference to weight. Instead
the adaptation or contrast effects involved other types of sensations (taste, darkness, smell, temperature, & sound). This was employed to determine whether general cueing on adaptation would generalize to the weight test items. This condition was expected to be less effective than the one in which the cueing questions made direct reference to weight as the only similarity between these cueing items and the test items was the fact that both involved perceptual adaptation or contrast-like effects. A third cueing condition incorporated five questions, three of which were on non-weight items and two on weight items. This should be intermediate in effectiveness in comparison to the previously described cueing conditions. Finally a last cueing condition was designed to give the subjects practice on making distinctions between appearance and reality. This was employed to determine whether performance on the weight illusion test was related to a more general or basic understanding of the appearance-reality distinction.

The expectations described above were tentative. Other possibilities could be imagined. For instance, it is possible that the non-weight cueing items would be less difficult than the weight items. This raises the possibility that there would be greater facilitation in the condition with the non-weight items. Or it is even
possible that the condition with the combination of weight and non-weight items would prove to be the most effective. For instance, perhaps children would transfer knowledge from the non-weight to the weight items, developing an understanding of weight in the cueing condition itself. Or perhaps the combination of items would lead to a more general awareness of adaptation or contrast effects involving weight.

Subjects

There was a total of 140 subjects: 49 third graders, 50 sixth graders and 41 college students. Twenty-nine of the third graders were boys, and 20 were girls. The mean age of the third graders was 9 years, 4 months (range = 8 years, 5 months to 11 years, 3 months). There were 21 boys and 29 girls among the sixth graders. The mean age of the sixth graders was 12 years, 2 months (range = 11 years, 5 months to 12 years, 8 months). There were 21 males and 20 females among college students. The mean age of the college students was 20 years, 9 months (the range = 18 years, 6 months to 29 years, 9 months). Both the third graders and the sixth graders were randomly chosen from Beacon Elementary and Hilliard Elementary Schools, which were in Hilliard, a middle-class suburb of Columbus, Ohio.
College students were volunteers from the psychology 100 pool which included all the students taking psychology 100 and participating in research to receive course credits. Each subject was randomly assigned to each condition. Each subject was tested individually in a separate room for about 15 minutes.

Materials

The materials used in the study were ten pill bottles filled to different degrees with BBs, so as to have different weights. Each bottle was 3.5 inches high and 2 inches in diameter. The heaviest bottle (H) was 674 g. The lightest one (L) was 26 g. The other eight bottles were divided into four groups. Each group had two bottles with the same weights: 162 g, 242 g, 321 g, 401 g.

A simple balance scale was used to show subjects that bottles weighed the same. It was 12 inches long and 3 inches high and constructed out of plexiglass. Each pair of bottles with equal weights could be balanced on the scale.

There were five items used in one condition. The condition was named the "Appearance-Reality condition" and was based on Flavell's work. This was one of the experimental treatments to examine whether the general
concept of appearance and reality can cue subjects to understand the weight illusion. The procedure of administering the condition will be described in the section of "Treatments". One item in this condition was a white index card and a red transparent piece of plastic. The red transparent sheet made the white index card look red when placed over it. The second item used in this treatment was a paper clip and a magnifying glass. A third item used was a chopstick and a beaker, 5.5 inches high and 4 inches in diameter. The beaker was half filled with water. When the chopstick was put in the water there was an illusion that the chopstick was bent. A piece of paper was taped to the side of the beaker. The paper covered the beaker from the bottom to water level. This forced subjects to look down through the top of the beaker to see the illusion. The fourth item in the condition was a candle that looked like an ice-cream cone. It was about 5 inches high. The fifth item in the condition was a piece of plastic, that looked like a part of a Hershey candy bar. It was about 2 square inches.

Procedures

There were five treatments in this study. In each condition subjects received cueing questions of
different kinds and followed by the four main weight-illusion test trials. The main test trials will be described first.

Main Test Trials. Each of the four main test trials had four phases. In the first phase the subject witnessed that two bottles weighed the same. In the second, an illusion was created that two equal weights, one in each hand, felt different. That is one felt heavier than the other. There were then two critical questions that served to determine how subjects interpreted the illusion. One, termed the "feel-really test" asked the subject whether the experience of a weight difference was real or whether it just felt like there was a difference. For the second test, the "rotation test", the experimenter asked the subject what would happen if the bottles were switched, so that the hand holding the heavier bottle now held the lighter appearing one. If the subject believed that the perceived difference was due to an illusion then switching hands should make no difference. That is, the hand holding the heavier bottle should still feel like it is holding the heavier one after the bottles are switched. If the subject believed that there was a real difference then he/she should assert that reversing the bottles will change the experience, just as if there were a real weight
difference. The following presents a more specific description of each trial.

On each trial the experimenter first selected one of the four sets of pill bottles that weighed the same, and demonstrated to the subject that they were identical in weight by placing them on the simple balance scale that appeared directly in front of the subject. Each subject was asked, "Does one weigh more or do they both weigh the same?". If the subject was correct the experimenter recorded the response; otherwise the subject was corrected. These two bottles that both weighed the same were removed from the scale and placed between the scale and the subject. At this point the subject was given the set of unequal weights, one heavy and one light, and asked to lift up one in each hand. (The hand, right or left, to which the heavier of the two bottles was presented was randomized across trials). While holding the two bottles that were of different weights the subject was asked which was heavier than the other. There was an obvious difference between these. Subjects held these bottles for about 15 seconds.

To create the illusion, the subjects were asked to place the two unequal bottles down and to pick up the two bottles that had previously been on the scale and identified (by the subject or experimenter) as equal in
weight. The exact instructions required the subjects to pick the two equal bottles up and put them down and indicate which one of the two felt heavier. (The reason for putting the bottles down is that the weight illusion is transitory, vanishing after a few seconds.)

The first main test question, the "feel-really" item was introduced at this point. For this question the subject was asked, "Is it really heavier or does it just feel heavier?" or "Does it just feel heavier or is it really heavier?". The exact version presented was randomized across trials.

The second main test question, the "rotation test", was then administered. Here the experimenter took the two equal bottles of weights, rotated them so that the bottle that was in front of the left hand was now in front of the right hand (and vice versa) and asked the subjects what would happen if the bottles were picked up (e.g., "If you pick this one up in this hand, and his one in this hand, which will feel heavier?").

Each subject received four weight-illusion tests as described. The only difference between these trials involved the sets of equal weights that were used. One of the four different sets was used on each trial. Remaining constant across trials was the weight of the
heavy and light bottles that the subject held to create the illusion.

Following the administration of the last weight-illusion trial, subjects were asked for explanations. At this point there was also added a "conflict-test" designed as a diagnostic means to determine some of the thoughts of the subjects. The gist of the test was to ask subjects what would happen if, following the rotation trial, the weights were to be placed back upon the scale. The correct answer is that the scale shouldn't tilt as the weights are still the same. If the subject believes that there is a real difference, though, he or she might say that the scale would tilt. In any event, some sort of dissonance or conflict should be created between a belief that there is a real difference in weight and the prior knowledge that the bottles weighed the same.

In fact, the question was designed to create conflict. Immediately after the explanations given on the last, i.e., the fourth rotation trial, the subject was told to remember which bottle was the heavier. The subject was asked, "What would happen if the bottles are placed back on the scale?". If the subject said the scale would tilt he or she was reminded that they originally were the same. How come they were different now? If the subject predicted there would be no tilt,
he or she was told that one was different from the other (in the rotation trial) and what made the discrepancy.

Treatments

The experimental treatments consisted of five cueing or control questions that subjects received prior to four main tests on the weight illusion. Each subject received one of the five sets of questions in a block followed by the four test questions. The treatments were as follows:

Weight-cueing condition. In this condition there were five questions which were related to the concept of "weight adaptation". For example, one of the cueing questions was that, "Lots of times we pick something up and put it down. Then we pick something else up. Sometimes the things we pick up feel heavier than they do at other times. Tell me, when does a ball feel heavier if you lift a feather up and put it down before you pick up the ball or if you lift up a brick first?". The four remaining questions were similar but referred to different objects that were lifted up. The order of the key phrases in the sentences were randomized (in this condition and in the remaining ones as well). For example, some of the subjects were asked, "When does a
ball feel heavier, if you lift a feather up and put it down or if you lift up a brick first?". Some of the subjects were asked, "When does a ball feel heavier, if you lift a brick up first and put it down or if you lift up a feather first?". All of the cueing questions in this condition are presented in Appendix A.

Non-weight-cueing condition. In this condition, subjects received five cueing questions which were not related to the concept of weight adaptation. However, the cueing questions in this condition were related to perceptual adaptation of other sensory modalities. For example one question asked about adaptation to darkness: "When we go out of the house we see the sunshine and it is bright. Sometimes it seems like it is brighter than it is at other times. Let's say you go outside at lunch time. When does it seem like it is brighter, if you come out of your house or out of a movie?". The other kinds of cueing questions in this experimental treatment were related to taste, smell, temperature, and sound, and they are presented in Appendix B.

Weight plus Non-weight-cueing condition. In the third experimental condition, subjects received cueing questions which combined the concept of non-weight adaptation with the concept of weight adaptation. In this condition each subject received five questions,
two involving weight and three involving other dimensions (temperature, taste, and darkness). Each weight item was preceded and followed by a nonweight item, i.e., nonweight, weight, nonweight, etc. This order was employed to allow the subjects the opportunity to improve on the weight items due to experience with the other cueing questions. Except for the order in which the weight and non-weight questions were presented, the order of the cueing questions was randomized. The questions used in the condition are presented in Appendix C.

Appearance-Reality condition. The fourth condition which was based on the work of Flavell and his colleagues (Flavell, Flavell, & Green, 1983). In their study, Flavell et al. used different kinds of materials to test the concept of appearance and reality of children. In present study, the experience in distinguishing appearance and reality was used to facilitate children's understanding of weight illusion. Three objects whose appearances were distorted and two other objects whose real natures were distorted were used in present study. On one item a white index card was placed behind a red transparent piece of plastic causing the card to appear to be red. Then, subjects were asked: "Is it really really white or is it really really red?". Another item involved a paper clip that
was enlarged by a magnifying lens. Subjects were asked: "Is it really really big or is it really really small?". For a third item a chopstick was placed in a beaker that was half filled with water. This made the chopstick appear to be bent. Subjects were asked: "Is it really really bent or is it really really straight?". (To achieve this effect the bottom half of the beaker, i.e., that containing the water was covered with a piece of paper taped onto the beaker. Subjects were forced to look down through the top, uncovered half of the beaker, at the chopstick which was immersed in the water.) A fourth cueing question was related to the reality of an object. A wax candle, designed like an ice cream cone was used in this question. Subjects were asked to touch the ice cream cone before they answered the question which was, "Is it really really an ice cream cone or is it really really wax?". For a fifth question, subjects were shown plastic fashioned to look like a piece of a Hershey candy bar. After touching the plastic subjects were asked: "Is it really really a candy bar or is it really really plastic?". In this question, as in all the questions in this condition, the order of key phrases in the question was randomized. The questions are presented in Appendix D. Control condition. A control condition was used as the fifth condition. All the questions in this condition
were irrelevant to the adaptation concept; for example, "How many days are there in a year?" or "Who is the President of the United States of America?". All these questions are presented in Appendix E.

Results

A 2 (sex) by 3 (grade) by 5 (condition) analysis of variance was used to analyze the data. The number of errors each subject had in each test was used. In the following section, results will be presented according to the tests that subjects were administered to test their understanding in weight illusion.

(1) Feel-really Test

Subjects who reported that it was really heavier than the other were scored as "incorrect". Subjects who reported that it just felt heavier than the other were scored as "correct". However, subjects who answered that it felt heavier than the other were scored as "correct" also.

The results of analysis of variance on the feel-really test indicated that there were no sex differences (F(1,110) = 3.82, p>.05), no age differences (F(2,110) = 2.20, p>.10), and no
condition effects \( (F(4,110) = 1.99, p > .10) \) on subjects' understanding of weight illusion.

(2) Rotation Test

The scores subjects had on the rotation test were dependent on the subjects' responses to the illusion. If subjects reported that the bottle at the "left-hand" side felt heavier at the period of experiencing the illusion, subjects who reported that the bottle at the "left-hand" side would still feel heavier after rotation were scored as "correct". Subjects who reported that the bottle at the "left-hand" side felt heavier at the period of experiencing the illusion, then reported that the bottle at the "right-hand" side would feel heavier were scored as "incorrect".

Subjects who did not experience the illusion, that is, they reported that both of the bottles felt the same at the period of experiencing the illusion, had to report that both bottles would still feel "the same" after rotation to be scored as correct. Subjects who experienced the wrong illusion (e.g. "right-hand" side felt heavier), had to report that the bottle at the same side (e.g. right-hand side) would feel heavier after the rotation to get credit. In this study, 80% of the subjects experienced the correct illusion, 12%
of the subjects did not experience the illusion and 8% of the subjects experienced the "wrong" illusion. Results of the analysis of the number of errors are shown in Table 1.

The results of the Tukey test showed that the third graders performed significantly lower than both the sixth graders and college students did (mean errors of the third graders = 3.18, mean errors of the sixth graders = 2.26, and mean errors of college students = 2.12). Both of the differences were significant at 0.05 level. The sixth graders did not perform significantly differently from college students on the rotation test.

However, the grade effects were the only significant main effects in this study. Condition effects, on the other hand, were not significant (F(4,110) = 1.51, p>.2). In other words, the cueing questions in each condition made no differences in improving subjects' performance on the rotation test.

(3) Conflict Test

In this analysis, subjects who answered that the scale would tilt if the bottles were put back on the scale were scored as incorrect. Subjects who reported that the scale would be balanced if the bottles were put back on the scale were scored as correct. A chi
Table 1: The ANOVA Table on Rotation Test in Study One

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.471</td>
<td>1</td>
<td>0.471</td>
<td>0.23</td>
</tr>
<tr>
<td>Grade</td>
<td>24.975</td>
<td>2</td>
<td>12.488</td>
<td>6.07  ***</td>
</tr>
<tr>
<td>Cond</td>
<td>12.431</td>
<td>4</td>
<td>3.108</td>
<td>1.51</td>
</tr>
<tr>
<td>Sex * Grade</td>
<td>4.198</td>
<td>2</td>
<td>2.099</td>
<td>1.02</td>
</tr>
<tr>
<td>Sex * Cond</td>
<td>4.444</td>
<td>4</td>
<td>1.111</td>
<td>0.54</td>
</tr>
<tr>
<td>Grade * Cond</td>
<td>7.168</td>
<td>8</td>
<td>0.896</td>
<td>0.44</td>
</tr>
<tr>
<td>Sex * Grade * Cond</td>
<td>15.042</td>
<td>8</td>
<td>1.880</td>
<td>0.91</td>
</tr>
<tr>
<td>Error</td>
<td>226.262</td>
<td>110</td>
<td>2.057</td>
<td></td>
</tr>
</tbody>
</table>

*** p < .001
square test was used to examine the independence between grade and subjects' responses on conflict test. The relation between conditions and subjects' responses on conflict test was also examined. Results of the analysis indicated that the null hypothesis that grade and subjects' responses on conflict tests were independent of each other was accepted (Chi square (140) = 3.17, p>.20). Subjects' responses were not related to the cueing questions in each condition (Chi square (140) = 4.43, p>.35). Results of the analyses are shown in Table 2 and Table 3.

(4) Experiencing the Illusion

As mentioned above, subjects who either did not experience the illusion (by reporting that both of the bottles felt the same) or experienced the wrong illusion (by reporting the hand lifting up the heavy weight beforehand felt heavier when lifting up the equal weights) were scored as incorrect at the period of experiencing the illusion. The total number of errors subjects made at this period was analyzed. Results showed that there were no grade effects (F(2,110) = 0.55, p>.5) and no condition effects (F(4,110) = 0.13, p>.9) on experiencing the illusion.
Table 2: Frequency Table of Subjects' Responses to Conflict Test in Study One

<table>
<thead>
<tr>
<th>Response to Conflict Test</th>
<th>0 (correct)</th>
<th>1 (incorrect)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 3</td>
<td>34</td>
<td>15</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>69.39 %</td>
<td>30.61 %</td>
<td></td>
</tr>
<tr>
<td>Grade 6</td>
<td>38</td>
<td>12</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>76.00 %</td>
<td>24.00 %</td>
<td></td>
</tr>
<tr>
<td>College</td>
<td>35</td>
<td>6</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>85.37 %</td>
<td>14.63 %</td>
<td></td>
</tr>
</tbody>
</table>

107 33

Chi square (140) = 3.17  p > .20
Table 3: Subjects' Responses to Conflict Test in Each Condition in Study One

<table>
<thead>
<tr>
<th>Condition</th>
<th>0 (correct)</th>
<th>1 (incorrect)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>25</td>
<td>4</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>86.21 %</td>
<td>2.86 %</td>
<td></td>
</tr>
<tr>
<td>Non-weight</td>
<td>22</td>
<td>6</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>78.57 %</td>
<td>21.43 %</td>
<td></td>
</tr>
<tr>
<td>Weight +</td>
<td>18</td>
<td>10</td>
<td>28</td>
</tr>
<tr>
<td>Non-weight</td>
<td>64.29 %</td>
<td>35.71 %</td>
<td></td>
</tr>
<tr>
<td>Appearance and Reality</td>
<td>21</td>
<td>8</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>72.41 %</td>
<td>27.59 %</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>21</td>
<td>5</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>80.77 %</td>
<td>19.23 %</td>
<td></td>
</tr>
</tbody>
</table>

Chi square (140) = 4.34  \( p > .30 \)
(5) Responses to Cueing Questions

The number of errors subjects made in answering the cueing questions in each condition was calculated. Results of an analysis of variance are shown in Table 4.

The results of analyzing the errors of the cueing questions indicated that age groups were different in responding to these questions. College students made the fewest errors on the cueing questions compared to the other two groups (mean errors of college students = 0.44, mean errors of the sixth graders = 0.92, and mean error of the third graders = 1.78, Tukey test, both p<.05). And the sixth graders made significantly fewer errors on cueing questions than the third graders did (Tukey test, p<.05).

In addition to grade effects, there were significantly different degrees of difficulty on the questions in different conditions. Subjects made the most errors on the questions in the Weight condition and made the least errors on the irrelevant, i.e. control, questions. Apparently, the cueing questions in the Weight condition were the most difficult ones among all questions. Subjects made significantly more errors on the questions in this condition than on the
Table 4: ANOVA Table for Subjects' Responses to Cueing Questions in Study One

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.041</td>
<td>1</td>
<td>0.041</td>
<td>0.03</td>
</tr>
<tr>
<td>Grade</td>
<td>40.765</td>
<td>2</td>
<td>20.383</td>
<td>15.47***</td>
</tr>
<tr>
<td>Cond</td>
<td>19.603</td>
<td>4</td>
<td>4.901</td>
<td>3.72 **</td>
</tr>
<tr>
<td>Sex * Grade</td>
<td>1.237</td>
<td>2</td>
<td>0.619</td>
<td>0.62</td>
</tr>
<tr>
<td>Sex * Cond</td>
<td>2.647</td>
<td>4</td>
<td>0.662</td>
<td>0.73</td>
</tr>
<tr>
<td>Grade * cond</td>
<td>24.349</td>
<td>8</td>
<td>3.044</td>
<td>2.31 *</td>
</tr>
<tr>
<td>Sex * Grade * Cond</td>
<td>15.534</td>
<td>8</td>
<td>1.942</td>
<td>1.47</td>
</tr>
<tr>
<td>Error</td>
<td>144.940</td>
<td>110</td>
<td>1.318</td>
<td></td>
</tr>
</tbody>
</table>

*** p < .001
**  p < .01
*   p < .05
questions in the Non-weight condition (mean error on the questions in the Weight condition = 1.62, mean error on the questions in the Non-weight condition = 0.79, p<.05). Subjects also made significantly more errors on the questions in the Weight condition than on the questions in control condition (mean error on the questions in the Weight condition = 1.62, mean errors on the questions in control condition = 0.62, p<.05).

(6) Practice Effects

As was mentioned above, there were four trials in the study. Therefore, an analysis was conducted to determine whether trials had an effect on subjects' understanding of the weight adaptation. In the analysis, a 2 (sex) by 3 (grade) by 5 (condition) by 2 (trials) mixed designed was used. Subjects' performance on the first two trials and the last two trials were calculated. The number of errors subjects made on the feel-reality test and the rotation test was calculated respectively.

Feel-reality Test

There was no significant practice effect in the analysis. Subjects neither performed better under
different cueing conditions nor improved their performance over the trials.

Rotation Test

The results of the analysis of variance showed that there were significant grade effects and significant practice effects over the trials ($F(2,110) = 6.57, p<.01$; $F(1,110) = 4.76, p<.05$). The summary Table of the analysis will be in Appendix F. Results showed that subjects performed better on the last two trials than on the first two trials (mean errors of the first two trials $= 1.34$, mean errors of the last two trials $= 1.20$, $p<.05$).

Summary and Discussion

In general, there were significant grade effects on subjects' performance on the rotation test. College students and sixth graders performed significantly better than the third graders. However, performance among the college students and sixth graders was extremely poor. The mean number of errors was over 2 (out of four). Clearly, many of these students were not showing success on the rotation trials. This high
degree of difficulty possibly accounts for the failure of the condition effect.

The overall level of difficulty on the rotation trials is particularly surprising in light of work on the original Locke illusion which suggested that by about the sixth grade level success was being achieved on rotation trials. Work on the understanding of illusions also (Winer, 1987) suggests that students have the basis for understanding perceptual adaptation in the early years. In his research most students were answering simple questions about the adaptation experiences correctly by the sixth grade. Winer (1987) also showed that college students achieved high degrees of success on trials in which they were asked to predict the illusion, although this occurred when the questions provided implicit cues that there would be some adaptation. In short there is reason to believe that results of this study understated what the subjects really know about adaptation. However, by the same token, the poor performance of the adults and the failure of the cueing questions also make it clear that the subject's awareness of adaptation is not particularly keen.

There were a few other findings of interest. There was some evidence that subjects performed better on the last two trials than on the first two trials. There
were no age differences in experiencing the illusion. However, subjects with different ages did have different abilities in answering those cueing questions. Cueing questions in different condition had different degrees of difficulty also.
CHAPTER FOUR

STUDY TWO

Introduction

The results of Study One were very surprising, especially in comparison to work on the original Locke illusion and unpublished work on the child's understanding of weight adaptation (Arnold, Moye, & Winer, 1986, Winer, 1987). For instance, work on the Locke illusion showed that children were achieving success by the sixth grade level (Arnold et al., 1986). And prior work has shown that many children and college students can predict the effects of weight adaptation (Winer, 1987). Thus, if a student is holding a heavy weight in one hand, and a light one in the other, and is asked to predict what will happen if he places these unequal weights down and picks up two equal ones, the student can respond with an understanding of adaptation, i.e., that the weights will not feel the same, but that the hand holding the heavier weight will feel like it is holding a lighter one.

The results of that prior research also suggested (Winer, 1987), though, that the extent to which children and adults predict the effects of adaptation
depend upon the cues provided. If a forced choice question is used, asking which will feel heavier —essentially telling the subject that there will be some effect— the subjects will answer correctly. But if the question asks, "Which will feel heavier or will they weigh the same?", many students err, indicating that they will weigh the same. These results suggest that subjects might have the knowledge but that they might not use it unless the task is structured so as to almost force them to give the correct answer.

It thus appears that with more definition to the task, improvements might occur. The present study is essentially a replication of the procedure of the first study, with one added cue designed to provide assistance to the subjects. This cue came on the test trials only.

Recall that in the original test trials subjects first held unequal weights, and then lifted (and placed down) the equal weights where they experienced the illusion. After they experienced the illusion, the feel-really and rotation trials were given.

The procedure of the present study replicated the procedure of Study One with one exception. Prior to the reversal of the positions of the heavy and light weights on the rotation test, the subjects were asked to pick up the two bottles of unequal weights and hold
them. That is, each rotation test occurred while the subjects were holding the unequal weights. This added cue of holding the weights should serve as a reminder to the subjects, signalling the importance of earlier experiences to later ones.

The idea for this manipulation derived from some of the responses of the subjects in the first study. They simply indicated that reversing the orientation of the bottles would change the experience. It seemed so obvious to them that they did not apparently reflect before answering the question. Holding the bottles of unequal weights at the very least reminds the subjects of one phase of the procedure which might influence later performance.

Subjects

There were three age groups in this study: third graders (mean age = 9 years, 4 months, range = 8 years, 5 months to 10 years, 6 months), sixth graders (mean age = 12 years, 4 months, range = 11 years, 4 months to 13 years, 9 months), and college students (mean age = 20 years, 4 months, range = 18 years, 4 months to 37 years, 7 months). The third graders were randomly chosen from J. C. Sommer Elementary School in Grove City, a middle class suburb of Columbus, Ohio. The
sixth graders were chosen from Brookpark Middle School, which is in Grove City also. College students were volunteers from the subject pool in the Ohio State University, which included all the students who took psychology 100 and had to participate in research to receive credits. There was a total of 225 subjects which included 79 third graders, 79 sixth graders and 67 college students. Among the third graders, there were 39 boys and 40 girls. There were 41 boys and 38 girls among the sixth graders. There were 42 males and 25 females among college students. Each non-college subject was individually tested in a separate room near the classroom and randomly assigned to experimental conditions. College students were tested individually in the Department of Psychology. Each subject received five cueing questions, four trials on the weight illusion test, one explanation trial and one conflict test. It took about fifteen minutes for each subject to finish the study.

Methods

The materials and procedure in this study were identical to those of Study One with one exception which occurred on each test trial. Here, as was described, subjects were required to lift up the
unequal bottles after the feels-really trial and before the rotation trial. That is, the subjects were required to lift up the heavy and light bottles, one in each hand. Then the experimenter reversed the orientation of the bottles of equal weights and asked the subject what would happen if he or she put the bottles down and lifted the others up.

As in the first study, five different cueing conditions were employed. Each subject received questions from one cueing condition, or control questions, followed by the four test items.

Results

The results of Study Two are presented below. The first results will be presented are major findings and included subjects' responses to A) the "feel-really tests", B) the "rotation tests", and C) the "conflict test"; a category of additional findings, which includes subjects' responses to experiencing the illusion and their responses to cueing questions, will next be presented.

Major Findings

The major objective of this study was to examine whether different kinds of cueing questions, as were
operationalized in five experimental treatments, would facilitate children's or even adults' performances on the tasks of weight illusion. Also important were the presence and absence of grade effects. This was a significant variable insofar as Study One proved that there were meager differences. That is the age differences in Study One were not as striking as were expected, based upon the prior research with the Locke illusion (Arnold et al., 1986).

The measurements of understanding of weight illusion in this study were the "feel-really test", the "rotation test" and the "conflict test". Among all these measures, the rotation test was regarded as the most powerful. Prior research suggested that the feel-really test was particularly ambiguous (Arnold et al., 1986). It should also be noted that a pilot study suggested that the presence and absence of the "feel-really" test had no impact on performance on the rotation trials.

In the following section, results will be presented according to the measurements used to test the concept.

(1) Feel-really Test

The feel-really test was administered after subjects experienced the illusion in each trial.
Subjects were asked whether one of the bottles of equal weights was really heavier than the other or whether it just felt heavier than the other. Subjects who answered that it was really heavier than the other were scored as incorrect. On the other hand, subjects who answered that it just felt heavier or that it felt heavier were scored as correct. The number of errors each subject made was used in the analysis. The maximum number of errors that could be made was 4.

Table 5 presents the summary table of the results of the analysis of variance on errors from the feel-really test.

Grade Effects

Table 5 shows that there were no condition effects in understanding the weight illusion when the feel-really test was used as the measure. However, there was an effect due to grades (F(2,195) = 5.64, p<.01). When post-hoc paired-comparisons were used to test the difference between grades, college students were found to make significantly fewer errors than the sixth graders and third graders (Tukey test, p<.001). The mean number of errors that college students made was 0.66, which was significantly fewer than those of the third graders (mean number of errors = 1.18) and the sixth graders had (mean number of errors = 1.25).
Table 5: ANOVA Table of Feel-Really Test Using Sex, Grade and Condition as Independent Variables and Number of Errors Subjects Responded to Feel-Really Test as Dependent Variable in Study Two

<table>
<thead>
<tr>
<th>Sources of Variances</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.085</td>
<td>1</td>
<td>0.085</td>
<td>0.07</td>
</tr>
<tr>
<td>Grade</td>
<td>13.767</td>
<td>2</td>
<td>6.884</td>
<td>5.64 **</td>
</tr>
<tr>
<td>Cond</td>
<td>7.486</td>
<td>4</td>
<td>1.872</td>
<td>1.53</td>
</tr>
<tr>
<td>Sex * Grade</td>
<td>1.396</td>
<td>2</td>
<td>0.698</td>
<td>0.57</td>
</tr>
<tr>
<td>Sex * Cond</td>
<td>5.346</td>
<td>4</td>
<td>1.337</td>
<td>1.10</td>
</tr>
<tr>
<td>Grade * Cond</td>
<td>10.157</td>
<td>8</td>
<td>1.270</td>
<td>1.04</td>
</tr>
<tr>
<td>Sex * Grade * Cond</td>
<td>4.389</td>
<td>8</td>
<td>0.549</td>
<td>1.22</td>
</tr>
<tr>
<td>Error</td>
<td>237.869</td>
<td>195</td>
<td>1.220</td>
<td></td>
</tr>
</tbody>
</table>

** p < .01
(2) Rotation Test

The difference in the procedure employed in this study, namely having the subjects lift up the unequal weights prior to each rotation trial, permitted a number of different ways of scoring the data. On the one hand the data could be scored as they were in the first study. That is, the response to rotation trial could be scored in terms of how the subject responded to the illusion. However, it was also possible to score the responses to the rotation trials independent of the subject's response to the illusion. In this case, instead of using the illusion to determine the correctness or incorrectness of the performance, the position of the weights as they were being held could be employed. To illustrate this procedure, let us assume that a subject fails the illusion trial by saying that the two weights felt the same. The subject next lifts the unequal weights up, one in each hand, during the rotation. In this instance the correct response on the rotation trial can be determined by the position of the heavy and light weights as the subject is holding them. If the subject is holding the light weight in the left hand, then the correct response would be that the weight will feel heavier in that hand when he or she picks it up. In this instance the
correct response is independent of whether the subject responded correctly on to the illusion.

One reason for employing this method of scoring is the possibility that subjects who said they were not experiencing an illusion really were in fact doing so. They might have asserted that there was no perceived difference when asked whether there was an illusion, because they initially knew that the weights were equal. (This type of error, incidentally, can be characterized in different ways but it represents what might be termed a realism error, that is, the subject responds in terms of what is really the case instead of what is perceived.) Upon experiencing the rotation itself, however, the subjects might have responded to what they perceived instead of what they reported on the phase where the illusion should have been felt. (This same method of scoring might have been employed in Study One. That is, I could have scored the result of the rotation trial in terms of the position of the heavy and light weights, independently of how the subject experienced the illusion. However, it made little sense to do it in that study, and more sense to do so in the present investigation. That is, in Study Two it was assumed that holding the weights during the rotation would have an impact on the subject serving as
an additional cue. It is thus meaningful to examine responses scored in this way.)

Several different analyses follow. In these analyses we examine the impact of the conditions under different methods of scoring and different combinations of these methods. In this study, 78% of the subjects experienced the correct illusion, 12% of them reported no illusion, and 10% of them experienced the "wrong" illusion.

A. Responses Scored in Terms of the Illusion

In the analyses presented in these section, the correctness or incorrectness of the response to the rotation trials depended upon the subject's response to the illusion. This was the way I scored the rotation tests in Study One. If the subject, for instance, erroneously claimed on the illusion trial that the right hand was holding a heavier weight, the correct response on the rotation trial was that the right hand would continue to be holding the heavier of the two. A 3 (grade) by 2 (gender) by 5 (condition) ANOVA was conducted on responses to the rotation trials scored in this fashion. The summary appears in Table 6. There were significant grade effects and condition effects.
Table 6: ANOVA Table of Rotation Test When Responses Were Scored According to the Illusion They Experienced in Study Two

<table>
<thead>
<tr>
<th>Sources of Variances</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.944</td>
<td>1</td>
<td>0.944</td>
<td>0.58</td>
</tr>
<tr>
<td>Grade</td>
<td>55.116</td>
<td>2</td>
<td>27.558</td>
<td>16.80***</td>
</tr>
<tr>
<td>Cond</td>
<td>23.242</td>
<td>4</td>
<td>5.811</td>
<td>3.54 **</td>
</tr>
<tr>
<td>Sex * Grade</td>
<td>1.732</td>
<td>2</td>
<td>0.862</td>
<td>0.53</td>
</tr>
<tr>
<td>Sex * Cond</td>
<td>2.373</td>
<td>4</td>
<td>0.593</td>
<td>0.36</td>
</tr>
<tr>
<td>Grade * Cond</td>
<td>19.095</td>
<td>8</td>
<td>2.387</td>
<td>1.46</td>
</tr>
<tr>
<td>Sex * Grade * Cond</td>
<td>8.006</td>
<td>8</td>
<td>1.001</td>
<td>0.61</td>
</tr>
<tr>
<td>Error</td>
<td>319.808</td>
<td>195</td>
<td>1.640</td>
<td></td>
</tr>
</tbody>
</table>

*** p < .001
**  p < .01
Grade Effects

The information presented in Table 6 indicates that there were highly significant grade differences (F(2,195) = 16.80, p<.001). As expected, college students made significantly fewer errors on the rotation test than the sixth graders did (mean errors of college students = 1.37, mean errors for sixth graders = 2.06, p<.01). College students performed significantly better than the third graders did also (mean errors for college students = 1.37, mean errors for third graders = 2.63, p<.001). Finally the sixth graders were significantly better than the third graders were on the rotation test (mean errors for sixth graders = 2.06, mean errors for third graders = 2.63, P<.01; Tukey test).

Figure 1 shows the relationship between age groups and their performance on rotation test.

Condition Effects

As shown in Table 6, the conditions had significant effects on subjects' judgments on weights also (F(4,195) = 3.54, p<.01). Post hoc analyses revealed that subjects who received the cueing questions related to the concept of weight adaptation performed better than all those subjects who were in other conditions. The mean number of errors on weight
Figure 1: Grade Effects on Rotation Test When Responses Were Scored in Terms of Illusions Only
judgments in the Weight-cueing condition was 1.51, which was significantly fewer than those in the Non-weight-cueing condition (mean errors = 2.17, \( p < .05 \)), the Appearance-Reality condition (mean error = 2.47, \( p < .001 \)), and the control condition (mean error = 2.12, \( p < .05 \)). Furthermore, subjects in the Weight+nonweight-cueing condition made significantly fewer errors than those who were in the Appearance-Reality condition (mean errors in Weight+nonweight-cueing condition = 1.80, mean errors in Appearance-Reality condition = 2.47, \( p < .05 \)).

Figure 2 shows the condition effects on subjects' performance on the rotation test when the responses were scored in terms of illusion.

Although the interaction effect in the ANOVA was not significant I examined the data for an age by condition effect. On theoretical grounds, one might expect such an effect. For example, the conditions should not be too effective among college students as they should have high levels of performances under all treatments. And the third graders might be too young to benefit from the cueing. Thus the effects of cueing should be most evident among the sixth graders.

Figure 3 presents the interaction effects between grades and conditions. Analysis of the means showed
Figure 2: Condition Effects on Rotation Test
When Responses Were Scored in Terms of Illusion Only
Figure 3: Interaction Effects on Rotation Test When Responses Were Scored in Terms of Illusion Only
that the effects among conditions were significant for the sixth graders as predicted.

B. Rotation Trials Scored in Terms of the Illusion and in Terms of the Positions of the Weights

For the next analysis a combination of the two previously described means of scoring the rotation trials was used, namely scoring according to how the subject perceived the illusion and according to the positions of the unequal weights during the rotation trials. The response of only certain subjects were included, though.

For the purpose of this analyses I excluded the scores of any subjects who erred on any illusion trial by giving a response that was in essence an inaccurate illusion, that is by asserting that the wrong hand should feel the heavier weight. Thus, if the subject should have experienced the heavier weight as occurring in the right hand and actually claimed that the heavier weight was in the left hand, the subject's data were excluded. In the analysis, 71 out of 225 subjects were excluded. However, if the subject erred on the illusion trial by asserting that there was no illusion, that subject's data were included. Also, the data were included from all those subjects who were correct on
all illusion trials. In short, the analyses were performed on responses from those subjects who correctly responded to all illusion trials, or whose error on the illusion trials were such as to claim that they experienced no illusion.

Scores of subjects who responded accurately on the four illusion trials had their responses scored in the standard fashion. Subjects who erred by giving a "same" response on any one illusion trial had their responses to those trials scored differently. Here subjects were given credit as being correct if they showed one of two types of response: a correct response defined in terms of the illusion (i.e., as the subject experienced no illusion or no weight difference, the correct response to the rotation trial should be that there would be no difference experienced after the rotation); or a correct response defined in terms of the position of the unequal weights.

The rationale for this admittedly complex scoring procedure has already been given: Those subjects who said that they experienced no difference on the illusion trial might have actually experienced one and then, on the rotation trial, responded in accordance with the perceived illusion. (Analysis of the illusion trials, to be presented in a later section, provides
some very indirect evidence in support of this rationale).

To illustrate this method of scoring, let us take an example of the responses of a subject who correctly perceived the illusion on the first two trials out of the four. Let us also assume that on the remaining two trials, the subject failed the illusion by claiming that there was no difference when lifting the two equal weights. Finally, for the sake of the example, let us assume that the subject was holding the heavier weight in the left hand on the last two trials.

It is these remaining trials that were subject to the procedure being described. The response to the rotation trial would be defined as correct if the subject claimed either A) that the rotation would make no difference, i.e., the weights would be identical or B) that the weight in the right hand should be the heavier of the two. Response "B" is precisely the response that would have been if the subject correctly responded to the illusion trial and subsequently correctly responded to the rotation trial. However, if the subject claimed that the weight in the right hand would feel heavier the response was scored as incorrect.
Grade Effects

Table 7 shows the results of the analysis of variance on rotation test when responses were scored in terms of illusion and in terms of positions of the weights. Grade effects were highly significant as expected ($F(2,124) = 19.5$, $p<.001$). The analysis also showed that the third graders and the sixth graders were significantly different from college students (mean errors of college students = 0.70, mean errors of the sixth graders = 1.97 and mean errors of the third graders = 2.34). College students did better than the third graders and the sixth graders ($p<0.001$). However, the sixth graders did not significantly understand the concept better than the third graders ($p>.1$).

Figure 4 plots the differences between each grade on their performance on rotation test when it was scored in terms of either the illusion or positions of the weights.

Condition Effects

Similar to the findings when responses were scored in terms of illusion only, the scores of the present analysis showed significant condition effects in Table 7 ($F(4,124) = 4.25$, $p<.01$). Subjects in the Weight-cueing condition were the best performers (mean
<table>
<thead>
<tr>
<th>Sources of Variances</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>3.257</td>
<td>1</td>
<td>3.257</td>
<td>2.09</td>
</tr>
<tr>
<td>Grade</td>
<td>60.763</td>
<td>2</td>
<td>30.382</td>
<td>19.50***</td>
</tr>
<tr>
<td>Cond</td>
<td>26.485</td>
<td>4</td>
<td>6.621</td>
<td>4.25**</td>
</tr>
<tr>
<td>Sex * Grade</td>
<td>0.679</td>
<td>2</td>
<td>0.340</td>
<td>0.22</td>
</tr>
<tr>
<td>Sex * Cond</td>
<td>5.014</td>
<td>4</td>
<td>1.254</td>
<td>6.80</td>
</tr>
<tr>
<td>Grade * Cond</td>
<td>15.575</td>
<td>8</td>
<td>1.947</td>
<td>1.25</td>
</tr>
<tr>
<td>Sex * Grade * Cond</td>
<td>7.356</td>
<td>8</td>
<td>0.920</td>
<td>0.59</td>
</tr>
<tr>
<td>Error</td>
<td>193.209</td>
<td>124</td>
<td>1.558</td>
<td></td>
</tr>
</tbody>
</table>

*** p < .001
** p < .01
Figure 4: Grade Effects on Rotation Test When Responses Were Scored in Terms of Either Illusion or the Positions of the Weights
errors of Weight-cueing condition = 1.04), significantly outscoring subjects in the Non-weight-cueing condition, the Appearance-Reality condition, and the control condition (mean errors of Non-weight-cueing condition = 1.85, p<.05, mean errors of Appearance-Reality condition = 2.25, p<.001, and mean errors of control condition = 1.89, p<.05). However, subjects in the Weight+nonweight-cueing condition were not different from those in the Weight-cueing condition (mean errors in Weight+nonweight-cueing condition = 1.33, p> .3).

Besides the findings that Weight-cueing condition made subjects perform better than those who were in the Non-weight-cueing condition, the Appearance-Reality condition, and the control condition, subjects in the Weight+nonweight-cueing condition also performed better than those who were in the Appearance-Reality condition. The mean number of errors that subjects made in the Weight+nonweight-cueing condition was 1.33 which was significantly less than 2.25, which was the average number of errors subjects made in the Appearance-Reality condition. The p value was significant at 0.01 level. The relationship among these conditions on subjects' performances on the rotation test is shown in Figure 5.
Figure 5: Condition Effects on Rotation Test When Responses Were Scored in Terms of Either Illusion or the Positions of the Weights
As presented earlier, I examine for the effect of an interaction although the interaction effect in the ANOVA was not significant. Figure 6 plots the interaction effects between grades and conditions. Specific analysis of differences between the means showed that the sixth graders under the weight cueing condition outscored other sixth graders and that there was not a similar difference in other age groups.

C. Rotation Trials after Successful Illusion Trials

A third scoring method was also employed. The rationale of this particular scoring method was to ensure that the responses that subjects gave on the rotation questions reflected their true understanding. To avoid any ambiguity in scoring, the rotation trials on which there were prior errors in experiencing the illusion were excluded from the analysis. That is, the only rotation trials analyzed in this section were those on which the subjects had correct scores on the prior illusion tests.

There is a need to have a constant number of trials across subjects, of course. In order to maximize the number of subjects whose data could be used, a decision was made to analyze the rotation response for
Figure 6: Interaction Effects on Rotation Test When Responses Were Scored in Terms of Either Illusion or the positions of the Weights
those subjects who had at least two trials correct on the illusion tests. Subjects who made more than two errors at the period of experiencing the illusion were excluded from the analysis. Only ten subjects were excluded for this reason leaving the total number of subjects at 215.

If the subject was correct on only two illusion trials, only responses to the corresponding rotation tests were analyzed. If the subject was correct on more than two illusion trials, corresponding rotation trials to only the last two such trials were employed in the analysis. This procedure was based upon the fact that there was some improvement across trials. Using the last two rotation trials would provide scores showing any gains due to improvement. (The summary tables of the practice effects in each analysis are shown in Appendix F).

The maximum errors that subject could make in the analysis, whose summary appears in Table 8, were 2. A discussion of the specific findings from the analysis appears below.

Grade Effects

Table 8 indicates that there were significant grade effects when the analysis was used \( F(2,186) = 19.36, \text{P}<.001 \). The results of the preceding two
Table 8: ANOVA Table of Rotation Test When only 2 Trials with Corresponding Successful Illusions were Analyzed in Study Two

<table>
<thead>
<tr>
<th>Sources of Variances</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.001</td>
<td>1</td>
<td>0.001</td>
<td>0.00</td>
</tr>
<tr>
<td>Grade</td>
<td>24.959</td>
<td>2</td>
<td>12.480</td>
<td>19.36**</td>
</tr>
<tr>
<td>Cond</td>
<td>9.607</td>
<td>4</td>
<td>2.402</td>
<td>3.73**</td>
</tr>
<tr>
<td>Sex * Grade</td>
<td>0.127</td>
<td>2</td>
<td>0.064</td>
<td>0.10</td>
</tr>
<tr>
<td>Sex * Cond</td>
<td>2.861</td>
<td>4</td>
<td>0.715</td>
<td>1.11</td>
</tr>
<tr>
<td>Grade * Cond</td>
<td>5.180</td>
<td>8</td>
<td>0.648</td>
<td>1.00</td>
</tr>
<tr>
<td>Sex * Grade * Cond</td>
<td>2.297</td>
<td>8</td>
<td>0.287</td>
<td>0.45</td>
</tr>
<tr>
<td>Error</td>
<td>119.884</td>
<td>186</td>
<td>0.645</td>
<td></td>
</tr>
</tbody>
</table>

*** p < .001
** p < .01
analyses of the rotation trials showed that the three age groups were significantly different from one another (while the data were scored in terms of illusion), or that college students performed significantly better than both the third and the sixth graders (when data were scored in terms of either the illusion or positions of the weights). The results of the present analysis had similar age trends. The mean number of errors that college students had was 0.39 which was significantly lower than 0.99 and 1.26 which were the mean errors of the sixth graders and the third graders, respectively (both of the p value were significant at 0.001 level). The sixth graders also performed significantly better than the third graders (p<.05).

Figure 7 shows the age differences on the rotation test.

Condition Effects

Table 8 also indicates that the condition effects were significant (F(4,186) = 3.73, p<.01). However, the results of pairwise comparisons between conditions were a little different from those found in previous analyses. Although subjects in the Weight-cueing condition performed better than those who were in the
Figure 7: Grade Effects on Rotation Test When Only 2 Successful Illusion Trials Were Scored
Appearance-Reality and the control conditions (p<.01 and p<.05 respectively), they did not perform significantly better than those who in the Nonweight-cueing condition (p>.1, mean errors of Weight-cueing condition = 0.66, mean errors of Nonweight-cueing condition = 0.81, mean errors of Appearance-Reality condition = 1.23 and mean errors of control condition = 1.02). It is also interesting to note that the cueing questions in the Non-weight condition served as a performance facilitator on the rotation tasks in that there was a difference between responses to the Non-weight condition and the Appearance-Reality condition, based upon Flavell's work (p<.05). Figure 8 presents the results of condition effects on the rotation test.

Figure 9 indicates the interaction effects between grade and condition in this analysis. Specific analyses within each grade showed no differences due to condition.

Summary

In summary, on the rotation trials there were differences among age groups in understanding weight adaptation no matter what kinds of scoring methods were used to analyze the data. As expected, college students
Figure 8: Condition Effects on Rotation Test When Only 2 Successful Illusion Trials Were Scored
Figure 9: Interaction Effects Between Grade and Condition When Only 2 Successful Illusion Trials Were Scored
made the fewest errors on the judgments, the sixth graders were the second, and the third graders had the most errors on rotation trials.

With respect to condition effects, all these scoring methods resulted in significant condition main effects on subjects' performance on the rotation test. In all three scoring methods the weight cueing condition yielded higher performance than was observed in the control condition. The first two scoring methods, that in which the total scores were considered in terms of the performance on the illusion trials and that in which credit was given for responses scored either in terms of the illusions or in terms of the positions of the weights that were being held yielded essentially identical findings: In both of these analyses the weight condition, aside from improving performance over the control group, enhanced performance over both the Appearance-Reality and Non-weight groups. Moreover, there was a difference between the Weight+nonweight group in comparison to the Appearance-Reality condition based upon Flavell's work. That is, scores were higher in the Weight plus non-weight condition. (The analysis suggests that this difference was due to an improvement due to the Weight+nonweight condition, rather than to a depression
due to the Flavell condition. That is, performance in
the Weight+nonweight condition was nearly significantly
better than scores in the control group, -- p < .06 --
while there was no hint of a significant
difference between the Appearance-Reality and the
control groups).

In the last scoring method based on responses to
two rotation trials for which there was correct
performance on the corresponding illusion trials, all
of the prior findings also emerged as significant. In
addition there was a difference between the Non-weight
condition and the Appearance-Reality condition, with
higher performance on the Non-weight condition. This
last finding suggests, at first blush, that there might
be some improvement from adaptation cueing questions
that make no reference to weight. However, it is
equally as likely that the Appearance-Reality condition
hindered performance.

(3) Conflict Test

This test was administered at the end of the
experiment after the last weight-illusion trial.
Subjects were asked what would happen if the
experimenter put the bottles with equal weights back on
the scale. Subjects who answered that the scale would
be balanced were scored as correct. Subjects who claimed that the scale would be tilted over were scored incorrect. Since there was only one such test each subject was scored as either correct or incorrect. A chi square test, to test the relation between grades and subjects' responses on the test and between conditions and subjects' responses on the test, was used to analyze the data. The results of grade effects are shown in Table 9.

Table 9 indicates that age had an effect. For the third graders, there were almost equal numbers of subjects who said that the scale would be even (53%) and the scale would be tilted over (47%). However, the sixth graders seemed to understand the concept better than the third graders did (approximately 73% correct and 27% incorrect). College students had almost perfect performance. Only 7% of the subjects said that the scale would be tilted while approximately 93% of the subjects responded correctly.

Table 10 shows the relation between conditions and the subjects' responses on the conflict test.

The results in Table 10 showed that the conditions made no difference on subjects' responses to the conflict test. The trends that did exist were in the expected direction though. Interestingly, the most discrepant condition was the Appearance-Reality
### Table 9: Grade Differences in Responses to Conflict Test in Study Two

<table>
<thead>
<tr>
<th></th>
<th>0 (correct)</th>
<th>1 (incorrect)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grade 3</strong></td>
<td>42</td>
<td>37</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>53.16 %</td>
<td>46.84 %</td>
<td></td>
</tr>
<tr>
<td><strong>Grade 6</strong></td>
<td>58</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>73.42 %</td>
<td>26.58 %</td>
<td></td>
</tr>
<tr>
<td><strong>College</strong></td>
<td>62</td>
<td>5</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>92.54 %</td>
<td>7.46 %</td>
<td></td>
</tr>
</tbody>
</table>

**Chi square (225) = 27.99**

*** p < .001
Table 10: Subjects' Responses to Conflict Test in Each Condition in Study Two

<table>
<thead>
<tr>
<th></th>
<th>0 (correct)</th>
<th>1 (incorrect)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weight</strong></td>
<td>34</td>
<td>12</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>73.91%</td>
<td>26.09%</td>
<td></td>
</tr>
<tr>
<td><strong>Non-weight</strong></td>
<td>35</td>
<td>10</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>77.78%</td>
<td>22.22%</td>
<td></td>
</tr>
<tr>
<td><strong>Weight +</strong></td>
<td>37</td>
<td>9</td>
<td>45</td>
</tr>
<tr>
<td><strong>Non-weight</strong></td>
<td>88.43%</td>
<td>19.57%</td>
<td></td>
</tr>
<tr>
<td><strong>Appearance and Reality</strong></td>
<td>25</td>
<td>19</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>56.82%</td>
<td>43.18%</td>
<td></td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>31</td>
<td>13</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>70.45%</td>
<td>29.55%</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>162</td>
<td>63</td>
<td></td>
</tr>
</tbody>
</table>

Chi square (225) = 7.535 \( p > .05 \)
condition, which seemed to slightly depress performances.

Additional Findings

Some minor findings were also discovered. These additional analyses indicated that A) there was an age trend on experiencing the illusion, B) there were age differences and condition differences in responding to cueing questions, and C) there were practice effects among four weight-illusion trials on the feel-really and rotation tests. In the following sections, results of the illusion trials will be considered first, followed by analyses of the cueing questions and analyses for practice effects.

(1) Experiencing the Illusion

In analyzing responses to the illusions trials subjects were defined as having made an error if they asserted that they experienced no difference between the weights or if the difference they did experience was the opposite of that expected via the adaptation. The analyses of these errors showed that the age differences were significant \( F(2,195) = 3.78, \)
p<.05). Table 11 shows the summary of the analysis of variance. The Tukey test was used to analyze the differences between age groups. Surprisingly, the results indicated that the third graders made significantly fewer errors than college students did (mean errors of experiencing illusion for the third graders = 0.72, mean errors of experiencing illusion for college students = 1.11, p<.01). There was no condition effects and of course none was expected.

Figure 10 shows the number of errors that subjects made on the illusion trials by grades. The maximum number of errors that each subject could make was 4.

To say the least, it was surprising that college students committed more errors than third graders. There is a possible explanation for this effect though. Recall that there were two types of errors, one in which the subject asserted that he or she felt no difference and one in which there was a perceived difference but this perceived difference was opposite to what was expected. The explanation for the "poorer" performance of the college students rests with the possibility that they committed more of the first error relative to the third graders, that is with the possibility that the college students were inclined to perceive no difference when there was one. We might expect this response of the more mature subjects if we
Table 11: ANOVA Table for Number of Errors That Subjects Made on Experiencing the Illusion

<table>
<thead>
<tr>
<th>Sources of Variances</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.652</td>
<td>1</td>
<td>0.652</td>
<td>0.90</td>
</tr>
<tr>
<td>Grade</td>
<td>5.501</td>
<td>2</td>
<td>2.751</td>
<td><strong>3.78</strong></td>
</tr>
<tr>
<td>Cond</td>
<td>2.843</td>
<td>4</td>
<td>0.711</td>
<td>0.98</td>
</tr>
<tr>
<td>Sex * Grade</td>
<td>0.288</td>
<td>2</td>
<td>0.144</td>
<td>0.20</td>
</tr>
<tr>
<td>Sex * Cond</td>
<td>2.452</td>
<td>4</td>
<td>0.613</td>
<td>0.84</td>
</tr>
<tr>
<td>Grade * Cond</td>
<td>5.576</td>
<td>8</td>
<td>0.697</td>
<td>0.96</td>
</tr>
<tr>
<td>Sex * Grade * Cond</td>
<td>5.275</td>
<td>8</td>
<td>0.659</td>
<td>0.91</td>
</tr>
<tr>
<td>Error</td>
<td>141.949</td>
<td>195</td>
<td>0.728</td>
<td></td>
</tr>
</tbody>
</table>

**p < .01
Figure 10: Number of Errors Subjects Made on Experiencing the Illusion
assume that they are dominated by their knowledge that the weights were equal. That is, both groups of subjects initially saw that the weights were equal on the scales, prior to the illusion test. The college students might have been more strongly influenced by this knowledge than the sixth graders. That is, this knowledge might have overridden their experiences of or admission to the illusion. It is interesting, in this connection to note that this type of response is what Flavell terms a realism response. That is, the subject is responding in terms of what he or she knows to be the case rather than what appears to be so, when the task calls for responding to appearance.

In fact, analysis of the number of errors supports the aforementioned explanation. When separately analyzing the frequency of each type of error, the responses of subjects reporting that the bottles are the same, vs. experiencing the wrong effect altogether, the results show that there were practically the same number of errors of the second type among college students and the third graders. That is, when it came to experience the incorrect illusion, there was a total of 24 errors among the third grade vs. 22 among the college students. (This trend is the opposite of what was found in the ANOVA.) However, when analyzing the
"same" responses, we find that college students showed 48 such responses as compared to 23 among the third graders. The effect becomes more dramatic when we consider the percentage of errors. Forty-nine percent of the illusion errors of third graders were of the "same" sort vs. 69% among the college students. The significant difference on the ANOVA then must be due to the effect of the "same" responses. Table 12 presents the percentage rate of the error patterns across grades.

(2) Responses to Cueing Questions

Grade Effects

The number of errors that subjects made on cueing questions in each condition was analyzed. The maximum number of errors that each subject could make in responding to the cueing questions was 5. Results showed that both the grade and condition effects were significant at the 0.001 level (F(2, 195) = 11.26, p < .001; F(4, 195) = 6.22, p < .001). Table 13 shows the results of the analysis of variance for all the variables.

The Tukey test was used to differentiate the differences among age groups. Both the third graders and sixth graders made significantly more errors on
Table 12: The Percentage Rate of the Error Patterns Subjects Made at the Period of Experiencing the Illusion among All the Age Groups in Study Two

<table>
<thead>
<tr>
<th></th>
<th>THE SAME</th>
<th>WRONG ILLUSION</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRADE 3</td>
<td>23</td>
<td>24</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>49 %</td>
<td>51 %</td>
<td></td>
</tr>
<tr>
<td>GRADE 6</td>
<td>41</td>
<td>15</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>73 %</td>
<td>27 %</td>
<td></td>
</tr>
<tr>
<td>COLLEGE</td>
<td>48</td>
<td>22</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>69 %</td>
<td>31 %</td>
<td></td>
</tr>
</tbody>
</table>
Table 13: ANOVA Table for the Errors Subjects Made in Response to the Cueing Questions

<table>
<thead>
<tr>
<th>Sources of Variances</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>1.308</td>
<td>1</td>
<td>1.308</td>
<td>0.83</td>
</tr>
<tr>
<td>Grade</td>
<td>35.594</td>
<td>2</td>
<td>17.797</td>
<td>11.26***</td>
</tr>
<tr>
<td>Cond</td>
<td>39.340</td>
<td>4</td>
<td>9.835</td>
<td>6.22***</td>
</tr>
<tr>
<td>Sex * Grade</td>
<td>3.544</td>
<td>2</td>
<td>1.772</td>
<td>1.12</td>
</tr>
<tr>
<td>Sex * Cond</td>
<td>1.318</td>
<td>4</td>
<td>0.330</td>
<td>0.21</td>
</tr>
<tr>
<td>Grade * Cond</td>
<td>10.160</td>
<td>8</td>
<td>1.270</td>
<td>0.80</td>
</tr>
<tr>
<td>Sex * Grade * Cond</td>
<td>2.603</td>
<td>8</td>
<td>0.325</td>
<td>0.21</td>
</tr>
<tr>
<td>Error</td>
<td>308.173</td>
<td>195</td>
<td>1.580</td>
<td></td>
</tr>
</tbody>
</table>

*** p < .001
cueing questions than college students did (mean errors of the third graders = 1.56, mean errors of the sixth graders = 1.16, and mean errors of college students = 0.54). The difference between the third graders and college students was significant at 0.001 level. However, the difference between the sixth graders and college students was significant at 0.01 level. Figure 11 indicates the errors they made among each age group in responding to the cueing questions.

Condition Effects

There was a difference in the difficulty level of the five different sets of cueing question also. Not surprisingly, questions in the control condition were the least difficult ones among all five conditions (mean errors = 0.46). However, questions in the Appearance-Reality condition were generally the most difficult ones among all conditions (mean errors = 1.72). The mean number of errors on the cueing conditions are graphically presented in Figure 12. Analyses of specific differences between the means showed that the questions in the control condition were significantly less difficult (all < .05 at least) than those in the Weight, the Weight+nonweight and the Appearance-Reality cueing conditions. The Appearance-Reality items were more difficult than all
Figure 11: Number of Errors Subjects Made in Response to Cueing Questions among Age Groups
Figure 12: Number of Errors Subjects Made in Response to Cuing Questions in Each Condition
the items but the Nonweight questions. The mean numbers of errors for each condition were as follows: Weight (mean errors = 1.30); Non-weight (mean errors = 0.82); Weight+nonweight (mean errors =1.14); Appearance-Reality (mean errors = 1.72); and control condition (mean errors = 0.46).

(3) Practice Effects

As it was pointed out before, each subject received four weight-adaptation trials. Hence it is worth examining whether subjects improved their weight-illusion judgments over the trials. The analyses revealed that subjects improved on both the feel-really measure as well as on the rotation trials. Summaries for these analyses appear in the Appendix F.

Overall Summary

Generally speaking, the major findings were as was expected. Subjects of different ages understood the weight-adaptation concept differently. College students seemed to understand the concept best. Subjects within different cueing conditions understood the concept differently. Those cueing questions which related to the concept of weight, including those in the
Weight-cueing and Weight+nonweight-cueing conditions, seemed to be the best performance facilitators in assisting subjects to understand the weight-illusion tasks. There did exist practice effects over four weight-illusion trials. Subjects performed better on weight-illusion tasks on the last two trials than they did on the first two trials.

Subjects in different age groups experienced the illusion differently on weight-illusion trials. College students made more errors than the third graders and the sixth graders.

Subjects also showed different levels of performances in responding to the cueing questions. The third graders made the most errors on cueing questions among all the age groups. The cueing questions in different conditions had different degrees of difficulty also. Questions used in the Appearance-Reality condition seemed to be the most difficult ones. The questions in the Weight condition were rated the second in order of difficulty. However, questions used in the control condition were the easiest ones among all the conditions.
CHAPTER FIVE

GENERAL DISCUSSION

Two studies were conducted to determine subjects' understanding of a weight illusion as a function of age and experimental cueing conditions. Of particular concern to this study was how subjects experienced an illusion wherein one of two equal weights was made to appear heavier than the other. Would subjects treat them as really unequal or would they realize that the weights were the same after all? Answers to these questions were determined mainly by analyzing subjects' responses on what were termed rotation trials. On these trials the experimenter asked the subjects what would happen if the subject reversed the weights being held in each hand, so that the hand previously holding what appeared to be the lighter weight would be holding the heavier appearing one. If the subject believed that the experience of a weight difference was really due to a difference in weight, then the subject should claim that reversing the position of the weights should alter the experience of which hand is holding the heavier weight. If subjects believed that the experience of the weight differences was due to an adaptation-like effect, that is a product of the sensory apparatus,
then the subject should claim that reversing the weights should make no difference.

In the first study there were some effects due to age, but analysis of the results suggested that even many college students were having difficulty with the task. Moreover, there were no condition effects. The age trends were particularly discrepant with earlier findings that suggested that by the sixth grade, and certainly, college students, should have knowledge of perceptual adaptation and this knowledge should have facilitated their performances.

For this reason, Study Two was devised. Study Two was identical to the first study, except for the addition of one step in the procedure. Prior to the rotation trials, the subject was asked to lift up the unequal weights. This step was added because it was thought that lifting up these weights would constitute a reminder to the subjects of the possibility that holding the two unequal weights was the cause of the illusion.

The results of Study Two stood in stark contrast to those of Study One. In this study there were marked age differences. Moreover, the condition effects proved significant. Overall, the condition in which subjects were asked questions about prior weight-adaptation-like effects enhanced performance.
There are many questions that arise concerning these results. One concerns the meaning of the failure of the subjects of Study One and the reason for this failure. With respect to the meaning of the failure it is interesting to speculate that in a very real sense, the failure on the rotation trials signaled an absence of conservation of weight. That is, subjects knew that the bottles weighed the same. They saw this demonstrated to them on the scales. Yet they apparently ignored this information upon experiencing the illusion. Here they asserted that the difference in the weights was real. A possible reason for this apparent lack of conservation is that the subjects are not associating the information on the equality phase of the trials with what appears later. This finding shows how much conservation can be said to depend upon task demands. If the demands are such as to place a strain on the ability of the subject to connect different phases of the trials, failure will result even among adults.

Another question concerns the meaning of the pattern of effects of the cueing trials of Study Two. There were several points worthy of discussion here. For one, as was expected, the weight cueing condition proved the most effective. This was not surprising in that the weight cueing questions bore the strongest
relation to the subsequent test items. The condition involving weight plus non-weight adaptation questions showed the enhancing effects also. However, results showed that subjects could not gain from practice on general knowledge of adaptation.

Why did the cueing effects occur? One answer to this question is that the experiences activated a schema or type of knowledge that the subjects could later use on the test phase of the trial. This schema could be used in one of two ways. First, the schema might represent knowledge of a more general principle, of a construct, that the subject might have and this construct might simply generalize to the test item. For example, when the subjects are asked about weight adaptation, they might be reminded of the general principle of weight adaptation and weight contrast effects. The later rotation trials might then be seen as a case or example of this more general principle. On the other hand, it might also be assumed that activating memory of several experiences might provide a base wherein subjects could infer the more general principle of weight adaptation. This latter interpretation assumes that at the time of success the subjects undergo some sort of cognitive restructuring. In other words, they only come to understand the principle of adaptation as a result of the combination
of the test and cueing trials. There is no way of deciding which of these explanations is most satisfactory.

These explanations can be seen as very similar to Holyoak's theory of analogical thinking (Holyoak, 1983) which is proposed to be a type of human intelligence (Sternberg, 1983). Holyoak (1983) hypothesized that subjects can "converge" a schema out of two analogies to solve a type of problem. The definition of a schema in his theory is "a representation or the abstracted core meaning of the type of problem for which convergence solution is possible". Mapping between two analogies is the main process of forming the schema in analogical thinking. He assumed that subjects will inspect two analogous problems, element by element, to find the corresponding similarities and differences when they are asked to do the analogical problems. Sometimes, one to one correspondence is impossible because of the wording of each analogy. Therefore, a higher level thinking process is necessary. In this process the subject abstracts out the core meaning of the analogies to form a high level cognitive structure. In this case, a schema is formed which stands for the analogy.

However, the mapping process between analogies is not always successful. Failure occurs if the
corresponding similarities are not sufficient to form a schema or subjects are unable to execute the thinking process of a higher level. Moreover, the subject has to have an awareness that one problem can be used to solve another. In other words, the subject must think at the outset that the solution or principle contained in one problem can be used in another.

In his theory there are two subtypes of unsuccessful mapping. The first one is named "structure-violating differences" and occurs when no consistent core meaning between analogies can be found underlying corresponding similarities in two analogies. However, "indeterminant correspondences" between two analogs is defined as "failure to map conclusively some elements between the analogies" (Holyoak, 1983). In other words, at the beginning of the process of mapping, subjects found that the structures of the problems were different. Therefore, the mapping process was interrupted. No schema was eventually formed.

In his theory, Holyoak assumed that everyone has the ability to form a schema if the structures of the problems can provide sufficient information. However, the cognitive extracting process should be impossible for children of some ages.

Mapping may explain the performance on the rotation test in the present study. To answer the
question why and how it happens that the cueing questions in the Weight and Weight+nonweight conditions are the best performance facilitators, the
successfulness of mapping is important. When subjects receive the cueing questions related to weights, the mapping process succeeds because of the high similarities between cueing questions and rotation tasks. The corresponding one to one matching enables subjects to form the schema. In present study, a schema of weight adaptation could have formed in both the Weight and Weight+nonweight conditions. However, as compared to those questions in Weight condition, questions in Nonweight conditions are much more difficult to map.

As was mentioned in discussing Holyoak's notion of awareness, the possibility that one should apply the principle of one problem to that of another is also important. In his work, if the subject saw or mapped the similarities, then they should have had the awareness. But the awareness could also be provided by instructing the subject to try to connect one problem to the next. In the present study, the differential effectiveness of the cueing conditions might have been due to difference in awareness on the part of the subject, that he or she should even try to connect one
experience with the other, in particular on the Non-weight and the Appearance-Reality conditions.

In this light, it should be noted, that with respect to the Appearance-Reality condition based upon Flavell's work, there was not even a clear correspondence between the cueing and test items. That is the concept of the appearance-reality distinction underlying the questions in Flavell-cueing condition was somewhat different from that of perceptual adaptation. There would be what Holyoak termed "indeterminant correspondences".

There were several other questions with the results. For example, even in Study Two, the performance in the rotation trials was not similar to what Arnold et al. (1986) reported on the temperature tasks. The weight tasks of this study were more difficult. One reason for this difference probably involves the nature of the tests. In the study by Arnold et al. (1986), subjects experienced the illusion of a temperature difference using water that was in one container. The fact that the water was in a single container might have cued children to believe that there were not two temperatures.

It was also surprising that the Appearance-Reality condition failed to improve performance; indeed, there were even consistent trends suggesting that it might
have indeed hindered performance! There is no reason why this should have occurred unless subjects were confused by the cueing. Moreover, the relatively low success rate on the cueing questions themselves deserves some comment. Flavell's results would have us believe that success on this type of question should be experienced by preschoolers (Flavell, 1986; Flavell et al., 1983; Flavell et al., 1986). Yet many of the subjects in this study, including college students, were having difficulty on these cueing items. There were some inherent ambiguities on some of these items, to be sure. For example, when covering the card, the question read, "Is it really really white or is it really really red?". The term "it" is ambiguous and might refer to the card, but obviously many subjects thought that it did not. Despite this and other possible ambiguities, the difficulties on these items were surprising.

The age trend on the illusion trials in Study Two was also surprising, with college students showing more errors than the third graders. This was earlier explained as due to the college student's reliance on prior knowledge. It should be noted, though, that this trend was not replicated in the results of Study One. There is no apparent reason for the discrepancies in the results of these studies.
In conclusion, there is much opportunity for additional work. There is a question as to how other types of feedback might work. There are also questions as to how modifications in the testing procedure might affect performance such as having the subjects determine the equality of weight by having them experience the equality through lifting the two weights.
REFERENCES


Appendix A: Questionnaires Used in

Weight-cueing Condition
(REMEMBER) Lots of times we pick something up and put it down. Then we pick something else up. Sometimes the things we pick up feel heavier than they do at other times. Tell me, when does a ball feel heavier

A. if you lift a feather up and put it down before you pick up the ball or if you lift up a brick first?

B. if you lift a brick up and put it down before you pick up the ball or if you lift up a feather first?

score ________
(REMEMBER) Lots of times we pick something up and put it down. Then we pick something else up. Sometimes the things we pick up feel heavier than they do at other times. Tell me, when does a carton of milk feel heavier ......

A. if you lift a bag of popcorn up and put it down before you pick up a carton of milk or if you lift up a bag of potatoes first?

B. if you lift a bag of potatoes up and put it down before you pick up a carton of milk or if you lift up a bag of popcorn first?

score ______
(REMEMBER) Lots of times we pick something up and put it down. Then we pick something else up. Sometimes the things we pick up feel heavier than they do at other times. Tell me, when does a bag of grocery feel heavier .......

A. if you lift a bag of straw up and put it down before you pick up a bag of grocery or if you lift up a bag of sand first?

B. if you lift a bag of sand up and put it down before you pick up a bag of grocery or if you lift up a bag of straw first?

score ________
(REMEMBER) Lots of times we pick something up and put it down. Then we pick something else up. Sometimes the things we pick up feel heavier than they do at other times. Tell me, when does a book feel heavier

A. if you lift a piece of cotton up and put it down before you pick up a book or if you lift up a stone first?

B. if you lift a stone up and put it down before you pick up a book or if you lift up a piece of cotton first?

score _____
(REMEMBER) Lots of times we pick something up and put it down. Then we pick something else up. Sometimes the things we pick up feel heavier than they do at other times. Tell me, when does a doll feel heavier

A. if you lift a big book up and put it down before you pick up a doll or if you lift a piece of paper first?

B. if you lift a piece of paper up and put it down before you pick up a doll or if you lift a big book up first?

score ________
Appendix B: Questionnaires Used in
Non-weight-cueing Condition
When we go out of the house we see the sun shine and it is bright. Sometimes it seems like it is brighter than it is at other times. Let's say you go outside at lunchtime. When does it seem like its brighter ..........

A. if you come out of your house or out of a movie?

B. if you come out of a movie or out of your house?

score __________
Lots of time we take a bath. Sometimes the bath water feels warmer than it does at other times. When does the bath feel warmer ......

A. if you take a bath after playing outside in the snow or after playing inside in the kitchen?

B. if you take a bath after playing inside in the kitchen or after playing outside in the snow?

score ________
Sometimes things taste better than they do at other times. Tell me, when you eat a candy bar which tastes better ........

A. the first bite or the second bite?

B. the second bite or the first bite?

score ______
Many times we smell cookies baking in the kitchen. Sometimes we can smell them better than we can at other times. When do you smell the cookies better .......

A. if you have just come into the kitchen from playing outside or if you have been playing in the kitchen for a long time?

B. if you have been playing in the kitchen for a long time or if you have just come into the kitchen from playing outside?

score ________
Lots of time we hear noise outside the house. Sometimes the noise seems louder than it does at the other times. Tell me when does the sound outside your house seem louder .......

A. after you come home from a football game or after you take a walk in the woods?

B. after you take a walk in the woods or after you come home from a football game?

score _________
Appendix C: Questionnaires Used in

Weight+nonweight-cueing Condition
When we go out of the house we see the sun shine and it is bright. Sometimes it seems like it is brighter than it is at other times. Let's say you go outside at lunchtime. When does it seem like its brighter .......

A. if you come out of your house or out of a movie?

B. if you come out of a movie or out of your house?

score
Lots of times we pick something up and put it down. Then we pick something else up. Sometimes the things we pick up feel heavier than they do at other times. Tell me, when does a ball feel heavier ......

A. if you lift a feather up and put it down before you pick up the ball or if you lift up a brick first?

B. if you lift a brick up and put it down before you pick up the ball or if you lift up a feather first?
Lots of time we take a bath. Sometimes the bath water feels warmer than it does at other times. When does the bath feel warmer .......

A. if you take a bath after playing outside in the snow or after playing inside in the kitchen?

B. if you take a bath after playing inside in the kitchen or after playing outside in the snow?

score _____
Lots of times we pick something up and put it down. Then we pick something else up. Sometimes the things we pick up feel heavier than they do at other times. Tell me, when does a carton of milk feel heavier ......

A. if you lift a bag of popcorn up and put it down before you pick up a carton of milk or if you lift up a bag of potatoes first?

B. if you lift a bag of potatoes up and put it down before you pick up a carton of milk or if you lift up a bag of popcorn first?

score _____
Sometimes things taste better than they do at other times. Tell me, when you eat a candy bar which tastes better ........

A. the first bite or the second bite?

B. the second bite or the first bite?

score _______
Appendix D: Questionnaires Used in

Appearance-Reality Condition
1. A white index card that looked red when placed behind a piece of plastic.

   Look at this card. Now watch what I am doing
   ................................

   (A) Is it really really white or is it really really red?

   (B) Is it really really red or is it really really white?

   really really red _____

   really really white  ________
2. A plastic candy bar that looks like a real Hershey candy bar.

I want you to touch this.........

(A) Is it really really a candy bar or is it really really plastic?

(B) Is it really really plastic or is it really really a candy bar?

really really plastic ______

really really candy bar _________
3. A paper clip that looked through a magnifying lens.

Look at the paper clip. Now watch what I am doing..................

(A) Is it really really big or is it really really small?

(B) Is it really really small or is it really really big?

really really small
really really big
4. A chopstick in a glass of water.

I want you to look at this. Now watch what I am doing .............

(A) Is it really really bent or is it really really straight?

(B) Is it really really straight or is it really really bent?

really really bent

really really straight
5. An ice cream cone that made of wax.

I want you to touch this......

(A) Is it really really an ice cream cone or is it really really wax?

(B) Is it really really wax or is it really really an ice cream cone?

really really wax __________

really really ice cream __________
Appendix E: Questionnaires Used in Control Condition
Who is the president of the United States of America?

How many days are there in a year?

Where does the president live?

When do the leaves fall off the trees?

What color is a Robin?
Table 14: ANOVA Table for Rotation Test in Study One When Trials as Another Dependent Variable

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>SS</th>
<th>df</th>
<th>Mf</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.122</td>
<td>1</td>
<td>0.12</td>
<td>0.12 **</td>
</tr>
<tr>
<td>Grade</td>
<td>13.459</td>
<td>2</td>
<td>6.730</td>
<td>6.57 *</td>
</tr>
<tr>
<td>Cond</td>
<td>5.244</td>
<td>4</td>
<td>1.311</td>
<td>1.28</td>
</tr>
<tr>
<td>Sex * Grade</td>
<td>1.988</td>
<td>2</td>
<td>0.994</td>
<td>0.97</td>
</tr>
<tr>
<td>Sex * Cond</td>
<td>2.640</td>
<td>4</td>
<td>0.660</td>
<td>0.64</td>
</tr>
<tr>
<td>Grade * Cond</td>
<td>3.202</td>
<td>8</td>
<td>0.400</td>
<td>0.39</td>
</tr>
<tr>
<td>Sex * Grade * Cond</td>
<td>7.735</td>
<td>8</td>
<td>0.967</td>
<td>0.94</td>
</tr>
<tr>
<td>Error</td>
<td>112.714</td>
<td>110</td>
<td>1.025</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>SS</th>
<th>df</th>
<th>Mf</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trials</td>
<td>1.272</td>
<td>1</td>
<td>1.272</td>
<td>4.76 *</td>
</tr>
<tr>
<td>Trials * Sex</td>
<td>0.155</td>
<td>1</td>
<td>0.155</td>
<td>0.58</td>
</tr>
<tr>
<td>Trials * Grade</td>
<td>0.103</td>
<td>2</td>
<td>0.051</td>
<td>0.19</td>
</tr>
<tr>
<td>Trials * Cond</td>
<td>1.192</td>
<td>4</td>
<td>0.298</td>
<td>1.11</td>
</tr>
<tr>
<td>Trials * Sex * Grade</td>
<td>1.124</td>
<td>2</td>
<td>0.562</td>
<td>2.10</td>
</tr>
<tr>
<td>Trials * Sex * Cond</td>
<td>0.794</td>
<td>4</td>
<td>0.198</td>
<td>0.74</td>
</tr>
<tr>
<td>Trials * Grade * Cond</td>
<td>0.889</td>
<td>8</td>
<td>0.111</td>
<td>0.42</td>
</tr>
<tr>
<td>T * S * G * C</td>
<td>1.923</td>
<td>8</td>
<td>0.240</td>
<td>0.90</td>
</tr>
<tr>
<td>Errors</td>
<td>29.414</td>
<td>110</td>
<td>0.267</td>
<td></td>
</tr>
</tbody>
</table>

** p < .01
* p < .05
Table 15: ANOVA Table for The Results of Feel-Really Test When Trials as One of the Independent Variables in Study Two

<table>
<thead>
<tr>
<th>Sources of Variances</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.013</td>
<td>1</td>
<td>0.013</td>
<td>0.02</td>
</tr>
<tr>
<td>Grade</td>
<td>6.429</td>
<td>2</td>
<td>3.215</td>
<td>5.22</td>
</tr>
<tr>
<td>Cond</td>
<td>4.284</td>
<td>4</td>
<td>1.071</td>
<td>1.74</td>
</tr>
<tr>
<td>Sex * Grade</td>
<td>0.394</td>
<td>2</td>
<td>1.197</td>
<td>0.32</td>
</tr>
<tr>
<td>Sex * Cond</td>
<td>2.696</td>
<td>4</td>
<td>0.674</td>
<td>1.09</td>
</tr>
<tr>
<td>Grade * Cond</td>
<td>5.919</td>
<td>8</td>
<td>0.740</td>
<td>1.20</td>
</tr>
<tr>
<td>Sex * Grade * Cond</td>
<td>2.796</td>
<td>8</td>
<td>0.349</td>
<td>0.57</td>
</tr>
<tr>
<td>Error</td>
<td>120.083</td>
<td>195</td>
<td>0.616</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sources of Variances</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial</td>
<td>4.777</td>
<td>1</td>
<td>4.777</td>
<td>17.63</td>
</tr>
<tr>
<td>Trial * Sex</td>
<td>0.363</td>
<td>1</td>
<td>0.363</td>
<td>1.34</td>
</tr>
<tr>
<td>Trial * Grade</td>
<td>0.814</td>
<td>2</td>
<td>0.407</td>
<td>1.50</td>
</tr>
<tr>
<td>Trial * Cond</td>
<td>0.624</td>
<td>4</td>
<td>0.156</td>
<td>0.58</td>
</tr>
<tr>
<td>Trial * Sex * Grade</td>
<td>0.708</td>
<td>2</td>
<td>0.354</td>
<td>1.31</td>
</tr>
<tr>
<td>Trial * Sex * Cond</td>
<td>2.334</td>
<td>4</td>
<td>0.584</td>
<td>2.15</td>
</tr>
<tr>
<td>Trial * Grade * Cond</td>
<td>1.012</td>
<td>8</td>
<td>0.127</td>
<td>0.47</td>
</tr>
<tr>
<td>Trial * S * G * C</td>
<td>2.502</td>
<td>8</td>
<td>0.313</td>
<td>1.15</td>
</tr>
<tr>
<td>Error</td>
<td>52.846</td>
<td>195</td>
<td>0.271</td>
<td></td>
</tr>
</tbody>
</table>

*** p < .001
**  p < .01
Appendix F: ANOVA Summary Tables of Practice

Effects in Study One and Study Two
Table 16: ANOVA Table for the Results of Rotation Test When Responses Scored According to the Illusion They Experienced and Trials as One of the Independent Variables in Study Two

<table>
<thead>
<tr>
<th>Sources of Variances</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.097</td>
<td>1</td>
<td>0.097</td>
<td>0.12</td>
</tr>
<tr>
<td>Grade</td>
<td>25.041</td>
<td>2</td>
<td>12.520</td>
<td>15.24</td>
</tr>
<tr>
<td>Cond</td>
<td>12.490</td>
<td>4</td>
<td>3.123</td>
<td>3.80</td>
</tr>
<tr>
<td>Sex * Grade</td>
<td>0.875</td>
<td>2</td>
<td>0.438</td>
<td>0.53</td>
</tr>
<tr>
<td>Sex * Cond</td>
<td>1.369</td>
<td>4</td>
<td>0.342</td>
<td>0.42</td>
</tr>
<tr>
<td>Grade * Cond</td>
<td>9.790</td>
<td>8</td>
<td>1.223</td>
<td>1.49</td>
</tr>
<tr>
<td>Sex * Grade * Cond</td>
<td>2.634</td>
<td>8</td>
<td>0.329</td>
<td>0.40</td>
</tr>
<tr>
<td>Error</td>
<td>160.233</td>
<td>195</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sources of Variances</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial</td>
<td>5.703</td>
<td>1</td>
<td>5.703</td>
<td>14.78</td>
</tr>
<tr>
<td>Trial * Sex</td>
<td>0.350</td>
<td>1</td>
<td>0.350</td>
<td>0.91</td>
</tr>
<tr>
<td>Trial * Grade</td>
<td>0.114</td>
<td>2</td>
<td>0.057</td>
<td>0.15</td>
</tr>
<tr>
<td>Trial * Cond</td>
<td>0.888</td>
<td>4</td>
<td>0.222</td>
<td>0.58</td>
</tr>
<tr>
<td>Trial * Sex * Grade</td>
<td>2.094</td>
<td>2</td>
<td>1.047</td>
<td>2.71</td>
</tr>
<tr>
<td>Trial * Sex * Cond</td>
<td>0.820</td>
<td>4</td>
<td>0.205</td>
<td>0.53</td>
</tr>
<tr>
<td>Trial * Grade * Cond</td>
<td>1.345</td>
<td>8</td>
<td>0.168</td>
<td>0.44</td>
</tr>
<tr>
<td>Trial * S * G * C</td>
<td>1.746</td>
<td>8</td>
<td>0.218</td>
<td>0.57</td>
</tr>
<tr>
<td>Error</td>
<td>75.221</td>
<td>195</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*** p < .001
**  p < .01
Table 17: ANOVA Table for the Results of Rotation Test When Responses Were Scored either According to the Illusion or According to Positions of the Weights and Trials Were Used as One of the Independent Variables in Study Two

<table>
<thead>
<tr>
<th>Sources of Variances</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>1.434</td>
<td>1</td>
<td>1.434</td>
<td>1.90</td>
<td>***</td>
</tr>
<tr>
<td>Grade</td>
<td>30.178</td>
<td>2</td>
<td>15.089</td>
<td>20.03</td>
<td>**</td>
</tr>
<tr>
<td>Cond</td>
<td>11.956</td>
<td>4</td>
<td>2.989</td>
<td>3.97</td>
<td></td>
</tr>
<tr>
<td>Sex * Grade</td>
<td>0.821</td>
<td>2</td>
<td>0.411</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>Sex * Cond</td>
<td>1.181</td>
<td>4</td>
<td>0.295</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>Grade * Cond</td>
<td>7.722</td>
<td>8</td>
<td>0.965</td>
<td>1.28</td>
<td></td>
</tr>
<tr>
<td>Sex * Grade * Cond</td>
<td>3.710</td>
<td>8</td>
<td>0.464</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>106.965</td>
<td>195</td>
<td>0.753</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Trial                | 11.052| 1  | 11.052| 31.11 | ***|
| Trial * Sex          | 0.085| 1  | 0.085| 0.24  |   |
| Trial * Grade        | 0.899| 2  | 0.449| 1.27  |   |
| Trial * Cond         | 0.527| 4  | 0.132| 0.37  |   |
| Trial * Sex * Grade  | 0.594| 2  | 0.297| 0.84  |   |
| Trial * Sex * Cond   | 0.563| 4  | 0.141| 0.40  |   |
| Trial * Grade * Cond | 2.570| 8  | 0.321| 0.90  |   |
| Trial * S * G * C    | 1.775| 8  | 0.222| 0.62  |   |
| Error                | 50.446| 195 | 0.355|       |   |

*** p < .001
** p < .01