INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

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

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
The role of computer-based instruction in the development of strategies for computational estimation with middle school children

Whiteman, Frederick Cookson, Ph.D.

The Ohio State University, 1988

Copyright 1988 by Whiteman, Frederick Cookson. All rights reserved.
THE ROLE OF COMPUTER-BASED INSTRUCTION IN THE
DEVELOPMENT OF STRATEGIES FOR COMPUTATIONAL ESTIMATION
WITH MIDDLE SCHOOL CHILDREN

DISSERTATION

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

by
Frederick C. Whiteman, B.S., M. A.

*****
The Ohio State University
1988

Dissertation Committee:
Suzanne K. Damarin
Lorren L. Stull
William D. Taylor

Approved by:
Suzanne K. Damarin
Advisor
Department of Educational Policy and Leadership
Copyright by
Frederick C. Whiteman
1988
To My Loving Wife, Rebekah
ACKNOWLEDGEMENTS

In the years since I first began my graduate "career," I have benefitted greatly from the many people that have touched my life—professors, fellow graduate students, university staff, friends and family members. In some way this acknowledgement provides an opportunity to recognize those individuals who have been influential.

My advisor, Dr. Suzanne Damarin; respected mentor, affectionate advocate, and cherished friend; has constantly involved me in projects that have extended me to the limits of my ability. No words flowing forth from this pen will ever prove an adequate appreciation for what Professor Damarin has done for me.

In the darkest hours of frustration, personal difficulty, and self–doubt, Dr. Lorren Stull always found a way to comfort, encourage, cajole, and amuse in such a way as to put me back on the path to enlightenment. Graduation, although a happy event, is somewhat diminished as it will almost certainly involve a loss of intimacy with this truly great man.

Although his name does not appear on the title page of this document, Professor Richard Howell was a valued member of my dissertation committee. His contribution to my development has been both intellectual and spiritual. His awareness of, and sensitivity for the graduate student "condition" is unparalleled.

My conversations with Dr. William Taylor almost always led to a greater insight of the intricacies of the educational process, particularly uniquely human concerns that grow in significance as our culture focuses increasingly upon
technological innovation. I am a much wiser man as a result of his constant probing.

Although he was not a member of my dissertation committee, Dr. John Belland was a constant source of support and inspiration and his influence has been profound throughout my candidacy.

In addition to my esteemed advisors, numerous graduate student colleagues have provided much needed support. My dear friends Vince Lasnik, Jane Johnsen, Marty Grant, Jim Clarke, Larry Hoffman, and Jim Schwartz have all been instrumental in helping me maintain a proper intellectual and psychological perspective.

I would also like to extend my appreciation to other cherished friends: Don Renner, Rick and debi Keller, Liz Stull, Paul Hinrichs, and Jeff Gourlay for continued encouragement and unwavering support.

My cousin, Dr. James Drake, deserves special mention as his sensitivity, love, and kindness have provided me with the courage, self-confidence, and inspiration to broaden myself. His support during the most difficult of life's experiences may never be returned in kind, however, it has not gone unappreciated—Thanks, Jim, for being there when I needed you the most. My mother, brother, and aunt have been equally influential and persevering in their support of my academic pursuits.

Rebekah, my long suffering wife and love of my life, has made this endeavor bearable through her love, compassion, patience, and willingness to sacrifice a large part of our time together these past few years. I trust that the future will provide ample opportunity to help her fulfill her dreams.

iv
VITA

August 24, 1950

B. S. in Elementary Education,
The Ohio State University,
Columbus, Ohio.

1979 - 1980

Mathematics and Science Teacher
Welsh Hills School,
Granville, Ohio.

1984

M. A. Elementary Education
The Ohio State University,
Columbus, Ohio.

1981 - 1987

Graduate Research Associate,
The Ohio State University,
Columbus, Ohio.

1987 - 1988

Senior Researcher - Project Director,
Guild Associates,
Columbus, Ohio.

FIELDS OF STUDY

Major Field: Instructional Design and Educational Technology

Studies in Instructional Design: Professors Suzanne Damarin, John Belland,
William Taylor, and Richard Howell.

Minor Filed: Early and Middle Childhood Mathematics:

Studies in Early and Middle Childhood Mathematics: Professors Suzanne Damarin,
Lorren Stull, John Reidl, and Richard Shumway.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ................................................................. iii

VITA ........................................................................................................... v

LIST OF TABLES .................................................................................. x

LIST OF FIGURES .................................................................................. xvi

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Qualitative Estimation</td>
<td>2</td>
</tr>
<tr>
<td>Quantitative Estimation</td>
<td>3</td>
</tr>
<tr>
<td>Computational Estimation</td>
<td>5</td>
</tr>
<tr>
<td>Need</td>
<td>5</td>
</tr>
<tr>
<td>Computational Estimation Processes</td>
<td>7</td>
</tr>
<tr>
<td>Reasonableness</td>
<td>8</td>
</tr>
<tr>
<td>Manageability</td>
<td>10</td>
</tr>
<tr>
<td>Appropriateness</td>
<td>12</td>
</tr>
<tr>
<td>Compensation</td>
<td>14</td>
</tr>
<tr>
<td>Estimation and Problem Solving</td>
<td>15</td>
</tr>
<tr>
<td>Computers in Instruction</td>
<td>19</td>
</tr>
<tr>
<td>Research Questions</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II</th>
<th>LITERATURE REVIEW</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Introduction</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Assessment Issues</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Assessment Difficulties</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Timing</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Question Format</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Question Context</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Numbers</td>
<td>28</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Estimation Processes and Related Abilities</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Research in Teaching Computational Estimation</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>Instructional Issues</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Teaching Computational Estimation</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Sequencing</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Timing</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Cognitive Goals</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td>Affective Goals</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Computer-based Instruction in Computational Estimation</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Interactivity</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Graphics</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>Practice and Problem Variation</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Computer-based Instruction and the Goals of Estimation</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Summary of Literature Review</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>III METHODS AND PROCEDURES</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Procedures</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Sampling</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Instructional Setting</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Treatments</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>Instruments</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Adjusting Estimates-Multiplication Test</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>Data Coding and Scoring</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>Data Coding</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>Response Categories</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>Scoring</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Experimental Design</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>Internal Validity</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>External Validity</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>IV FINDINGS</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>Major Hypotheses</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>Minor Hypotheses</td>
<td>78</td>
<td></td>
</tr>
</tbody>
</table>

vii
### LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A Comparison of Computational Estimation Processes and Strategies Defined by Different Researchers</td>
<td>31</td>
</tr>
<tr>
<td>2. The Instructional Sequence for the Treatment Groups</td>
<td>60</td>
</tr>
<tr>
<td>3. ACE Posttest Items by Subscale</td>
<td>65</td>
</tr>
<tr>
<td>4. AET Posttest Items by Operation Typ</td>
<td>66</td>
</tr>
<tr>
<td>5. Categories of Response ACE Estimation Items</td>
<td>68</td>
</tr>
<tr>
<td>6. Categories of Response for AE-M Multiplication Subscale</td>
<td>70</td>
</tr>
<tr>
<td>7. Aspects of Scoring for the AE-M and ACE Tests</td>
<td>71</td>
</tr>
<tr>
<td>8. A Summary of the Results of Two-Factor ANOVAS for All Pretest Measures</td>
<td>84</td>
</tr>
<tr>
<td>10. Two-Factor Analysis of Variance for the AE-M Pretest Adjusted Estimates on the Multiplication Subscale by Group and Gender</td>
<td>87</td>
</tr>
<tr>
<td>11. Means and Standard Deviations for the AE-M Pretest Items With No Response</td>
<td>90</td>
</tr>
<tr>
<td>12. Two-Factor Analysis of Variance for the AE-M Pretest Total Number of Items with No Response by Group and Gender</td>
<td>90</td>
</tr>
<tr>
<td>TABLE</td>
<td>PAGE</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>13. A Summary of the Results of Two-Factor ANOVAS for All Posttest Measures</td>
<td>93</td>
</tr>
<tr>
<td>14. Means and Standard Deviations for the AE–M Posttest Total Score</td>
<td>95</td>
</tr>
<tr>
<td>15. Means and Standard Deviations for the AE–M Posttest Multiplication Subscale</td>
<td>95</td>
</tr>
<tr>
<td>17. Two-Factor Analysis of Variance for the AE–M Posttest Total Score and Multiplication Subscale by Group and Gender</td>
<td>97</td>
</tr>
<tr>
<td>18. Two-Factor Analysis of Variance for the AE–M Posttest Predicted Initial Estimate on the Multiplication Subscale by Group and Gender</td>
<td>97</td>
</tr>
<tr>
<td>19. Means and Standard Deviations for the AE–M Posttest Numerical Subsection</td>
<td>100</td>
</tr>
<tr>
<td>20. Two-Factor Analysis of Variance for the AE–M Posttest Numerical Subsection Scores by Group and Gender</td>
<td>101</td>
</tr>
<tr>
<td>21. A Summary of Paired T-Tests for the AE–M Instrument by Treatment Group</td>
<td>103</td>
</tr>
<tr>
<td>22. Means and Standard Deviations for Good, Fair, and Poor Estimators on Selected AE–M Measures</td>
<td>104</td>
</tr>
<tr>
<td>23. A Summary of Paired T-Tests for Good, Fair, and Poor Estimators in the Two Experimental Groups on Selected Measures of the AE–M Instrument</td>
<td>105</td>
</tr>
<tr>
<td>24. Means and Percentage of Use of Adjusted Estimates on the AE–M Pretest Multiplication Subscale</td>
<td>111</td>
</tr>
<tr>
<td>TABLE</td>
<td>PAGE</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>25. Means and Standard Deviations for the ACE Pretest Total Score</td>
<td>188</td>
</tr>
<tr>
<td>26. Means and Standard Deviations for the ACE Pretest Multiplication Subscale</td>
<td>188</td>
</tr>
<tr>
<td>27. Two-Factor Analysis of Variance for the ACE Pretest Total Score and Multiplication Subscale Score by Group and Gender</td>
<td>189</td>
</tr>
<tr>
<td>28. Means and Standard Deviations for the AE-M Pretest Total Score</td>
<td>189</td>
</tr>
<tr>
<td>29. Means and Standard Deviations for the AE-M Pretest Multiplication Subscale</td>
<td>190</td>
</tr>
<tr>
<td>30. Two-Factor Analysis of Variance for the AE-M Pretest Total Score and Multiplication Subscale Score by Group and Gender</td>
<td>190</td>
</tr>
<tr>
<td>32. Two-Factor Analysis of Variance for the AE-M Pretest Predicted Initial Estimates on the Multiplication Subscale by Group and Gender</td>
<td>191</td>
</tr>
<tr>
<td>33. Means and Standard Deviations for the ACE Pretest Numerical Subsection</td>
<td>192</td>
</tr>
<tr>
<td>34. Means and Standard Deviations for the ACE Pretest Application Subsection Score</td>
<td>192</td>
</tr>
<tr>
<td>35. Two-Factor Analysis of Variance for the ACE Pretest Numerical and Application Subscale Scores by Group and Gender</td>
<td>193</td>
</tr>
<tr>
<td>TABLE</td>
<td>PAGE</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>36. Means and Standard Deviations for the ACE Pretest Addition Subscale</td>
<td>193</td>
</tr>
<tr>
<td>37. Means and Standard Deviations for the ACE Pretest Subtraction Subscale</td>
<td>194</td>
</tr>
<tr>
<td>38. Two-Factor Analysis of Variance for the ACE Pretest Addition and Subtraction Subscale Scores by Group and Gender</td>
<td>194</td>
</tr>
<tr>
<td>39. Means and Standard Deviations for the ACE Pretest Division Subscale</td>
<td>195</td>
</tr>
<tr>
<td>40. Means and Standard Deviations for the ACE Pretest Multiple Operations Subscale</td>
<td>195</td>
</tr>
<tr>
<td>41. Two-Factor Analysis of Variance for the ACE Pretest Division and Multiple Operations Subscale Scores by Group and Gender</td>
<td>196</td>
</tr>
<tr>
<td>42. Means and Standard Deviations for the AE-M Pretest Numerical Subsection</td>
<td>196</td>
</tr>
<tr>
<td>43. Means and Standard Deviations for the AE-M Pretest Application Subsection</td>
<td>197</td>
</tr>
<tr>
<td>44. Two-Factor Analysis of Variance for the AE-M Pretest Numerical and Application Subsection Scores by Group and Gender</td>
<td>197</td>
</tr>
<tr>
<td>45. Means and Standard Deviations for the ACE Posttest Total Score</td>
<td>198</td>
</tr>
<tr>
<td>46. Means and Standard Deviations for the ACE Posttest Multiplication Subscale</td>
<td>198</td>
</tr>
<tr>
<td>TABLE</td>
<td>PAGE</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>47. Two–Factor Analysis of Variance for the ACE Posttest Total Score and Multiplication Subscale Score by Group and Gender</td>
<td>199</td>
</tr>
<tr>
<td>49. Two–Factor Analysis of Variance for the ACE Posttest Adjusted Estimates on the Multiplication Subscale by Group and Gender</td>
<td>200</td>
</tr>
<tr>
<td>50. Means and Standard Deviations for the ACE Posttest Numerical Subsection</td>
<td>200</td>
</tr>
<tr>
<td>51. Means and Standard Deviations for the ACE Posttest Application Subsection Score</td>
<td>201</td>
</tr>
<tr>
<td>52. Means and Standard Deviations for the ACE Posttest Addition Subscale</td>
<td>201</td>
</tr>
<tr>
<td>53. Means and Standard Deviations for the ACE Posttest Subtraction Subscale</td>
<td>202</td>
</tr>
<tr>
<td>54. Means and Standard Deviations for the ACE Posttest Division Subscale</td>
<td>202</td>
</tr>
<tr>
<td>55. Means and Standard Deviations for the ACE Posttest Multiple Operations Subscale</td>
<td>203</td>
</tr>
<tr>
<td>56. Two–Factor Analysis of Variance for the ACE Posttest Numerical and Application Subscale Scores by Group and Gender</td>
<td>203</td>
</tr>
<tr>
<td>57. Two–Factor Analysis of Variance for the ACE Pretest Addition and Subtraction Subscale Scores by Group and Gender</td>
<td>204</td>
</tr>
<tr>
<td>TABLE</td>
<td>PAGE</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>58. Two–Factor Analysis of Variance for the ACE Pretest Division and Multiple Operations Subscale Scores by Group and Gender</td>
<td>204</td>
</tr>
<tr>
<td>59. Means and Standard Deviations for the AE–M Posttest Application Subsection</td>
<td>205</td>
</tr>
<tr>
<td>60. Means and Standard Deviations for the AE–M Posttest Items With No Response</td>
<td>205</td>
</tr>
<tr>
<td>61. Two–Factor Analysis of Variance for the AE–M Posttest Application Subsection and Items With No Response by Group and Gender</td>
<td>206</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Postal Service Estimation Problem</td>
</tr>
<tr>
<td>2.</td>
<td>An Example Estimate for the Postal Service Problem</td>
</tr>
<tr>
<td>3.</td>
<td>Definition and Example of Translation and Reformulation Processes</td>
</tr>
<tr>
<td>4.</td>
<td>The Relationship Between Computational Estimation and Problem Solving</td>
</tr>
<tr>
<td>5.</td>
<td>The Influence of Instructional Activities on Computational Estimation Abilities</td>
</tr>
<tr>
<td>6.</td>
<td>An Example Numerical Exercise</td>
</tr>
<tr>
<td>7.</td>
<td>An Example Word Problem</td>
</tr>
<tr>
<td>8.</td>
<td>An Example of Extended Contextual Complexity</td>
</tr>
<tr>
<td>9.</td>
<td>A Hypothetical Continuum of Problem Context</td>
</tr>
<tr>
<td>10.</td>
<td>An Example Application Item</td>
</tr>
<tr>
<td>11.</td>
<td>Example Item from the Adjusting Estimates–Multiplication Test</td>
</tr>
<tr>
<td>12.</td>
<td>Example Response Interval for an AE–M Multiplication Item</td>
</tr>
<tr>
<td>13.</td>
<td>Nonequivalent Control Group Design</td>
</tr>
<tr>
<td>FIGURE</td>
<td>PAGE</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>14.</td>
<td>Graphs of Means by Treatment Group and Gender</td>
</tr>
<tr>
<td>15.</td>
<td>Graphs of Means by Treatment Group and Gender for the AE–M Numerical Subsection</td>
</tr>
<tr>
<td>16.</td>
<td>Example of Adjusting an Initial Estimate</td>
</tr>
<tr>
<td>17.</td>
<td>Adjusting Estimates–Multiplication Test Organization</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

The five sensory organs allow humans to perceive and appreciate the beauty that nature provides— the delicate fragrance of a rose; the splendor of a captivating sunset; the night-in-gale’s tranquil soliloquy; the silky softness of a kitten’s fur; the sweetness of a fresh-picked strawberry. The senses constantly supply the brain with a vast array of data to decode, process, encode, and store for future use. Much of the data collected in sensory memory, however, is forgotten unless it is identified, or recognized, and encoded in a more “durable form” (Bourne et al., 1986). Sensory data that is recognized and encoded in long term memory becomes information that is used in higher-order processing to plan actions, form judgements, make decisions, and solve problems.

The higher-order cognitive processes are of particular interest in this document, as they contribute significantly to the acquisition and application of concepts, skills and strategies for generating estimates. In the most general sense, the word estimate suggests the act of appraising, evaluating, and judging the nature or character of people, things; and events. Dewey (1933) suggested three “significant traits” of judgement: 1) a reason to judge— a controversy, doubt, difficulty, or curiosity to be resolved; 2) a process of problem definition— determining important data and interpreting meaning from the data; and 3) a final judgement— a conclusion or decision. A person possessing “sound judgement” is “a good judge of relative values; . . . can estimate, appraise, evaluate with tact and discernment” (p. 120). When viewed broadly, estimation is a process of
judgement that serves the individual in recognizing, interpreting, and analyzing important data with which to plan and make decisions.

Qualitative Estimation

In conversation, the speaker is motivated by a need or desire to communicate observations, ideas, or feelings. As a speaker, one assesses the listener’s knowledge, language ability and level of interest, among other things, in selecting the appropriate words with which to express a given message. The attentive listener will attempt to interpret the auditory information to derive meaning from the speaker’s words. Most mature listener’s however, go beyond the words to infer meaning regarding the speaker’s attitude, internal emotional state, and purpose for communicating a given message. External sources such as tone and volume of voice, rhythm and pattern of speech, facial expressions, and deportment provide invaluable clues to understanding the speaker’s state of mind and intended message.

The processes of word selection and message interpretation involve evaluating and judging, or estimating, phenomena that are not readily quantifiable. This form of judgement is referred to as qualitative estimation. Examples of qualitative estimation are many and varied but would include:

Law: A member of the jury interprets and weighs pertinent evidence regarding the circumstances of a crime including motive, opportunity, and disposition of the accused, to form opinions and ultimately decide whether the preponderance of evidence indicates guilt or innocence.
Business: An employer reviews and evaluates a prospective applicant’s resume and references to decide whether the applicant possesses the requisite skills for success and if s/he will fit into the organization well.

Quantitative estimation

Although the human sensory systems are remarkably capable of perceiving and processing data for use in higher-order cognitive functions, they are limited in their ability to provide information beyond the thresholds of their design. They are not well-suited for perception of microscopic entities such as cells or macroscopic phenomena such as the relationship between interference of radio and television signals and increased solar activity. Because of these limitations, scientists have developed other tools for observing and collecting data such as microscopes and telescopes. Similarly, mathematics has evolved as a valuable tool for extending knowledge and understanding of the objects and events in the world and beyond. In this capacity, it is responsible for illuminating many phenomena that can not be directly perceived by our sensory organs, for example, the orbiting of the earth around the sun.

Mathematics is valued for its intellectual rigor and the certainty that is often associated with precise calculation. While arithmetic can provide precise quantitative information, there are many instances where exact results are unnecessary, undesirable, or impossible to achieve.

A shopper wishing to buy a portable cassette player, a cassette album, and batteries priced at $34.95, $7.95, and $3.49 respectively, wants to know if the fifty hard-earned dollars in his/her pocket is enough to cover the cost of these items including tax. By rounding the prices to the next highest dollar amount and
adding, the shopper can estimate the cost of the items from which an estimate of
the total cost, including tax, can be derived. While the total cost can be easily
calculated, in such instances it is easier and more practical to estimate.

The label on the back of a can of paint often provides a formula for determin­ing
how much paint is required to cover the surface one wishes to paint. One
can use the formula to calculate the number of gallons, quarts, and pints of paint
needed with some degree of accuracy. However, good sense militates against
accuracy in this case as different surface conditions such as smoothness, shade
and color, and other factors affect the amount of paint that will be required to
produce a finished look. In most cases it is more desirable to overestimate the
amount needed so that some paint remains for touching-up.

Many cities host activities such as concerts, sporting events, art exhibits, and
conventions. These events affect the economic prosperity of the community in
diverse ways by increasing the demand for available goods and services. An
accurate accounting of the economic impact is, however, neither practical nor
possible. In assessing such phenomena, it is more appropriate to simplify the
problem and generate an estimate.

The problems posed in the preceding section are examples of what has
come to be known as computational estimation. A broader form of estimation
that involves the explicit use of number concepts, is referred to here as quantita­tive
estimation. In addition to computational estimation, the broader quantita­tive form includes several other varieties used for estimating time, measurement,
extent, and numerosity. Each of these varieties is associated with a specific set of
strategies for generating estimates in the domain, as well as a corresponding and
similar set of reasons why it is useful to do so.
Qualitative and quantitative estimation can be thought of as forming the two ends of a continuum of contexts in which estimation is an appropriate means for reaching "a final judgement." Between these two extremes lie a multitude of problems where a combination of the two forms are used. For example, in diagnosing a patient's illness, a doctor commonly uses both sensory information—obtained from visual (swelling, redness . . .), tactile (probing for tender, sore, or stiff regions), and auditory sources (heartbeat, breathing . . .)—and quantitative information—obtained from blood tests, urinalysis . . .—as clues to the location, source, and extent of distress. The problem confronting the physician is not unlike the problem confronting the speaker, the listener, the juror, or the employer. The results of mathematical and scientific investigation provide data that are used, like data from the sensory organs, to plan actions, make decisions, form judgements, and solve problems.

Computational Estimation

This study focuses on a subset of computational estimation strategies, as well as related mental computation skills and number concepts that support the successful use of these strategies. Moreover, the study explores important instructional issues regarding the use of computer-based activities designed to introduce and provide practice in the use of selected strategies.

Need

The ability to generate estimates in a variety of problem contexts, using a variety of numerical forms (e.g. whole number, decimal, fraction, and percent), operations, and number concepts and skills, is regarded as a very important
mathematical skill (NAEP, 1983; NCTM, 1980; NCSM, 1977). The importance of this skill is, in part, a by-product of the shift in emphasis away from computation and toward acquiring experience in applying concepts, skills, and strategies in problem solving contexts as suggested by the writers of the Agenda For Action (NCTM, 1980). The emphasis now placed on learning to use both calculators and computers as tools for approaching problems of greater sophistication and computational complexity places concomitant importance on the ability to predict, project, and estimate the results of such computations in order to assess their reasonableness.

Although estimation has been identified as an important skill area in mathematics education, studies have found that most students are deficient in their ability to generate estimates. Moreover, while the use of calculators and computers is filled with the possibility of both key-stroke and conceptual errors, it is the over-dependence on computing devices that should be of greatest concern. Timnick (1982) reports a disturbing tendency for people to believe their calculators and computers over their own better judgment.

"The smartest among us doubt our own abilities when pitted against a calculator. We take a machine's word over our own." (Timnick, 1982; p. 10)

This lack of mathematical judgement has been noted by other authors as well. "Number Numbness" or "Innumeracy" are phrases that suggest a certain inability to deal with numbers, especially large numbers (Hofstadter, 1985).

The availability of calculating and computing devices coupled with a lack of number sense describes only part of the need for a broader development of
estimation skills. Almost all businesses and public institutions have adopted computers as a means for achieving greater efficiency and cost effectiveness. Today computers are used to calculate everything from grocery bills to bank balances to income tax liability. It is necessary to be alert to potential errors that can result from such uses.

Regardless of the causes that result in a general lack of number sense and intuition, it is clear that instruction in estimation can be useful in providing tools that will help learners to understand complex computational environments, and help build a greater sense of number as well as contribute to enhanced conceptual development and problem solving ability. If one considers that mathematical literacy is based upon lower-order computational skills and higher-order abilities such as number sense, then the shift in instructional emphasis toward problem solving would seem to suggest a greater need for development of estimation skill.

**Computational Estimation Processes**

In discussing the search for an existing definition of computational estimation, Reys et al. (1980) noted that “no acceptable or complete definition of computational estimation has been found.” In response to this problem, that research team constructed the following operational definition:

The interaction and/or combination of mental computation, number concepts, technical arithmetic skills including rounding, place value, and less straightforward processes such as mental compensation that rapidly and consistently result in answers that are reasonably close to a correctly computed result. This process is done internally, without the external use of a calculating or recording tool. (Reys et. al., 1980; p. 6)
Although this definition identifies many of the most important elements of computational estimation, it lacks consideration of several key constructs that are critical to successful use of computational estimation strategies. The processes of computational estimation are guided and informed by three major constructs: appropriateness, manageability, and reasonableness.

Reasonableness

The word reason suggests the deliberate weighing of possibilities through an intelligent and objective approach with an awareness of the implications that a given solution may bring. In most models, the final stage of problem solving involves "looking back" at the problem statement, the plan, the solution process, and the results to confirm the reasonableness of the solution. The process of assessing the accuracy and sensibility of an answer provides, perhaps, the most common motivation for exploring computational estimation strategies in the mathematics curriculum.

A curious and well-informed reader might use estimation to check the reasonableness of the following claim:

**Postal Service is $30 Billion Business**

During 1986, the postal service employed 780,000 people to collect, process, and deliver 450,000,000 pieces of mail each day to businesses and households in the U.S. and abroad. In 1986, the cost of mailing a first class letter was 22¢ and the total revenues exceeded $30 billion. [This example was extracted from a U.S. Postal Service advertisement in the August 4, 1986 edition of Time magazine.]

Figure 1 Postal Service Estimation Problem
A brief review of this statement suggests that the postal service "handles" 450 million letters and packages each day of the year at an average rate of 22¢ per item. This review yields a simple calculation—365 days X 450 million items X $.22, or an annual revenue of $36,135,000,000 for 1986. In generating an estimate to test this claim, one might round 365 days to 300 and 450 million 500 and $.22 to $.20 to facilitate mental computation\(^1\). The estimate might be completed as follows:

\[
\begin{align*}
1) & \quad (450 \text{ million} \times 365) \times 0.22 \\
2) & \quad (500 \text{ million} \times 300) \times 0.20 \\
3) & \quad 150000 \text{ million} \times 0.20 \\
4) & \quad 15000 \text{ million} \times 2 \\
5) & \quad 30000 \text{ million} \\
6) & \quad 30 \text{ billion}
\end{align*}
\]

Figure 2 An Example Estimate for the Postal Service Problem

The resulting estimate, $30 billion, suggests that the postal service is indeed a 30 billion dollar business given that the numbers supplied reflect a reasonable degree of accuracy.

\(^1\) An example of mental processes involved in generating an estimate for the postal service problem: Treating the information in a problem of this sort as fact ignores certain realities that can be deduced from personal knowledge. First, it is clear that the postal service does not process or deliver mail on Sundays or holidays. Therefore, eliminating Sundays would reduce to about 300, the number of days that should be used in calculating the annual postal revenues. Second, 450,000,000 is a conspicuously round number. It is doubtful that the post office processes that much mail each day. It is more likely that that number reflects an
estimate of the average volume of mail delivered daily. Third, one might question whether 22¢ per item is a reasonable estimate of the cost to deliver the assortment of letters, post cards, bulk-rate mail, and packages that are mailed each day. Consideration of these factors militate against using the data contained in the problem statement to calculate revenue for 1986. This process would yield a better estimate reflecting these considerations. Using 22¢ is unavoidable as additional, more accurate, information that is not available would be needed to generate a more accurate estimate.

Assessing the reasonableness of other's claims, and one's own calculations, involves an evaluation of the problem statement to determine that the numerical data and operations were properly identified and that the mental computations were performed correctly. Although the motivation may differ, the process of determining whether a result is reasonable involves consideration of appropriateness issues and the management strategies employed and therefore is dependent upon them.

Manageability

Reys, Bestgen, Rybolt, and Wyatt (1980) identified two general processes—translation and reformulation—that good estimators commonly employ to make numerical data more amenable to mental computation. These are defined and illustrated in Figure 3.
Translation- “changing the equation or mathematical structure of the problem to a more mentally manageable form. This form is used to computationally process the numerical data...”

<table>
<thead>
<tr>
<th>347 X 6</th>
<th>347 X 5</th>
<th>347</th>
<th>350</th>
<th>=50</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>45</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Reformulation- “Changing the numerical data into a more mentally manageable form.”

| 5335 X 426 | 5000 X 400 | =2,000,000 |

Figure 3 Definition and Examples of Translation and Reformulation Processes

In the postal service example (see Figure 2 ), 365 days was transformed to 300 to better reflect the actual number of days that mail is processed. In line 3, the resulting phrase (150,000 million X $.22) was changed (translated) to 1500 million X $22. Steps two and three reflect the use of management strategies to make the numbers more conducive to mental computation. This notion of rearranging, or changing, numbers and/or the structure of the expression to a more manageable form is critical to success in mental computation and, therefore, estimation. Such alterations reduce the amount of information that must be stored in and retrieved from short-term memory, thus simplifying the computations (Atweh, 1983; Simon, 1982).

In addition to revealing the importance of management strategies, Figure 2 also illustrates the interdependence between these strategies and underlying
number concepts and arithmetic skills. In line 4 of Figure 2, $.20 is translated to $2 by recognizing that dividing 150,000 by 10 and multiplying $.2 * 10 simplifies the numerical expression without changing the result. B. Reys (1986) indicated that good estimators commonly used five different management strategies including front-end, clustering, rounding, compatible numbers, and special numbers. These strategies can be used individually or in combinations. For example, one might use a compatible numbers strategy to simplify the problem 4675/8 by recognizing that 4675 is close to 4800 and that 4800 is divisible by 8. However, one might use a rounding strategy, a compatible numbers strategy, and place value concepts to estimate the result to 46755/82.

\[
\begin{align*}
46755 / 80 & \quad \text{(round 82 to 80)} \\
48000 / 80 & \quad \text{(recognize the 46755 is close to 48000)} \\
4800 / 8 & \quad \text{(divide both numbers in the expression by ten)}
\end{align*}
\]

These examples underscore the interrelationship between management strategies and arithmetic skills and suggests the complexity that is involved in insightful use.

**Appropriateness**

In the preceding discussion, it was implied that determining the sensibility, or reasonableness, of one’s own calculations is the most common motivation for estimating. However, as Usiskin (1986) suggests, there are at least four other reasons for estimating including: 1) constraints force estimates; 2) estimates
increase clarity; 3) estimates are easier to use; and 4) estimating gives consistency.

In the postal service problem, the volume of mail delivered each day is given as 450 million pieces. This number undoubtedly represents an estimate of the average amount of mail processed each working day. The value varies; therefore, it makes sense to use an estimate. Using the 450 million estimate places an additional constraint on the accuracy of the result which Usiskin refers to as estimate-in estimate-out. Even if one were to calculate the revenue using the numbers given, the result would still be an estimate.

Some numbers, particularly large ones, are easier to understand and manipulate when reported in approximate terms. For example, 450 million is easier to read and is much clearer than 450,342,876. It is easier and more efficient to multiply 500 million X 300 days than it is to calculate 450,342,876 X 365. In addition to being easier, Usiskin insists that “answers derived using suitable estimates may be more reasonable and more realistic than those that attempt to be exact.”

It is clear then that one aspect of expertise in computational estimation involves the ability to recognize and interpret problem contexts in which estimation is an appropriate means for generating a solution. While there are no simple rules for determining issues of appropriateness, ill-defined problems and problems that contain ambiguous information or partial numerical data are candidates for estimation. The ability to interpret the accuracy of numerical data, resolve ambiguity, and recognize underlying assumptions are all important factors in the process of determining the appropriate solution path. Other factors
that may influence one’s decision to estimate include amount of time available, perception of accuracy requirements, and tolerance for error.

**Compensation**

The management strategies introduced earlier are valued because they provide a variety of means for reducing numerical complexity and facilitate rapid mental computation. Because of their nature and purpose, these strategies introduce an element of error that reduces the accuracy of estimates generated therewith. Researchers identified two methods that good estimators commonly used to increase the accuracy of their estimates—intermediate and final compensation.

Intermediate adjustments are made “on the fly” to compensate for errors introduced by changing the expression or numerical data therein. In considering the postal service problem, one might round 450 million up to 500 million since 365 was reduced to 300 as an estimate of the days mail is actually delivered. Although three hundred days is actually an underestimate, one might reduce the average cost per item from 22¢ to 20¢ to further compensate for raising the number of items from 450 million to 500 million. These intermediate refinements are performed on the numbers or expressions of a problem prior to generating a final estimate.

Final compensation reflects the same concern for accuracy; however, adjustments are performed after an initial estimate has been generated. In estimating the product of 42 X 72 for example, one might generate an initial estimate of 2800 (40 X 70) and recognize that adding 100 (approximately 2 X 70) would yield a more accurate estimate (2900).
Intermediate compensation is more appropriate in multiple-step problems where error is likely to be compiled through the course of many computations. Using intermediate compensation in such cases can relieve some of the memory overhead that would, inevitably, result if final adjustments were used instead. Techniques for adjusting estimates allow one to account for error that can creep into an estimate. In this way final estimates can reflect the level of accuracy desired without increasing the computation load on short-term memory.

**Estimation and Problem Solving**

The constructs introduced above—appropriateness, manageability, compensation, and reasonableness—can be thought of as phases in reaching a problem solution. In this context, the preceding discussion suggests that computational estimation involves many of the same cognitive and affective conditions that are commonly associated with mathematical problem solving. The phases of estimating, pictured below in Figure 4, have been placed beside corresponding steps in a simplified Polya four-stage problem solving model.
Identifying relevant information and operations, recognizing assumptions, and aspects of ambiguity and uncertainty are critical factors in interpreting and understanding a problem statement. Planning a solution involves elements of process such as identifying and selecting strategies, as well as a knowledge base. A broad knowledge base is necessary to support sophisticated planning and includes both declarative (number facts, relationships, and concepts) and procedural (arithmetic operations, management strategies, and cognitive strategies) knowledge (Das, 1987). Additionally, planning involves aspects of metacognition such as sequencing steps and evaluating the effectiveness of a plan.

A determination of the appropriateness of a solution strategy occurs at some point during the planning phase and is based on one’s knowledge base and
understanding and representation of the problem (Nickerson, Perkins, and Smith, 1985). Similarly, the selection of appropriate management strategies is made during the planning phase. In estimating, the process of carrying out the plan (Do) involves performing mental computations based on the management strategy that was selected during the planning phase.

Although the decision to estimate, the selection of management strategy and the generation of an estimate can be located in the problem solving model with relative ease, strategies for adjusting, or compensating, provide some difficulty as they are not so easily characterized. Adjustments may be considered during planning, execution, or looking back stages. The difficulty in placing compensation strategies may be more a reflection of the limitations of problem solving models as such models attempt to describe in discrete steps a phenomenon that is fluid and intuitive in nature. Compensation is based on a need or desire to produce a more accurate estimate and reflects consideration of error emanating from inaccurate data and mental manipulation of that data. Intermediate adjustments are more likely to occur during planning and execution, while final refinements are likely to occur during the execution and looking back stages.

Issues regarding the reasonableness of a problem solution are based on the problem context, the plan, and the processes involved in executing the plan. These issues are generally resolved during the looking back stage and involve the evaluation of the accuracy and the appropriate directionality of an estimate that is required. An assessment of reasonableness, however, is frequently made during planning (as a function of the executive processes) and may lead to the rejection of a given plan.
The view of computational estimation as a problem solving process is consistent with findings reported by B. Reys (1986). She indicates that good estimators used a variety of strategies and that "they chose these strategies to fit the context of the problem including the numbers and operations involved." Additionally, she observed that this behavior "parallels what we know about problem solving—no one problem-solving strategy is efficient for every problem." Finally, these observations led her to conclude that "part of the task of becoming a good problem solver (or estimator) is being able to select and use a strategy that fits the problems."

Trafton (1986) lends additional support for a broader view of estimation as he suggests that "estimation is a complex skill with many of the same subtleties and complexities as problem solving." He further suggests that expertise involves "developing student's reasoning, insight, and decision-making skills with this topic."

The broad view of estimation espoused here acknowledges its use in assessing the results of problem solutions reached through a variety of other quantitative techniques. Limiting study of computational estimation to considerations of reasonableness, however, ignores certain realities regarding how, when, and why mathematics is used in our daily lives. As R. Reys (1986) indicates, "a recent survey of non-school use of mathematics reported that in more than 80% of the situations surveyed an estimate of the answer was considered sufficient."

Several factors have limited the adoption of a broader view of computational estimation processes. First, mathematics is associated with accuracy and precision which has left estimation with the image of an illegitimate sibling. Second, estimation skills have been taught primarily as a means for checking the
results of other more precise techniques. Third, research has focused on aspects of estimation relating to management strategies. Fourth, because of the emphasis placed upon management skills, much of the literature presents a limited range of problem contexts that typify common word problems rather than richer contexts associated with problem solving. Finally, problem solving processes are among the most highly prized of all human cognitive capabilities; to suggest that computational estimation, a less rigorous form of computation, is related to problem solving appears to be a huge leap in faith.

This exploration of the relationship between computational estimation and problem solving was undertaken to illuminate the important supportive skills that are commonly ignored in discussions of computational estimation. Most authors have focused on aspects of estimation relating to management skills and determining reasonableness of results and have neglected to acknowledge related cognitive skills.

Computers in Instruction

By most accounts, computers have revolutionized aspects of business, government, science, and to a lesser degree education. In some areas, they have simplified common functions such as word processing and record keeping, enhancing both efficiency and productivity. In other areas, they have provided a means for constructing sophisticated mathematical models that simulate a diverse array of phenomena, often revealing important relationships and yielding greater understanding of "events" that are difficult, if not impossible, to analyze without the computational power that computers provide.
In the past decade, the microcomputer has emerged as a valuable tool for facilitating instruction in many areas of the curriculum. Instructional uses have proliferated in recent years and computers have become more accessible to a greater number of learners as a result of greater interactivity, ease of use, and inherent flexibility. Although computers are being used to provide a greater range of instructional opportunities, computer-based instruction remains among the most common uses.

The computer-based learning activities used in this study were designed to introduce students to three rounding strategies for reducing the numerical complexity of two-factor multiplication problems and provide practice in applying these or other strategies to mentally compute estimates in several contexts with numbers of varying size (See Appendix A for a detailed discussion of the the courseware). Additionally, students in one treatment group were introduced to strategies for refining initial estimates generated with two of the three rounding strategies that were taught.

Previously, the processes of computational estimation were shown to include appropriate uses, management strategies, compensation strategies, mental computation, and assessment of reasonableness. The estimation courseware was expected to influence the learning and application of these concepts and strategies as shown below in Figure 5.
The Processes of Computational Estimation

- Determine Appropriateness
- Select Management Strategies Compensation
- Mental Computation Compensation
- Reasonableness Compensation

Figure 5 The Influence of Instructional Activities on Computational Estimation Abilities

While the estimation courseware was designed specifically to teach the management strategies and provide practice in mentally computing estimates, the activities are likely to influence development of important concepts for determining appropriate uses of estimation and assessing the reasonableness of computed results as well, although to a lesser degree.

Two instruments were used to assess subject's estimation ability, both before and after the instructional sequence. The tests included addition, subtraction, multiplication, division, and multiple operation items. Several dimensions of estimation ability were used to provide a more accurate assessment of specific abilities that were taught in the instructional activities. Total score on both tests
is an indicator of general ability, while the numerical and application subsections and the operation subscales provide an index of specific skills.

**Research Questions**

These considerations led to the formation of the following questions regarding expected posttest differences between the treatment and control groups. The questions center on the effectiveness of the estimation courseware in helping students learn how to identify and select the appropriate rounding strategy to generate estimates.

Do students in the treatment groups:

1. generate a greater number of acceptable estimates on all test items?
2. generate a greater number of acceptable estimates on the multiplication subscale?
3. respond to a greater number of items?
4. use rounding strategies to generate estimates for other operation types?
5. generate a greater number of adjusted estimates on the multiplication subscale?

A substantial body of evidence exists indicating that males and females differ in their general mathematical ability. Similar evidence suggests differences in estimation ability based on gender. These considerations supported the investigation of the following questions.
When compared with females, do males:

1. generate a greater number of acceptable estimates across all test items?

2. generate a greater number of acceptable estimates on the multiplication subscale?

3. respond to a greater number of items?

4. use rounding strategies to generate estimates for other operation types?

5. generate a greater number of adjusted estimates on the multiplication subscale?
CHAPTER II
LITERATURE REVIEW

Introduction

Mathematics educators have long recognized the value of computational estimation as an important skill in daily life that should receive more attention in school mathematics. However, this recognition has not translated into more research, better methods of assessment or, most regretfully, broad acceptance in the practice of mathematics in the classroom. The review of literature was undertaken to establish the empirical and theoretical background of this study, exploring the following topics: 1) issues and difficulties in assessing computational estimation; 2) estimation processes and related abilities; 3) the teaching of computational estimation and related instructional issues; and 4) the role of computer-based instruction in teaching computational estimation.

Assessment Issues

Many researchers have chosen to construct their own measures of computational estimation ability because no test developed thus far has been found to be generally useful for exploring the many facets of this complex process with all groups of students.

In a study investigating the effects of systematic instruction with prospective elementary teachers, Bestgen et al. (1980) constructed a 60 item test of computational estimation skill (EST). The test consisted of straight computation, or
numerical, items in open-ended format distributed somewhat evenly across the operations of addition, subtraction, multiplication, and division involving both whole and decimal numbers. Subjects were given five minutes to respond to as many items as possible.

Levine (1980) constructed a twenty-item Test of Estimation Ability (TEA) that was administered with no time limit to 89 undergraduate non-mathematics majors. The test consisted of ten multiplication and ten division items of open-ended format and called for oral responses to items during interviews.

In a study designed to identify important estimation strategies and processes, Reys, Bestgen, Rybolt, and Wyatt (1980) constructed a fifty-five item test to measure the estimation skills of 1200 subject's. The Assessing Computational Estimation test (ACE) consisted of twenty-eight numerical and twenty-seven applied items, all in open-ended format. The test included addition, subtraction, multiplication, division, and multiple-operation items using whole, fractional, and decimal numbers. Test items were projected on a screen for an interval of time that varied with the difficulty of the problem. The researchers used the results of the ACE test to select 59 "good" estimators for think-aloud interviews. They used a fifteen item-test (five numerical and ten applied items) to conduct the interviews.

Rubenstein (1985) constructed a 64-item test to measure performance on four types of computational estimation tasks. She designed four sixteen item scales including open-ended, reasonable/unreasonable, reference number, and order of magnitude. The last three scales used a multiple-choice format. The items were distributed equally across three dimensions—form (32 each of numerical and applied), numbers (32 each of whole and decimal), and operation (16
each of addition, subtraction, multiplication, and division). Each item was projected for fifteen seconds and five seconds were allowed between items for students to record responses.

Threadgill-Sowder (1984) constructed a 12-item test representative of items used on NAEP tests. She used the test to conduct interviews exploring computational estimation procedures used by students ranging in grade from six through nine. The test involved both open-ended and multiple-choice formats and included verbal and applied items covering whole, fractional, and decimal numbers.

Wyatt (1985) constructed a fifty item timed test that he administered using a microcomputer. The test was based on the ACE test and used an open-ended format for numerical items and to explore reasonableness. The results of this test were used to classify ninth grade student's into three levels—good, average, and poor estimators. He selected eighteen subjects from each of the three levels for interviews.

Although researchers have yet to identify a set of standardized psychometric methodologies, efforts to assess computational estimation have served to illuminate important strategies and processes and uncover difficulties associated with different approaches to assessment.

Assessment Difficulties

Assessment of computational estimation skills has presented a number of "psychometric difficulties" that have hampered standardized measurement efforts, stifled research efforts, and slowed the development and acceptance of instructional activities (B. Reys, 1986; Benton 1986). Psychometric techniques
often associated with administering standardized and locally produced tests of mathematical skill are simply not adequate for assessing estimation skills (Reys and Bestgen, 1981). The problems with assessment are a function of the complex nature of computational estimation and the propensity of students to compute answers to estimation items. von Oech (1983) estimates that students take "over 2600 quizzes, tests and exams during their school careers." Homework papers, quizzes, and tests all require the individual to remember, search for, or compute the right answer for each question. In light of this emphasis upon correct answers, it is not surprising that researchers have reported that students commonly "try to work a problem quickly with pencil and paper, then round their answer to reflect an estimate" (Bestgen et al., 1980).

R. Reys (1986) suggested four factors that can threaten the validity of instruments designed to measure computational estimation ability: timing, question format, number complexity, and question context.

Timing

If the time allowed for completing an estimate is not tightly controlled, "a complete computation . . . may actually be performed" (Reys and Bestgen, 1981). Reys and Bestgen (1981) suggested that students should be made aware of the expectation to estimate and that items be administered one at a time to insure rapid pace and greater concentration, both of which will tend to curb the propensity to compute responses.

Question Format

Open-ended test items have been used on many of the tests constructed by researchers. The open-ended format allows students to respond more freely
which provides greater insight into the processes and strategies that they employ, as well as errors that they make. Although open-ended items allow more freedom, they require the consideration of response intervals within which an estimate must fall to be considered acceptable.

Multiple-choice questions have also been used to assess computational estimation skills, however, responses to such items may involve the use of strategies other than estimation (Reys et al., 1980). Although multiple-choice items have some limitations, they do provide a means for measuring other important estimation skills such as determining reasonableness, use of a reference number, and order of magnitude.

Question Context

Estimation items couched in applied contexts have been found to influence students' performance in comparison to numerical items (Reys et al. 1980). In some examples Reys and colleagues found that performance improved when the problem was embedded in an applied context. Although they sighted no counter examples, Reys & Bestgen (1981) suggested that in some cases applied contexts might have a negative influence on item performance which led them to conclude that both types of problems should be used to provide variety and promote realism.

Numbers

In addition to varying the context, researchers have recognized that "the numbers in the question should be complex enough to encourage and reward
estimation." If the numbers and not sufficiently complex, students are likely to attempt computation even when proper time constraints have been used.

**Summary**

While efforts to create effective objective measures of computational estimation ability have failed to produce a standardized instrument(s), they have contributed significantly to a better understanding of important assessment issues that have plagued this area of mathematics education, as well as a set of guidelines for creating instruments (Reys and Bestgen, 1981). Reys (1984) summarized the psychometric difficulties most elegantly when he said, "Validly measuring computational estimation skills clearly creates unique and challenging testing problems." Each of the factors discussed above raises questions about what skills are actually being measured— a factor which compromises the validity of such measures.

**Estimation Processes and Related Abilities**

Investigations of computational estimation have focused primarily on the areas of identification and characterization of estimation strategies and processes, relationships between estimation and other abilities, and the effects of instruction. Studies designed to explore dimensions of strategy use have been among the most fruitful in revealing the complex nature of the processes involved in estimating.
In their study, Reys et al. (1982) identified a set of cognitive and affective characteristics common among the good estimators which they interviewed. The analysis of 59 interviews indicated that good estimators:

1. Possessed a quick and accurate recall of basic facts;
2. Understood and used place value concepts;
3. Were quick and efficient in mental computation;
4. Used the three key processes reformulation, translation, and compensation;
5. Possessed a tolerance for error based on an understanding of the meaning and intent of estimation;
6. Understood and used a variety of arithmetic properties (e.g. distributive, associative, commutative);
7. Used a variety of strategies and approaches; and
8. Were generally "confident in their own estimation ability."

In addition to the findings resulting from interviews, these researchers reported several findings emanating from the analysis of the ACE test that was administered to the entire sample group. Their analyses indicated that: 1) performance increases across all age groups; 2) "marked differences" existed between males and females among all four age groups; and 3) subjects performed consistently better on applied items (Reys et al., 1980).

In an article discussing the teaching of computational estimation concepts and strategies, B. Reys (1986) elaborated on the strategies that had been found among the good estimators. She identified five broad-based strategies including: 1) Front-end; 2) Clustering; 3) Rounding; 4) Compatible numbers; and 5) Special numbers. She also indicated that there was a large degree of variance in the application of these strategies between individuals, as well as differences in how individuals applied the strategies to different problems.
In her study of computational estimation strategies used by non-mathematics majors, Levine (1982) identified eight different strategies that subjects used during think-aloud interviews to generate estimates for open-ended multiplication and division items. She reported that students used an average of four (4.25) strategies and that considerable variety existed in the application of each strategy type. The eight strategy types that she identified are shown in comparison to the key processes identified by Reys et al. (1980) and strategies outlined by B. Reys (1986) in Table 1.

**TABLE 1**

A COMPARISON OF COMPUTATIONAL ESTIMATION PROCESSES AND STRATEGIES DEFINED BY DIFFERENT RESEARCHERS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractions</td>
<td>Translation</td>
<td>Special numbers</td>
</tr>
<tr>
<td>Exponents</td>
<td>Translation</td>
<td>Special numbers</td>
</tr>
<tr>
<td>Rounding both numbers</td>
<td>Reformulation</td>
<td>Rounding</td>
</tr>
<tr>
<td>Rounding one number</td>
<td>Reformulation</td>
<td>Rounding</td>
</tr>
<tr>
<td>Powers of Ten</td>
<td>Reformulation</td>
<td>Special numbers</td>
</tr>
<tr>
<td>Known numbers</td>
<td>Reformulation and/or Translation</td>
<td>Special numbers</td>
</tr>
<tr>
<td>Incomplete Partial Products</td>
<td>Reformulation</td>
<td>?</td>
</tr>
<tr>
<td>Proceeding Algorithmically</td>
<td>Reformulation and/or Translation</td>
<td>?</td>
</tr>
</tbody>
</table>

In their study, Reys et al. (1980) identified the key processes: translation, reformulation, and compensation. Figure 2.1 indicates the relationship between Levine’s strategies, the broader processes identified previously (Reys, Bestgen, Rybolt, and Wyatt, 1980), as well as the strategies outlined by B. Reys (1986). As this comparison indicates, there is considerable overlap in the strategies defined
by these investigators. Levine studied only multiplication and division strategies while the others explored strategies used to generate estimates to a broader range of operations.

When viewed from this perspective, the two sets of findings are consistent. Levine’s finding that estimators employ a variety of strategies and that they apply them differently supports those reported by Reys, Bestgen, Rybolt, and Wyatt (1980).

Levine used the strategy classifications to explore further the relationships between quantitative ability and estimation ability, quantitative ability and the types of strategies used, and quantitative ability and the number of strategies used. She reported that quantitative ability accounted for 55% of the variance in estimation ability, and that there exists both a significant relationship between quantitative ability and the type of strategy used for twelve of the twenty test items, and a significant relationship between quantitative ability and the number of strategies used. These findings also support similar findings reported by Reys, Bestgen, Rybolt, and Wyatt (1980).

Rubenstein used a 64-item estimation test and an 80-item related factors test to explore relationships between computational estimation ability and eight dimensions of mathematical skill. The related factors test was developed to measure mathematical skills believed to be related to computational estimation including: 1) selecting operations; 2) making comparisons; 3) knowing number facts; 4) operating with tens; 5) operating with multiples of ten; 6) knowing place value; 7) rounding; and 8) judging relative size.
The analysis of the estimation test was used to evaluate student's performance on the four dimensions of the test, to determine the relative difficulty of each dimension, and to identify differences between subjects based on gender. With respect to the four types of estimation ability the order of difficulty was open-ended (most), reference number, and order of magnitude (least). The fourth scale, reasonable vs. unreasonable, was excluded from subsequent analyses because it was unreliable. She found that decimals were more difficult than whole numbers and the order of difficulty for the operations was division (most), multiplication, subtraction, and addition (least). Moreover, she found that males scored higher than girls across the three dimensions of estimation ability; however, when evaluated separately, differences were evident only for the order of magnitude scale.

In addition to the analysis of estimation abilities, Rubenstein explored the relationship between computational estimation ability and other mathematics skills. Of the eight dimensions of math ability, she observed that operating with tens, making comparisons, and judging relative size were highly related to estimation ability. Of these, she reported that operating with tens had an "especially strong relationship."

Summary

Despite numerous differences in methodology, population, and research focus, the studies reviewed here support the following generalizations:
1) Students are generally deficient in estimation skills.

2) There is a persistent tendency to compute.

3) Significant differences have been found between males and females.

4) Computational estimation strategies and processes are complex in nature.

Research in Teaching Computational Estimation

The problems that have plagued research in computational estimation generally have limited research investigating instructional methods as well. Benton (1986) indicated that "The most pressing need is for experimental research in estimation and how it should be taught." The need for research on teaching estimation strategies is punctuated by a dearth of instructional materials. Damarin et al. (1988) reported that a review of basic mathematics texts (performed by Freeman et al., 1980) identified estimation as one of the lowest priority topics in mathematics. Studies investigating learning of estimation strategies have employed a variety of psychometric and instructional approaches with different populations to explore both affective and cognitive effects.

In a twelve-week study of 187 preservice elementary teachers, Bestgen et al. (1980) investigated both cognitive and affective dimensions of learning. Each week one treatment group received instruction in the use of one strategy, a quiz, and a short debriefing period following the quiz. The second treatment group received only the weekly quiz while the control group participated in regular classroom activities.
In an analysis of the pretest, the researchers found that students attempted fewer than one-half of the sixty items. Analysis of the posttest results indicated that students receiving instruction showed significantly greater gains. Moreover, they found that students were more successful with whole numbers than with decimals, and that the difficulty of operations, ordered from easy to hard, was addition, subtraction, multiplication, and division. Both results are consistent with those found by Levine (1980) and Rubenstein (1984). Additionally, they found that better estimators "held favorable attitudes and perceived the processes of computational estimation as more understandable and less complicated than did" less able estimators. They summarized their findings by saying:

This research documents the value of providing preservice elementary teachers with computational estimation activities. Brief systematic practice over a 10-week period increased performance on computational estimation activities, furthermore, when such practice was accompanied with specific estimation strategies, greater understanding and respect for estimation processes occurred.

Damarin et al. (1988) investigated the effectiveness of using a sequence of computer-based activities with five-intact high-school classes including: two ninth grade general math classes, two tenth-grade algebra classes, and one twelfth-grade geometry and trigonometry class. Three computers were placed in the classroom and the estimation activities were provided as an alternative to beginning a new assignment, reflecting an "implicit assumption of the study that computer-based instruction on estimation could be added to the classroom with minimal disruption to the progress of the existing mathematics curriculum."

Pretest and posttest scores were analyzed and compared using paired t-tests for
each class individually, and collectively. The analyses indicated that “the treatment was appropriate for students at each grade level, and each class improved their estimation scores significantly,” although “the amount of improvement varied across classes.” This study was based on a one-group pretest/posttest design, which places certain limitations on inferences drawn from it. However, as the authors report, “the improvement was so robust that a genuine treatment effect seems likely.”

Benton (1986) reported on three additional studies conducted by Schoen et al. (1981), Nelson, (1967), and Sutherlin (1977). Nelson taught fourth and sixth-grade students a rounding procedure for generating estimates, and he found that the experimental groups performed better than control groups on a self-constructed test of estimation skill. Sutherlin taught sixth-grade students a procedure for producing estimates with decimal numbers. The results indicated no differences between the treatment and control groups. Finally, Schoen et al. (1981) taught fourth-graders to estimate the products of exercises involving numbers of varying size. In this study Schoen and colleagues found that students became better estimators and that “estimation in whole number computation can be taught in a short period of time.”

The review of the literature regarding instruction in computational estimation, although limited, indicates that there is little overlap among the studies that have been conducted in terms of population, instructional approach, and strategies taught. However, most of these studies have shown that teaching estimation strategies can have a measurable impact on student’s ability to estimate within the domain taught.
Instructional Issues

There is little debate that computational estimation strategies and processes need to be taught as estimation has been characterized as "one of the most neglected areas of the mathematics curriculum" (Reys et al., 1980). The existence of the NCTM Yearbook, Estimation and Mental Computation (Schoen and Zweng (eds.), 1986), suggests that much institutional support exists for research and development of instructional activities. Many researchers have offered reasons why estimation should be taught, however, none have done so more elegantly, or effectively, than Usiskin (1986) who offered the following reasons why estimation should be integrated into the curriculum:

1) Today's instruction is misleading;
2) Estimation is often the preferred alternative or the only alternative possible in a situation;
3) Estimation can lead to new insights about exact procedures;
4) Wrong procedures may give very good estimates;
5) Estimation ideas affect student grading and placement; and
6) Ignoring estimation gives students a distorted view of mathematics.

In the preface to Estimation and Mental Computation, Schoen offers the following additional support for increased integration of estimation activities:

One important theme that arises frequently throughout the volume is the interdependency between the process of estimating in a particular domain and the understanding of mathematical concepts within that domain. Estimation aids in concept development, but at the same time a solid conceptual understanding improves one's ability to make good estimates. It follows as a corollary to this theme that an important by-product of learning to estimate is better conceptual understanding, and, conversely, concepts must be
understood in order to acquire the flexible set of processes and decision-making rules needed by the proficient estimator. (pg. vii)

The interdependency between estimation and other mathematical concepts and abilities, as Schoen suggests, is well established in the literature. The empirical and theoretical work represented by the NCTM 1986 Yearbook indicates that there is no shortage of ideas regarding how estimation can and should be taught. However, mathematics educators are still confronted with the reality that the development of estimation activities is complicated by many of the same difficulties that have hampered research generally.

The view of estimation that was advanced in Chapter I, which is supported by the literature, suggests that estimation is a complex ability involving a set of related skills, concepts, processes, and affective traits which are necessary to support its insightful use. In order for instruction to be effective, it must be sensitive to both the cognitive and affective conditions that influence learning.

According to Trafton, there are four broad areas that must be explored to support the development of an “estimation mindset.” These areas are: legitimacy and usefulness; flexible thinking and decision making; adjusting initial estimates; and recognition of sensible answers. Trafton offers a variety of suggestions for facilitating learning in each of these four areas of which the following subset is presented:

Establishing Legitimacy and Usefulness

• use real world examples
• emphasize the language of estimation
• use a variety of instructional approaches
Develop Flexible Thinking and Decision Making

• use examples that require students to determine the type of estimate needed
• present examples showing different approaches

Learning to Adjust Initial Estimates

• ask students to evaluate the relationship of their estimate to the actual answer

Building Recognition of Sensible Answers

• use non-computational examples which require students to provide a reasonable estimate

A number of strategies for generating estimates have been identified as appropriate for instruction, including those elaborated by B. Reys (1986): front-end; clustering; rounding; compatible numbers; and special numbers. The following examples are presented to explore a subset of the instructional issues that must be addressed if learning activities are to be meaningful and successful.

The front-end strategy is a useful technique for generating estimates, particularly for addition and subtraction. In using the front-end method (shown below in Figure 6), one generates an estimate by using only the front-end, or most significant, numbers.

![Figure 6 An Example Numerical Exercise](image)
Trafton suggests the extensive use of real world examples as a factor in establishing the legitimacy and usefulness of estimation. The use of contexts, as illustrated in Figure 7, provides opportunities for emphasizing the language of estimation, an additional factor related to usefulness and legitimacy.

Figure 7  An Example Word Problem

Responses such as, "at least $50," "less than $55," and "more than $52," indicate a sensitivity to accuracy and the direction of the estimate, both of which are traits that should be encouraged. Figure 7 also illustrates an alternative approach to
adding the front-end numbers—B. Reys (1986) calls this clustering. The use of real world contexts also provides opportunities to explore other aspects of the problem. The alternative questions, also shown in Figure 7, explore considerations that are attendant with such real world problems.

The problem presented in Figure 8 provides the opportunity for students to supply estimates of non-computational phenomena—the cost of dinner for two at a medium-priced restaurant, for example—and can be used to highlight alternative approaches, flexible thinking, and decision making processes.

Suppose that you want to treat your parents to an evening out, including dinner at a medium-priced restaurant and a trip to the movie theater. About how much money will you need to give them to cover the evening's costs?

Figure 8  An Example of Extended Contextual Complexity

The preceding examples (Figures 6, 7, and 8) suggest the existence of a continuum of problem context ranging from simple to complex. Figure 9 illustrates the hypothetical continuum and the relative position of the different examples on the continuum. There is a correlation between the problem complexity and the intellectual demands that are placed on the learner/estimator—as complexity increases, so do the demands.
In Figure 7, the problem is well-defined— it is clear that an estimate of the total expenses for the evening is required. The question confronting the estimator is “What strategy should I use?” The problem in Figure 8 is, on the other hand, ill-defined, requiring more of the learner in terms of interpreting, analyzing, and understanding the problem, as well as the use of background knowledge, planning, and decision making. In order to produce an estimate of the amount adequate to cover the expenses that may be incurred in Figure 8, one might begin by recognizing:

- an estimate is the only reasonable approach.
- an overestimate is preferable.
- a need to define what is meant by the term “medium priced restaurant.”
- that prices of dinners will vary considerably regardless of the definition.
- that “a trip to the movies” should be defined to include expenses for confections such as popcorn, soft drinks, and candy.

After clarifying the problem by resolving these issues, one might proceed by planning the sequence of steps that would lead to a solution, for example: Estimate the cost of 1) dinner for two; 2) movie tickets; 3) snacks and select an
appropriate management strategy (front-end, perhaps) for estimating the total cost. Finally, after executing the plan and producing an initial estimate, one might look back to see if the estimate is, in fact, reasonable. It is important to acknowledge that the preceding analysis is only one of many possible alternatives that might be used to approach this problem. For example, one might begin by asking the question— “Is $100 enough?”— and continue to reduce the estimate by approximating the cost of just dinner— “Is $50 enough to cover dinner expenses?”— and so on. Regardless of the approach, problems like the example in Figure 8 involve greater intellectual sophistication in terms of analysis, planning, and decision making than do word problems such as shown in Figure 7.

**Teaching Computational Estimation**

Although different authors have used different words to express the idea, the phrase “use real world problems” is a common refrain (Reys & Bestgen, 1981; Trafton, 1979; Reys, 1984). While “real world problems” do provide certain instructional advantages, it is obvious that problems of greater complexity are not appropriate for the introduction and development of specific management strategies. The question of how estimation should be taught has been addressed in part, at least on a theoretical level, by a number of researchers. Trafton (1986) observed that Instructional programs must do two things:

First, students must become aware of and develop proficiency with several useful procedures for finding estimates. Second, instruction needs to address a cluster of variables whose purpose is to establish an “estimation mindset.” (pg. 16)
Sequencing

His point is clear—strategy development should precede the exploration of other important estimation ideas. This indicates that there are different levels at which instruction should take place, and that instruction within each level should be sensitive to particular issues. The interdependency between estimation strategies and related mathematical skills further suggests that the introduction of estimation strategies should follow the development of related concepts—which is assumed for the purposes of this discussion. A three level model for instruction is indicated by the examples in Figures 6, 7, and 8. For convenience, these are labeled strategy introduction, practice, and application.

The development of a broad set of estimation strategies imposes a number of other instructional conditions that must be addressed as well, including sequencing, timing, and feedback. Estimation strategies, such as those introduced by B. Reys (1986), allow one to manipulate the numbers in an equation, or the equation itself, in order to reduce numerical complexity to a level suitable for mental computation. It is apparent that strategies should be introduced using examples that highlight appropriate numerical manipulation and de-emphasize issues relating to context such as appropriateness, compensation techniques, and even reasonableness to some degree. Subsequent practice sessions should begin to increase the complexity by slowly introducing additional considerations.

In addition to considerations of problem context, there are several other aspects of instructional sequencing that remain ill-defined including:
• what order should be used to teach strategies for the different arithmetic operations?
• what order should be used for strategies within an operation?
• when should considerations of accuracy and strategies for adjusting estimates be introduced?

These represent but a subset of the questions that need to be addressed regarding the sequencing of instructional activities. Additionally, it is obvious that exercises involving the identification and selection of an appropriate strategy can not be addressed until learners have a sufficient number of strategies in their repertoire to choose from. Moreover, there are issues relating to analyzing the problem, planning, and decision making that deserve attention as well.

Timing

Earlier in this review, it was noted that when asked to estimate, students commonly respond by computing the answer rather than attempting an estimate. It was also noted that to reduce the tendency to compute, researchers often administered test items individually using a time limit. The literature on estimation is, at best, sketchy on the use of timing within instructional activities.

Bestgen et al. (1980) used weekly quizzes as a form of practice, and imposed a time limit “to encourage students to estimate mentally.” Damarin and Stull (1984) noted that preset time limits adopted for practice exercises “may be appropriate for some students,” but “not for others.” Moreover, they indicated that self-imposed time limits may negate potential “motivational value.”

From the standpoint of logic alone, it appears that timing would be counterproductive on both the strategy introduction and application levels. Until the specific procedures of a given strategy have been automatized, timing is likely to
interfere with the acquisition of those procedures. The goals associated with more sophisticated problem contexts (extended application) involve the development of "flexible thinking and decision making" (Trafton, 1986). The use of time constraints in these contexts is likely to reinforce "quick and dirty" strategies at the expense of broad-based well-conceived approaches. As an instructional variable, timing may be more appropriate during practice exercises after the learner has acquired the basic procedures (Damarin and Stull, 1984).

Feedback

Of the instructional issues covered in this review, feedback has, undoubtedly, received the least attention. The only practical suggestion that has been offered, one echoed frequently, is "accept a variety of answers" (Trafton, 1986; Reys, 1984). The lack of attention to feedback can be traced to the complexity of the instructional task and ultimately to the lack of studies investigating instruction in computational estimation.

In a reference to the problems associated with feedback, Damarin and Stull (1984) noted that the "design of appropriate feedback... is, at best, a delicate business." The "delicacy" stems from a conflict between the goals for different levels of instruction. At the introduction and practice levels, instruction focuses on the acquisition of a specific strategy. Therefore, feedback at these levels should be designed to emphasize and reinforce the use of that strategy, although, it should be sensitive enough to accommodate reasonable estimates generated with other strategies as well.

A primary goal that is pursued at the application level is to foster strategy diversification, therefore, the definition of reasonableness should be less rigid
than for other levels. Timing and context are additional factors in determining reasonableness. The first alternative question in Figure 2.2 (How much money does Rick need to cover these expenses?) illustrates the limits that context can place on estimates. The problem calls for an overestimate, hence underestimates are really unreasonable and, therefore, unacceptable.

It is clear that increasing problem complexity will support the development of "flexible thinking and decision making skills." However, Trafton (1986) suggests that instructional activities should also involve students in the analysis of the type of estimate that is needed for a given problem, as well as examples demonstrating various approaches to the same problem.

Summary

This review of issues suggests that the development of broad-based computational estimation ability is a complex instructional task involving the consideration of appropriate prerequisite conditions and cognitive and affective learning goals. R. Reys (1984) acknowledged the need for the development of a scope and sequence framework for instruction on estimation and suggested a "traditional spiral approach to ensure that students can apply computational estimation strategies used for whole numbers, decimals, fractions, and percentages." It is clear that such a framework needs to address the issues outlined above. Although no universally accepted framework exists, the literature does provide a foundation for identifying the major goals of instruction in computational estimation.
**Cognitive Goals**

Previous research has established the relationship between mathematical abilities and computational estimation abilities, indicating that related arithmetic skills and concepts are important cognitive prerequisites that need to be developed prior to learning computational estimation strategies (Reys et al., 1980; Rubenstein, 1983). In addition to supporting the development of these interrelated skills, Trafton (1986) suggests that the primary goal of instruction in computational estimation is the development of an "estimation mindset." The establishment of an "estimation mindset" involves the following cognitive goals.

Students should develop:

1) a variety of strategies for generating and refining estimates;

2) a set of procedures for identifying and selecting appropriate management and adjustment strategies;

3) a set of procedures for determining when and why it is appropriate to estimate;

4) a set of procedures for evaluating, or judging, reasonableness of results;

5) methods for determining what kind of estimate is needed, how accurate it should be, and what direction it will fall in comparison to the computed result; and

6) reasons why estimates are used.

Moreover, Trafton (1986) suggested flexible thinking and decision making processes as important corollary cognitive goals of instruction in estimation.
Affective Goals

In addition to the cognitive goals of instruction, the literature suggests the need for a corresponding set of affective goals (B. Reys, 1986; Trafton, 1986). If instruction is to be effective, it will have to support the development of critical attitudes, values, and self-development dimensions of learning. Important attitudes and values necessary for developing an "estimation mindset" were suggested by B. Reys (1986) and Trafton (1986).

Students need to develop:

1) an awareness of, and appreciation for, estimation;

2) a perception of its’ usefulness; and

3) comfort with the limitations of accuracy.

Consideration of instructional goals would not be complete without paying some attention to internal and external affective conditions that play an important role in self-development.

Computer-based Instruction in Computational Estimation

The 1984 NCTM Yearbook, Computers In Mathematics Education, serves as a testimony of the widespread perception that computers can play a valuable instructional role in mathematics education. Additionally, a variety of research studies have indicated that computer-based instruction can be an effective means for providing learning experiences. The central question addressed in this section relates to the potential contribution of computer-based instruction in teaching computational estimation strategies and processes.
Trafton (1986) suggested that teachers should "emphasize estimation regularly" and that activities be pursued using a variety of instructional modes. Presumably, computer-based learning activities would comprise one mode of individualized instruction.

Microcomputers possess an impressive array of features that support active involvement and can influence learner's affective and cognitive growth (Damarin and Stull, 1984). Instructional designers have acknowledged that computational speed, generation of random numbers, display of textual and graphical information, sound, and interactivity are among the most important characteristics underlying the computer's instructional potential. These characteristics can be exploited in the design of computer-based instruction to create learning activities sensitive to the goals of computational estimation (Damarin et al., 1988).

Interactivity

Actively engaging students in learning activities is an idea that is at least as old as the notion of organized schooling. Conventional wisdom suggests that a greater degree of learning occurs when students are actively involved in analyzing, synthesizing, and evaluating instructional tasks. It is not surprising, therefore, that interactive potential is considered a critical feature of exemplary courseware. The interactivity of a given courseware activity is a function of learner control of pacing and program sequence, as well as aspects of feedback. Although activities which facilitate active involvement are generally preferred, the degree of control which students are allowed in a given activity should be consistent with the content being explored and the instructional goals of the activity (Damarin and Stull, 1984).
Graphics and Animation

High-resolution color graphics, animated sequences, and sound generation are important features that can enhance computer-based learning activities (Malone, 1984; Loftus and Loftus, 1983). These features can be used to emphasize critical attributes of a learning task and can enhance learning by increasing attention and motivation (Malone, 1984). In addition to enhancing focus on critical instructional features, color and animated graphics can add “new dimensions to learning activities by providing situations and illustrations that were difficult, if not impossible, without the computer” (Damarin, 1984; p. 70).

Practice and Problem Variation

Several authors have noted the need to “emphasize estimation regularly” through exercises on daily assignments. (Trafton, 1986; Osborne, 1979). In the instructional issues discussion, it was observed that the acquisition of estimation strategies could be viewed as a three-step process involving increasingly complex problem contexts—strategy introduction, practice, and extended application. Computer-based activities are well suited for at least two of the three instructional levels. Damarin and Stull (1984) indicated that tutorial and drill and practice software could be used to provide initial instruction (introduction and practice). Moreover, they indicated that drill and practice activities and games could be used to provide opportunities to develop important procedures for identifying and selecting appropriate management strategies.

In addition to problem context, Reys and Bestgen (1981) indicated that the numbers in an exercise need to be sufficiently complex “to encourage and reward
estimation." Computer-based exercises are, perhaps, uniquely suited for providing a wide range of problems of varying levels of numerical complexity. The ability to randomly generate numbers provides additional advantages in terms of presenting a "unique set" of practice problems for each learner and also supports the reuse of an activity for learners requiring additional practice.

Feedback

One of the major criticisms that is commonly made of textbook estimation activities is that the problem key provides a single correct answer rather than a range in which estimates should fall. Generally, textbook activities provide an introduction to a specific strategy and exercises (predominantly numerical) with which to practice the strategy. Publishers are correct in assuming that if the strategy that is introduced is used properly, it will admit of only one answer. However, the broader goal of encouraging and reinforcing the development of a diverse set of strategies is subverted with a single correct answer approach to feedback.

Computer-based activities for teaching strategy introduction can be designed to provide a range of feedback that can be sensitive to the narrow goal (strategy introduction), as well as the broader goal (strategy diversification). Feedback can be sensitive to the narrow goal by informing students whether they used the expected strategy, and the broader goal by acknowledging when an alternative strategy produces a reasonable result. An additional advantage of computer-based instruction relates to its ability to provide corrective feedback.

The level of informativeness is an important characteristic of feedback. Studies have shown a positive relationship between highly informative feedback
and improved performance in learning (Carter, 1984). Textbook exercises, providing only one answer, provide a very low level of information to the learner—right and wrong. Computer-based activities can be designed to provide information regarding both the quality and directionality of a learner’s estimate.

Computer-based Instruction and the Goals of Estimation

In an earlier discussion, instruction in estimation was posited as involving three well-defined levels of instruction including strategy introduction, strategy practice, and extended application. These three levels of instruction involve somewhat different goals at each level, although it was indicated that estimation activities should support the development of critical higher-order cognitive processes such as flexible-thinking and decision-making, as well as an awareness of and appreciation for, the usefulness of estimation strategies.

The literature, limited though it may be, supports the use of computer-based estimation activities particularly at the introduction and practice levels (Damarin and Stull, 1984; Damarin et al., 1988). Although the Damarin and colleagues (1988) indicated that they used an estimation activity designed to involve extended application, no research was conducted to explore the influence of that activity upon student’s development of broader cognitive processes. The software that was used in their study was developed prior to 1986. At that time internal memory capacity of Apple II computers was limited by comparison to today’s somewhat more capable machines. One must wonder what impact a faster computer with greater memory capacity might have on the development of a new generation of courseware designed to provide opportunities for extended application with a sensitivity to the broader goals of estimation.
Summary of Literature Review

The empirical and theoretical foundations explored for this study involved areas of the computational estimation literature including assessment, identification and characterization of estimation strategies and processes, and issues relating to instruction and the use of computer courseware to teach estimation strategies. Although the research in these areas was sparse, the addition of conventional wisdom from theoretical sources was sufficient to provide guidance for further research in each of the areas mentioned.

With respect to assessment, the literature converges upon the difficulties associated with validly measuring computational estimation abilities. Much of the work in this area has been pursued as a result of interest in other phenomena such as exploration of the utilization of estimation strategies and related abilities or evaluation of instructional influences. Tests of estimation ability have varied greatly as a result of these primary interests.

Research designed to explore estimation processes and related abilities have involved a broad range of age and ability groups. Findings have generally indicated widespread deficiency in estimation ability. Moreover, researchers have found that ability to generate estimates is a complex process involving other aspects of mathematical ability.

The review of literature pertaining to instruction on computational estimation topics indicated that studies of this type are, indeed, rare. Although most studies have found positive instructional influences, there is little overlap either in strategies and processes taught or in the age and ability of participating students. Conventional wisdom drawn from previous research in computational
estimation and mathematics education generally, has served to bolster instructional methodology. However, the lack of research in the various areas of computational estimation has limited the organization and development of instructional activities.
CHAPTER III
METHODS AND PROCEDURES

Introduction

Many mathematics educators have suggested the need to provide instruction that facilitates the development of estimation strategies and related mental computation skills (Bestgen, 1986; Trafton, 1986; Reys, et al., 1982). Although research indicates that some people develop a set of rather sophisticated estimation strategies, little is known about how such skills are developed in the absence of a coherent instructional plan. Traditional approaches to teaching estimation have faced numerous difficulties that indicate the need to: 1) investigate a wide range of strategies for generating estimates with a variety of operations and sizes of numbers, as well as number types (e.g. whole, fractional, and decimal); 2) provide consistent, frequent and timed and untimed practice; 3) provide a variety of problem contexts where estimation is an appropriate "tool" for solving the "problem;" and 4) provide appropriate timely feedback.

The estimation software used in this study is designed to encourage students to develop strategies for refining their estimates as several of the activities include features that provide an opportunity to work at levels of increasing need for accuracy. The need for increased accuracy encourages students to try to develop methods for refining initial estimates. The feedback in most activities includes information regarding the quality and direction of estimates made, which provides an environment that rewards the refinement of estimates. Within the software explicit instruction is limited to rounding strategies, and in
the absence of instruction in methods for formulating other specific strategies for refinement, it is unclear what strategies will naturally evolve. It is possible that inherent characteristics of the software will facilitate the formation of legitimate compensation strategies. It is more likely, however, that students will formulate strategies that only approximate effective techniques for refining initial estimates.

The purpose of this study was to explore the influences of computer-based learning activities with respect to learners development of mental computation skills and estimation strategies. More specifically, the study was designed to investigate learner's development of strategies for generating and refining estimates of the results of two factor multiplication exercises, as well as the transfer of those skills and strategies to other arithmetic operations in numerical and problem contexts.

Procedures

Sampling

Subjects for this study consisted of 148 students in six eighth grade general mathematics classes selected from two middle schools in the Columbus Public School district. The two schools were selected for participation by the Columbus Public Schools based on their similarity in demographic characteristics, students' math ability, and availability of computer hardware necessary to support the instructional activities investigated in this study. The treatment and control conditions were randomly assigned to the six classrooms yielding two classrooms, or approximately 50 students, for each of the three conditions.
Enrollment at the one middle school was 700 students of which 42% were minority. The attendance rate at this school was 84% and 77% of the students received free or reduced price meals. Five hundred eighty-three students were enrolled at the second middle school of which 46% were minority. Attendance at this school averaged approximately 90% and 62% of the students received free or reduced price meals. For the purposes of comparison, system enrollment for middle schools was 14,418 students, 48.4% of which were minority. The school-wide attendance rate was 90% and 59% of the students received free or reduced price meals.

Instructional Setting

The computer laboratory at each school was equipped with fifteen Apple Ile computers, each with at least one disk drive, and a color monitor. The instructional plan required that each student work independently at a computer during each period in the lab. As a result of the limited number of computers and the instructional constraints, classes were divided into two groups of approximately twelve students each who accompanied this experimenter to the computer lab for each instructional session. The remaining sub-group stayed with the math teacher to continue normal instructional activities. Estimation was not pursued in regular math classroom activities during this study. Each sub-group received approximately thirty minutes of instruction in estimation twice weekly on alternating days—Monday and Wednesday for sub-group 1, and Tuesday and Friday for sub-group 2. Instruction was scheduled for six class periods over a three weeks period, or a total of approximately three hours.
During the course of this investigation, the experimenter was responsible for administering pretests, instructional activities, and posttests to the experimental groups. Because the schedule overlaps between the two schools, an assistant administered both pretests and posttests to the control group. The assistant was trained in the methodology for test administration, and the same set of instructions was used for introducing the test booklets and answer sheets to all groups. Both the experimenter and the assistant followed the procedures outlined in Appendix C.

Treatments

The treatments used in this investigation consisted of five sets of computer-based instructional activities and concise introductions to each set of these activities. The TABS Estimation Software is a series of four diskettes developed by the TABS-Math project at The Ohio State University under the direction of Dr. Suzanne K. Damarin. Each diskette contains instructional activities designed to help students develop mental computation skills and estimation strategies for addressing two-factor multiplication and division "problems." A fifth disk, the Adjusting Estimates activities, was developed by the investigator to explore the effects of explicit instruction designed to teach strategies for refining estimates. The adjusting estimates activities were based on two of the three management strategies addressed in the TABS Estimation Software. A complete description of the instructional activities has been included in Appendix A.

Subjects assigned to both treatment conditions used the TABS Estimation Software to learn three basic strategies for managing the numbers in a two-factor multiplication problem—rounding both numbers up, rounding both numbers
down, and rounding one number up and one down. Initial courseware activities were designed to familiarize students with these three basic numerical contexts, provide practice in identifying and selecting the appropriate management strategy, and completing the mental arithmetic to generate an estimate of the product. Subsequent activities provided further practice in identification, selection, and computation, as well as opportunities for applying these, or other, strategies in several different familiar contexts.

In addition to the TABS Estimation Software, treatment group 1 used the Adjusting Estimates activities to learn strategies for adjusting, or refining, two kinds of initial estimates—round both numbers up and round both numbers down. The sequence of instruction for the two groups is shown below in Table 2 in the order that the software was presented.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Treatment Group 1</th>
<th>Treatment Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round Up/Round down</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Bull's-Eye</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>At The Races</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Adjusting Estimates</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Estimation Invasion</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Your Choice</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

A typical instructional period began with a brief (two to five minutes) introduction to the courseware which highlighted appropriate methods for interaction
and other important courseware features, and outline student's responsibilities in terms of the sequence of activities and the minimum number of exercises that should be completed for each activity. The investigator avoided any references to specific estimation strategies or related techniques during introductory sessions. After the introduction, students were instructed to begin working on the day's activities. Students encountered very few problems generally, however, when problems did arise, the investigator offered only technical help regarding the courseware in an attempt to avoid extraneous instructional influences.

**Instruments**

Subjects in both treatment and control groups were tested for general estimation ability (ACE) and for more specific skills related to the instructional content (AE-M). The Accessing Computational Estimates test (ACE), a timed of test of computational estimation, was developed for use in the National Institute of Education study conducted at the University of Missouri (Reys, Bestgen, Rybolt, and Wyatt, 1980). The ACE was adapted to create two parallel fifty-four item tests, forms A and B. Adaptations were rendered by the TABS-Math project for use in evaluating field-test versions of the TABS-Estimation software. An item analysis was performed on the parallel forms during analysis of field test results and Alphas of .96 and .91 were found respectively for the pretest and posttest versions (Damarin et al., 1988). See Appendix B for examples of parallel test items.

The first half of the test is composed of twenty-eight arithmetic items (e.g. 458/16). These items included addition, subtraction, multiplication, and division operations as well as multiple operation items (e.g. 40 * 70/12). Some items
involve decimals and/or fractions as well. The second half of the test consists of twenty-seven applied items that were accompanied by a picture, or illustration, that provides a context in which the estimate is generated.

![A 12 Pack of Cola](image)

About how much does one cost?

Figure 10 An Example Application Item

For students in the seventh and eighth grades, the authors of the ACE test used response time intervals ranging from twelve to seventeen seconds. Twelve seconds were allowed for single operation items while seventeen seconds were allowed for items involving division or multiple operations. These time intervals were used for the administration of both forms. Additionally, three to four seconds were provided for the individual to finish recording item responses and to turn to the next page.

**Adjusting Estimates—Multiplication Test**

The original purpose of the ACE was to identify subjects who evidenced a high level of skill in computational estimation. While the ACE is a good test of general estimation ability, it is not sensitive enough to measure those skills which
were taught in the treatment conditions. In order to determine more precisely the effects of instruction in generating and refining initial estimates, a second test, the Adjusting Estimates–Multiplication test (AE-M), was created. Two parallel forms consisting of 34 items in each booklet were constructed for measuring estimation skills that were taught in the instructional sequence. Appendix C outlines the individual items, and the amount of time allowed for completion, as well as the response interval used for each item. Although similar in form and content to the ACE test, the AE-M consisted of items that were keyed to strategies introduced in the treatment conditions and, therefore, provided a better measure of those skills and strategies that were taught. Part one consisted of twenty arithmetic items that increase in difficulty. Addition, subtraction, and division items were used to introduce some variety, relieve computational strain, and promote feelings of accomplishment. Moreover, four of the twenty items force the student to compare two two-factor multiplication “problems” to determine which result would be larger, as shown below in Figure 11.

Which is LARGER?
187 X 34 or 190 X 30
A. B.

Figure 11 Example Item from the Adjusting Estimates–Multiplication Test
The second section of the AE-M consisted of fourteen applied items. Like the applied section of the ACE test, each item was embedded in a context suggested by an accompanying illustration (refer to Figure 10). The applied items were keyed to the strategies and some involved contexts familiar from the instructional sequence.

Inasmuch as the AE-M was designed to be more sensitive to strategies and skills that were introduced in the instructional activities, timing of individual items was considered an important element of test design. The multiplication items were created to measure student's learning of strategies for generating and refining initial estimates. Adjusting an initial estimate requires some additional time to identify the "parts" of the problem that should be used to refine the estimate and to complete the mental arithmetic. Therefore, fourteen seconds were allowed for each single operation multiplication item. Twelve seconds were allowed for students to complete addition and subtraction items and seventeen seconds for division and multiple operation items. A pilot study of the AE-M instrument indicated that the time limits established for the ACE instrument—which were used as a guideline for the AE-M—were somewhat narrow. As a result time intervals that were used in this study were expanded as indicated.

For the purposes of identifying specific aspects of student's performance on the tests, several subscales were identified for each instrument. Each test consisted of two subsections—numerical and application. The total score for each test and these subsection scores form one basis for evaluating learning that occurred. Although these scores are helpful for making general comparisons
between the groups studied, they provide no information and little insight regarding specific differences between treatment and control groups. In order to isolate and identify instructional effects more specifically, ACE test items were grouped by operation yielding five subscales—addition, subtraction, multiplication, division, and multiple operation. Table 3 lists the number of items for each subscale.

Table 3

ACE Posttest Items by Subscale

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Number of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical (part 1)</td>
<td>28</td>
</tr>
<tr>
<td>Application (part 2)</td>
<td>27</td>
</tr>
<tr>
<td>Addition</td>
<td>8</td>
</tr>
<tr>
<td>Subtraction</td>
<td>12</td>
</tr>
<tr>
<td>Multiplication</td>
<td>18</td>
</tr>
<tr>
<td>Division</td>
<td>11</td>
</tr>
<tr>
<td>Multiple Operations</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>55</td>
</tr>
</tbody>
</table>

The Adjusting Estimates–Multiplication test was constructed with a focus on multiplication as is suggested by the list of items in Table 4. The test consists of two subsections which were used for general comparisons. Additionally the seventeen multiplication items were used as a subscale to evaluate more specific aspects of expected learning phenomena including the number of estimates generated, the number of predicted initial estimates, and adjusted estimates.
The pretest battery was administered to all groups during the week preceding instruction and the posttest battery was administered during the week following the completion of instruction. Test administration required the majority of two class periods (consecutive days). The ACE was administered on the first day and took an entire class period (42 minutes) to complete. The AE-M took approximately 30 minutes and was administered on the second day for both pre and posttest assessments.

Data Coding and Scoring

Data Coding

With the exception of the five dichotomous items appearing on the AE-M, both tests used open-ended test items that required the student to mentally
calculate a response. Students were instructed to write their response to each item in the appropriate space on the answer sheet or to place a line, or an X, in the space for any test item where they were unable to generate a response. Moreover, they were instructed to pay close attention to recording the appropriate units (e.g. dollars, cents, miles per gallon) for application items.

Despite the various warnings regarding careful recording of item responses, several students neglected to keep track of the order of test items and/or appropriate units for the estimate. Although instructed to cross out items for which no response was made, some individuals neglected to do so and lost track of items and recorded some responses out of order. In most cases it was not possible, or proper, to reorder the items. In cases where students neglected to record units of the estimate and the unit was critical, the estimate was marked "out of range" as it was impossible to determine the student's intentions. In the example, "30 pencils for $1.55: How much for each one?", $.05, 5¢, or just .05 would be considered within the range of a reasonable estimate, while a response of just 5 would be considered out of range.

The answer sheets were used to record student's responses in the database management application of AppleWorks® Integrated Software package. The database was used to generate text files that could be read by an Applesoft BASIC program for the purposes of analyzing and scoring item responses.

Response Categories

In assessing computational estimation skills mathematics educators have acknowledged that different individuals use different strategies for generating
their estimates. Since there is no single correct answer to an estimation item, it is appropriate to establish a range within which an estimate should fall to be considered reasonable. Accordingly, the original developers of the ACE test established response intervals that ranged from an average of eight percent below to eight percent above the computed response for all items on the test. While this average appeared somewhat narrow, individual items varied considerably as they were designed to include responses based upon a broad range of strategies. Similar guidelines were used in constructing intervals for the two parallel forms of the ACE test used in this study.

A response to an estimation item was placed into one of five different categories based on the range, or interval, established for the item. Student’s responses to ACE test items were placed into one of the five categories listed in Table 5. Items where no response was made were marked as a zero. Responses that fell into categories two, three, or four were included in compiling the total and subscale scores.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Used for compiling total and subscale scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No response</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Over (out of range)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Over (within range)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Computed</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Under (within range)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Under (out of range)</td>
<td></td>
</tr>
</tbody>
</table>
In creating the response intervals for the Adjusting Estimates–Multiplication test, similar considerations were made as the range for responses to addition, subtraction, division, and multiple operation items was kept fairly narrow. The intervals for the multiplication items, however, were broadened to include both truncation and rounding up strategies. This broadening allowed responses based on the selection of an inappropriate management strategy to be included as estimates. Figure 12 illustrates the two strategies that form the range for an example item.

<table>
<thead>
<tr>
<th>Predicted strategy</th>
<th>Alternate strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>44 * 33 = 1452</td>
<td>40 * 30 = 1200</td>
</tr>
<tr>
<td>50 * 40 = 2000</td>
<td></td>
</tr>
</tbody>
</table>

Figure 12  Example Response Interval for an AE-M Multiplication Item

Any estimate falling between 1200 and 2000 was considered an appropriate response for this item. Response intervals for each item on the AE-M are included in Appendix C.

As with the ACE test, responses to items on the AE-M fell into six different categories. However, two additional categories were included to track responses to multiplication items with greater specificity. Table 6 includes predicted initial estimates and adjusted estimates that fall between the predicted initial estimate and the computed response.
Table 6

Categories of Response for AE-M Multiplication Subscale

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Used for compiling subscale scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No response</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Over (out of range)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Over (within range)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Computed</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Under (within range)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Predicted (round)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Adjusted (predicted)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Under (out of range)</td>
<td></td>
</tr>
</tbody>
</table>

**Scoring**

A set of Applesoft BASIC programs was created to analyze student’s responses to test items for both the ACE and Adjusting Estimates–Multiplication tests. The programs were designed to evaluate ACE responses based on the categories shown in Tables 5 and AE-M responses consistent with categories shown in Table 6. The programs were used to compile the following scores and totals:
Table 7
Aspects of Scoring for the AE-M and ACE Tests

<table>
<thead>
<tr>
<th>ACE</th>
<th>AE-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical</td>
<td>Numerical</td>
</tr>
<tr>
<td>Application</td>
<td>Application</td>
</tr>
<tr>
<td>Total</td>
<td>Total</td>
</tr>
<tr>
<td>Addition</td>
<td>Multiplication</td>
</tr>
<tr>
<td>Subtraction</td>
<td>Predicted Initial Estimate</td>
</tr>
<tr>
<td>Multiplication</td>
<td>Adjusted Estimate</td>
</tr>
<tr>
<td>Division</td>
<td>Items with no response</td>
</tr>
<tr>
<td>Multiple Operations</td>
<td></td>
</tr>
<tr>
<td>Number of items computed</td>
<td></td>
</tr>
<tr>
<td>Items with no response</td>
<td></td>
</tr>
</tbody>
</table>

The scores generated by the analysis and scoring programs were entered into StatView™, a full-range statistical package created by Abacus Concepts (1986). StatView was used to analyze the data and prepare the descriptive and ANOVA tables presented in Chapter 4 and appendices D and E of this document.

Experimental Design

The process of selecting an appropriate experimental design, one that would support and facilitate the evaluation of the problem, mirrored the development of the problem context. In conceptualizing the final experiment, it was obvious that in-tact classrooms would have to be used because of limitations relating to the availability and location of computer hardware within the schools. It is important to note that most public middle schools in the Columbus metropolitan
area have organized their microcomputer resources into laboratory environments, with very few schools having computers available for use in regular classroom settings. This virtually eliminated the possibility of assigning subjects to treatments. For these reasons, the Nonequivalent Control Group design (shown in Figure 13) was used to explore the potential effects of the the instructional sequence (Campbell & Stanley, 1963; Gay, 1981).

\[
\begin{array}{ccc}
O(1) & X_1 & O(2) \\
O(1) & X_2 & O(2) \\
O(1) & X_3 & O(2) \\
\end{array}
\]

Figure 13 Nonequivalent Control Group Design

O(1) and O(2) represent the pretest and posttest scores respectively for each subject for all measures. The treatment groups are represented by X1 and X2, while the control group is indicated by X3.

Sequence of instruction and gender were the independent variables, while total score and scores on subscales were used as dependent variables. A two-factor analysis of variance was conducted to establish initial estimation abilities and to determine the relative equivalence of all groups with respect to each of the scores indicated in Table 3.6. Identical analyses were conducted for posttest scores to evaluate the effects of the instructional sequence.

Internal Validity

An important aspect of planning and executing a research study involves demonstrating the linkage between treatment and measurement variables. If
proper controls are not exerted, the researcher may find that other hypotheses are equally able to explain any effects that were measured. The ability to eliminate competing hypotheses increases the certainty that the predicted effects resulted from treatment conditions, thereby strengthening the linkage (Krathwohl, 1985). Rival hypotheses present a threat to the internal validity of a study and limit the investigator's ability to demonstrate that such a relationship exists.

Campbell and Stanley (1963) outlined eight major sources of internal invalidity which include: 1) History; 2) Maturation; 3) Testing; 4) Instrumentation; 5) Statistical Regression; 6) Selection; 7) Mortality; and 8) Interaction of Selection and other sources (e.g. Maturation, etc.). The Nonequivalent Control Groups design, a quasi-experimental design, provides elements of control for six of the eight major sources of internal validity, including: History, Maturation, Testing, Selection, and Mortality. It does not, however, eliminate potential threats to validity stemming from statistical regression or interactions between subject selection and other sources. In elaborating these categories of threats, Campbell and Stanley noted that potential effects related to statistical regression are of a questionable nature with respect to the Nonequivalent Control Groups design. When subjects are selected on the basis of extreme pretest scores, there is some tendency for their posttest scores to regress toward the mean score for the sample group. Statistical regression, therefore, can create effects that may be confused with those of the treatment variable(s). In the present study, however, intact groups were used and pretests were evaluated to determine initial equivalence of groups. This procedure tends to strengthen the linkage between treatment and measurement variables by eliminating statistical regression as a competing explanation.
The Nonequivalent Control Groups design does not specifically control for interactions between selection and the other extraneous factors indicated above. However, competing explanations of this sort are commonly based on “selection differences that distinguish the experimental and control groups” (Campbell and Stanley, 1963). As with statistical regression, the combination of selection procedures and the process of establishing pre-intervention equivalence militate against rival explanations based on selection interactions.

In addition to the eight major threats to internal validity, Krathwohl (1985) suggests that Researcher Expectancy Effect, Diffusion, and Reactive Effects can be responsible for observed differences between groups. While these rival explanations can not be ignored, interactions with students were highly structured and planned to exclude any overt references to expectations or involvement in a special project. This would not exclude, however, other subtle expressions, either verbal or non-verbal, that might have been perceived by the participants. Diffusion effects appear least threatening, as the experimental and control groups attended different schools.

Krathwohl (1985) lists six reactive effects that could potentially confound experimental results. Of these six factors, Hawthorne and Novelty effects are most likely to present a confounding influence in the present study.

**External Validity**

In discussing aspects of threats to external validity, Krathwohl (1985) observed that “generality is always an inference; it is a leap of faith to generalize from any given instance to others like it.” As with internal validity, the ability to
eliminate rival explanations increases external validity and the degree of confidence one has that experimental results will, in fact, generalize to other instances.

Campbell and Stanley (1963) noted four sources of competing hypotheses that can threaten external validity including: 1) the reactive effects or interaction effects of testing; 2) the interaction effects of selection biases and the experimental variable; 3) reactive effects of experimental arrangements; and 4) multiple treatment interference. Multiple treatments were not used in this study and can not be considered to be a threat to the external validity of this study.

The Nonequivalent Control Groups design does not provide for control of reactive effects due to testing, indicating that either pre or posttests may have an influence on experimental results as generalized to other circumstances. Two separate tests of estimation ability were used in this study which, in some sense, increases the probability of such reactive effects. Campbell and Stanley (1963) observed that instruments bearing a high degree of similarity to tests commonly used in educational settings are less likely to involve interactive effects. Precisely the opposite is true of instruments used in this study.

Potential sources of invalidity due to selection bias and experimental arrangements are questionable regarding the Nonequivalent Control Groups design (Campbell and Stanley, 1963). Rival hypotheses regarding effects of selection bias are commonly based on questions of the representativeness of the chosen sample. The six intact classrooms used in this investigation were selected by the school district partially on the basis of available computer time. The larger population of eighth grade students may differ from the sample group in terms of computer experience, as the basis for selection might indicate, a variable which may have particular meaning in this study.
Reactive effects due to experimental arrangements may be found in settings where the circumstances of the study lead students to conclude that they are involved in an experiment— a feeling that "I am a guinea-pig" (Campbell and Stanley (1963). This possibility can not be excluded in the present study as instructional activities were conducted outside of the classroom.
CHAPTER IV
FINDINGS

This study was organized to investigate both general and specific effects of the instructional activities on the mental computation and estimation skills and strategies of participating eighth grade students. The research design required that subjects be tested before and after the treatment intervention. Two tests were used to measure subjects' estimation abilities, the Assessing Computational Estimation test (ACE) and the Adjusting Estimates-Multiplication Test (AE-M). The ACE test was used to measure general estimation ability while the AE-M focused on estimating the results of multiplication exercises.

In the process of evaluating the instructional goals, a set of eight hypotheses was generated regarding anticipated cognitive effects. The list of hypotheses, shown below, includes four major hypotheses of particular interest and four minor hypotheses of somewhat less significance, but important to understanding the cognitive impact of the instructional activities nonetheless.

**Major Hypotheses**

**HO: 1** There will be no difference between the experimental and control groups on the ACE posttest in:

a. total score or
b. multiplication subscale score

77
HO: 2 There will be no direct effects between gender, or interactional effects between gender and instructional treatments on the ACE posttest in:

a. total score or
b. multiplication subscale score

HO: 3 There will be no difference between the experimental and control groups on the AE-M posttest in:

a. total score;
b. multiplication subscale score;
c. Number of predicted rounded responses on the multiplication subscale; or
d. Number of adjusted responses on the multiplication subscale.

HO: 4 There will be no direct effects between gender, or interactional effects between gender and instructional treatments on the AE-M posttest in:

a. total score;
b. multiplication subscale score;
c. Number of predicted rounded responses on the multiplication subscale; or
d. Number of adjusted responses on the multiplication subscale.

Minor Hypotheses

HO: 5 There will be no difference between the experimental and control groups on the ACE posttest in:

a. Numerical
b. Application
c. Operation subscale scores; or
d. Number of items computed.
HO: 6  There will be no direct effects between gender, or interactional effects between gender and instructional treatments on the ACE posttest in:

   a. Numerical
   b. Application
   c. Operation subscale scores; or
   d. Number of items computed.

HO: 7  There will be no difference between the experimental and control groups on the AE-M posttest in:

   a. Numerical subscale score;
   b. Application subscale score;
   c. Total number of responses generated; or
   d. Number of items computed.

HO: 8  There will be no direct effects between gender, or interactional effects between gender and instructional treatments on the AE-M posttest in:

   a. Numerical subscale score;
   b. Application subscale score;
   c. Total number of responses generated; or
   d. Number of items computed.

Procedural Concerns

A two-factor analysis of variance (ANOVA) was performed on the ACE total score, subsection scores, and each of the operation subscale scores as indicated in the list of hypotheses. A total of eight analyses were conducted for the ACE. Similarly, a two-factor ANOVA was also performed on the AE-M pretest data for total score, subsection scores, multiplication subscale score, predicted initial estimates and adjusted estimates found on items in the multiplication subscale,
and the number of items with no response. A total of seven analyses were conducted on the AE–M pretest data.

In accordance with hypotheses 5d, 6d, 7d, and 8d, data was collected and analyzed for the number of items that subjects computed. The reader will remember that subjects' use of estimation strategies was expected to increase as a result of instruction. A reduced reliance upon computation was predicted for the treatment groups based on the increased use of estimation strategies. Although this hypothesis appears to be a logical consequence of instruction, these analyses were abandoned because of difficulties in defining and assessing computed responses. For example, the answer to problem number eleven on the ACE posttest, 22 divided by 73, is a repeating decimal (.301369863). What answer would be considered a computation? .30? .301? .3014? Students are generally taught to report fractional results to a fixed number of decimal places (frequently two) which complicates the categorization of responses as the individual's intent must be considered in determining whether or not the response was computed. It would appear that the only reasonable way to test a hypothesis of this sort would be through an interview, or verbal report, process—a process not pursued in this investigation.

**Level of Significance**

A total of fifteen separate analyses were conducted on both the pretest and posttest data. As a result of the rather large number of analyses that were conducted, it was considered appropriate to adopt a more rigid standard for the level of significance that each individual analysis should achieve in order for the
null hypothesis to be rejected. Adopting a more conservative level for alpha (.001 in this case) decreases the probability of committing a Type I error—falsely rejecting the null hypothesis when it is true (Minium, 1978)—across all analyses.

Following the logic of Kennedy (1978), fifteen separate tests conducted on an independent sample would yield the following probability for each of three levels of significance considered (.05, .01, and .001) where \( n = 15 \):

\[
\begin{align*}
(.05) & \quad P^m = .95^{15} = .46 \\
(.01) & \quad P^m = .99^{15} = .86 \\
(.001) & \quad P^m = .999^{15} = .98
\end{align*}
\]

For these levels of alpha .05 yields approximately a 54% chance, .01 a 14% chance and .001 a 2% chance of committing a Type I error across all analyses. Clearly the .05 and .01 levels do not provide a sufficient protection against commission of Type I errors.

While adopting a more conservative standard will reduce the probability of committing Type I errors, the practical implication of this decision may result in accepting the null hypothesis when it is actually false—Type II error. This phenomena could present a problem with respect to pretest data as some differences that actually exist may not be considered statistically significant. As the following analysis of the pretest data indicates, adopting the .01 level would have made no difference in the findings for the pretest data. However, using the .05 level, differences between the interaction of treatment group and gender would have been found in six instances.
Analysis of Pretest Data

The pretest battery was administered to students in all groups during the week prior to the beginning of instructional activities. One hundred and fifteen of the one hundred and forty-eight students were present for the ACE pretest and one hundred and twelve were present for the Adjusting Estimates-Multiplication pretest. The pretests were used to assess computational estimation skills and to determine what differences existed, if any, between the three groups prior to the treatment intervention.

Pretests

The ACE test consists of fifty-seven items, twenty-eight numerical and twenty-seven applied. The potential range of scores for the test was from 0 to 57. The observed range for ACE pretest scores was from 0 to 35. The AE–M consists of thirty-four items, twenty arithmetic and fourteen application. The potential range of scores for this test was from 0 to 34. The observed range for AE–M pretest scores was from 1 to 26.

Missing Data

Of the 148 students that participated in this study, only 115 were present for the ACE pretest and 112 were available for the AE–M pretest. Additionally, 109 and 100 students were present for the ACE and AE–M posttests respectively. Ninety-two students (62%) were present for both pre and posttest versions of the ACE test, while only 87 (59%) took both forms of the AE–M test. These numbers
reflect a high degree of absenteeism that was problematic throughout the course of this investigation. In order to compensate for part of the missing data, mean values were substituted for the appropriate pretest scores of all students who had taken the posttest. This procedure was discussed by Tabahnicht and Fidell (1983) as one of four potential approaches to handling missing cases. Although only 92 students actually completed both forms of the ACE test, 109 students were used in the analysis of the ACE data. Similarly, 87 students completed both forms of the AE-M test, and 100 students were used for data analysis.

Because of the large number of analyses performed, all descriptive statistics for the pretest measures are presented in Appendix D, except where a significant difference was indicated. Similarly, the two-factor ANOVA tables are reported in Appendix D, except where a significant difference was indicated. Table 8 provides a convenient summary of all the two-factor analyses that were performed on the pretest measures.
TABLE 8
A SUMMARY OF THE RESULTS OF TWO-FACTOR ANOVAS FOR ALL PRETEST MEASURES

<table>
<thead>
<tr>
<th>Pretest Measure</th>
<th>Treatment Group</th>
<th>Gender</th>
<th>Treatment by Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>p</td>
<td>F</td>
</tr>
<tr>
<td><strong>ACE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Score</td>
<td>.793</td>
<td>.455</td>
<td>.037</td>
</tr>
<tr>
<td>Multiplication Subscale</td>
<td>.195</td>
<td>.823</td>
<td>.079</td>
</tr>
<tr>
<td>Numerical Subsection</td>
<td>2.041</td>
<td>.135</td>
<td>.019</td>
</tr>
<tr>
<td>Application Subsection</td>
<td>.197</td>
<td>.821</td>
<td>.040</td>
</tr>
<tr>
<td>Addition Subscale</td>
<td>3.359</td>
<td>.039</td>
<td>.056</td>
</tr>
<tr>
<td>Subtraction Subscale</td>
<td>1.556</td>
<td>.216</td>
<td>.293</td>
</tr>
<tr>
<td>Division Subscale</td>
<td>1.254</td>
<td>.290</td>
<td>.539</td>
</tr>
<tr>
<td>Multiple Operations Subscale</td>
<td>1.400</td>
<td>.251</td>
<td>.028</td>
</tr>
<tr>
<td><strong>AE-M</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Score</td>
<td>.059</td>
<td>.943</td>
<td>.599</td>
</tr>
<tr>
<td>Multiplication Subscale</td>
<td>.649</td>
<td>.525</td>
<td>.855</td>
</tr>
<tr>
<td>Predicted Initial Estimates</td>
<td>.714</td>
<td>.493</td>
<td>.078</td>
</tr>
<tr>
<td>Adjusted Estimates</td>
<td>2.56</td>
<td>.083</td>
<td>7.295</td>
</tr>
<tr>
<td>Numerical Subsection</td>
<td>.462</td>
<td>.632</td>
<td>1.492</td>
</tr>
<tr>
<td>Application Subsection</td>
<td>.392</td>
<td>.677</td>
<td>.049</td>
</tr>
<tr>
<td>No Response Items</td>
<td>10.457</td>
<td>.000**</td>
<td>.370</td>
</tr>
</tbody>
</table>

* p < .01  ** p < .001

The ACE pretest provided two major points of interest regarding student’s performance, total score and multiplication subscale score. The total score provided an indication of student’s general estimation ability that existed prior to treatment intervention. The multiplication subscale was used to provide a specific index of student’s ability to estimate the results of multiplication exercises. Descriptive statistics for these measures appear in Tables 25 and 26 respectively in Appendix D.
A two-factor analysis of variance (ANOVA) was performed on the ACE pretest data to determine if initial differences existed between the three groups for these two measures. The following hypotheses were tested:

**HO: 1** There will be no difference between the experimental and control groups on the ACE pretest in:

a. total score or  
b. multiplication subscale score

**HO: 2** There will be no direct effects between gender, or interactional effects between gender and instructional treatments on the ACE pretest in:

a. total score or  
b. multiplication subscale score

Results from this analysis (presented in Table 27) indicated that no significant difference existed between subjects based on treatment group, gender, or the interaction between treatment group and gender. Hypotheses HO:1 and HO:2 can not be rejected at the .001 level of significance. Therefore, the groups can be considered to be similar with respect to general estimation ability and ability in multiplication as measured by the ACE test.

Student’s performance on the AE-M pretest provides four major sources of interest—total score and multiplication subscale score, and the number of predicted initial estimates and adjusted estimates generated on the multiplication subscale. Multiplication subscale scores provided a more specific index of student’s ability to estimate the results of multiplication exercises. Descriptive statistics for the total scores, multiplication subscale and predicted initial estimates scores are presented in Appendix D (Tables 28, 29, and 31 respectively).
Descriptive statistics for the number of adjusted estimates on the multiplication subscale are presented in Table 9.

A two-factor ANOVA was performed on the AE-M pretest data to determine what differences existed between the groups on the four measures discussed above. The results of analyses for total score, multiplication subscale score, and predicted initial estimates are shown in Tables 30 and 32 in Appendix D. The results for adjusted estimates are presented below in Table 10. The following hypotheses were tested:

**HO: 3** There will be no difference between the experimental and control groups on the AE-M pretest in:

a. total score;

b. multiplication subscale score;

c. Number of predicted rounded responses on the multiplication subscale; or

d. Number of adjusted responses on the multiplication subscale.

**HO: 4** There will be no direct effects between gender, or interactional effects between gender and instructional treatments on the AE-M pretest in:

a. total score;

b. multiplication subscale score;

c. Number of predicted rounded responses on the multiplication subscale; or

d. Number of adjusted responses on the multiplication subscale.
TABLE 9
MEANS AND STANDARD DEVIATIONS FOR THE AE-M PRETEST ADJUSTED ESTIMATES ON THE MULTIPLICATION SUBSCALE

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>14</td>
<td>.69</td>
<td>.908</td>
</tr>
<tr>
<td>T2</td>
<td>16</td>
<td>.41</td>
<td>.793</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>.64</td>
<td>.674</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>.58</td>
<td>.780</td>
</tr>
</tbody>
</table>

TABLE 10
TWO-FACTOR ANALYSIS OF VARIANCE FOR THE AE-M PRETEST ADJUSTED ESTIMATES ON THE MULTIPLICATION SUBSCALE BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Adjusted Estimates</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (A)</td>
<td>2</td>
<td>3.23</td>
<td>2.560</td>
<td>.083</td>
</tr>
<tr>
<td>Gender (B)</td>
<td>1</td>
<td>9.22</td>
<td>7.295</td>
<td>.008*</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>2.97</td>
<td>2.348</td>
<td>.101</td>
</tr>
<tr>
<td>Error</td>
<td>94</td>
<td>1.263</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < .01

Results from the two-factor ANOVAs for total score, multiplication subscale score, and predicted initial estimates indicated that no significant differences existed between the treatment groups for hypothesis 3. Therefore, hypotheses 3a through 3d can not be rejected at the .001 level of significance. Additionally, results regarding hypotheses 4a, 4b, and 4c indicate that no significant difference
existed on the basis of gender or interaction between gender and treatment
group. Hypotheses 4a through 4c cannot be rejected at the .001 level of signifi-
cance. However, differences were evidenced in the number of adjusted estimates
generated on the basis of gender. Therefore, hypotheses 4d was rejected at the
.001 level of significance with respect to the differences between gender. Al-
though Table 10 provides some evidence that some differences existed between
females and males regarding the number of adjusted estimates generated, the
groups can be considered to be similar with respect to estimation ability as mea-
ured by the AE-M.

In addition to the major elements of interest from the ACE pretest, students
performance was analyzed for both the numerical and application subsections,
and the four remaining arithmetic operation subscales. Descriptive statistics for
these analyses are presented in Appendix D (See Table 33, 34, 36, 37, 39, and 40).
A two factor ANOVA was conducted for each of the six dependent measures to
test the following hypotheses:

HO: 5 There will be no difference between the experimental and control
groups on the ACE pretest in:

   a. Numerical;
   b. Application; or
   c. Operation subscale scores.

HO: 6 There will be no direct effects between gender, or interactional
effects between gender and instructional treatments on the ACE
pretest in:

   a. Numerical;
   b. Application; or
   c. Operation subscale scores.
The results from these analyses, shown in Tables 35, 38, and 41, indicate that no significant differences existed between the groups on any of the six dependent measures. Hypotheses HO:5 and HO:6, therefore, can not be rejected at the .001 level of significance. Therefore, the groups can be considered to be similar with respect to general estimation as measured by both subsections and the four operation subscales of the ACE pretest.

Three minor points of interest from the AE-M pretest included the numerical and application subsections and the total number of items that individuals did not respond to across all test items. The descriptive statistics for the numerical and application subsections appear in Appendix D as Tables 42 and 43. The descriptive statistics for the “items with no response” appear below in Table 11.

A two-factor ANOVA was performed on the AE-M pretest data to determine if initial differences existed between the means of three groups in terms of subsection scores and number of items with no response. The following hypotheses were tested:

**HO: 7** There will be no difference between the experimental and control groups on the AE-M pretest in:

a. Numerical subscale score;
b. Application subscale score; or
c. Total number of items with no response.

**HO: 8** There will be no direct effects between gender, or interactional effects between gender and instructional treatments on the AE-M pretest in:

a. Numerical subscale score;
b. Application subscale score; or
c. Total number of items with no response.
TABLE 11
MEANS AND STANDARD DEVIATIONS FOR THE AE-M PRETEST ITEMS WITH NO RESPONSE

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Mean SD</td>
<td>N</td>
<td>Mean SD</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>14 9.12 8.070</td>
<td>18 8.81 4.906</td>
<td>32 8.95 6.367</td>
</tr>
<tr>
<td>T2</td>
<td>16 11.65 9.563</td>
<td>13 10.06 8.009</td>
<td>29 10.94 8.782</td>
</tr>
<tr>
<td>T3</td>
<td>18 4.14 5.124</td>
<td>21 3.60 3.254</td>
<td>39 3.85 4.170</td>
</tr>
<tr>
<td>Total</td>
<td>48 8.09 8.205</td>
<td>52 7.02 5.964</td>
<td>100 7.54 7.112</td>
</tr>
</tbody>
</table>

TABLE 12
TWO-FACTOR ANALYSIS OF VARIANCE FOR THE AE-M PRETEST TOTAL NUMBER OF ITEMS WITH NO RESPONSE BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Items with No Response</th>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment (A)</td>
<td>2</td>
<td>451.12</td>
<td>10.457</td>
<td>.000**</td>
</tr>
<tr>
<td></td>
<td>Gender (B)</td>
<td>1</td>
<td>15.95</td>
<td>.370</td>
<td>.545</td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>2</td>
<td>3.54</td>
<td>.082</td>
<td>.921</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>94</td>
<td>43.14</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** p < .01

The results shown in Table 44 in Appendix D indicated that no significant differences existed between the groups, between gender, or interactions between treatment group and gender for either subsection score. Therefore, hypotheses 7a, 7b, 8a, and 8b can not be rejected at the .001 level of significance. Additionally, Table 12 indicates that no significant difference existed between gender or interaction between treatment group and gender for the number of items with no
response. Therefore, hypothesis 8c can not be rejected at the .001 level. The results from Table 12 do indicate, however, that a significant difference existed between the treatment groups with respect to hypothesis 7c. Therefore, hypothesis 7c was rejected.

**Analysis of Posttest Results**

An analysis of the posttest data was conducted to identify differences that existed between treatment groups, gender, and the interaction between treatment groups and gender. These differences will be used to establish the relationship between the instructional activities and observed differences in subject's mental computation and estimation abilities as measured by the posttest measures. The analysis of pretest data revealed that the three groups possessed roughly equivalent abilities. Therefore, it was appropriate to use an identical framework for evaluating the posttest data.

A series of eight analyses (two-factor ANOVAs) were performed on the ACE posttest data and seven analyses were conducted for the AE-M posttest data. The .001 level of significance was used for the same reasons that were discussed above— to protect against the commission of Type I errors. As noted above, the probability of committing a Type I error increases exponentially as the number of comparisons increases.

**Posttests**

The posttest battery was administered to students in all groups during the week following instructional activities. One hundred and eight of the one hundred and forty-eight students were present for the ACE posttest and one
hundred were present for the Adjusting Estimates–Multiplication pretest. The posttests were used to assess computational estimation skills and to determine what differences existed, if any, between the three groups as a result of the treatment intervention.

The ACE posttest consists of fifty-seven items, twenty-eight arithmetic and twenty-seven applied. The potential range of scores for the test was from 0 to 57. The observed range for ACE posttest scores for the 108 students was from 1 to 42. The AE–M posttest consisted of thirty-four items, twenty arithmetic and fourteen application. The potential range of scores for this test was from 0 to 34. The observed range for AE–M posttest scores was from 3 to 28.

Because of the large number of analyses performed, all descriptive statistics for the posttest measures are presented in Appendix E, except where a significant difference was indicated. Similarly, the two-factor ANOVA tables are reported in Appendix E, except where a significant difference was indicated. Table 13 provides a convenient summary of all the two-factor analyses that were performed on the posttest measures.
TABLE 13
A SUMMARY OF THE RESULTS OF TWO-FACTOR ANOVAS FOR ALL POSTTEST MEASURES

<table>
<thead>
<tr>
<th>Posttest Measure</th>
<th>Treatment Group F</th>
<th>p</th>
<th>Gender F</th>
<th>p</th>
<th>Treatment by Gender F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ACE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Score</td>
<td>.765</td>
<td>.468</td>
<td>.679</td>
<td>.412</td>
<td>2.418</td>
<td>.094</td>
</tr>
<tr>
<td>Multiplication Subscale</td>
<td>.607</td>
<td>.547</td>
<td>1.593</td>
<td>.210</td>
<td>.896</td>
<td>.411</td>
</tr>
<tr>
<td>Numerical Subsection</td>
<td>1.101</td>
<td>.336</td>
<td>1.656</td>
<td>.201</td>
<td>2.994</td>
<td>.054</td>
</tr>
<tr>
<td>Application Subsection</td>
<td>.321</td>
<td>.726</td>
<td>.085</td>
<td>.771</td>
<td>1.585</td>
<td>.210</td>
</tr>
<tr>
<td>Addition Subscale</td>
<td>1.158</td>
<td>.318</td>
<td>.932</td>
<td>.337</td>
<td>2.387</td>
<td>.097</td>
</tr>
<tr>
<td>Subtraction Subscale</td>
<td>1.356</td>
<td>.262</td>
<td>.111</td>
<td>.740</td>
<td>1.216</td>
<td>.301</td>
</tr>
<tr>
<td>Division Subscale</td>
<td>.927</td>
<td>.399</td>
<td>.015</td>
<td>.902</td>
<td>5.485</td>
<td>.006*</td>
</tr>
<tr>
<td>Multiple Operations Subscale</td>
<td>.745</td>
<td>.477</td>
<td>.026</td>
<td>.872</td>
<td>1.194</td>
<td>.307</td>
</tr>
<tr>
<td><strong>AE-M</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Score</td>
<td>9.792</td>
<td>.000*</td>
<td>.000</td>
<td>.993</td>
<td>5.150</td>
<td>.008*</td>
</tr>
<tr>
<td>Multiplication Subscale</td>
<td>15.360</td>
<td>.000*</td>
<td>.085</td>
<td>.772</td>
<td>4.758</td>
<td>.011</td>
</tr>
<tr>
<td>Predicted Initial Estimates</td>
<td>13.257</td>
<td>.000*</td>
<td>1.923</td>
<td>.169</td>
<td>10.363</td>
<td>.000**</td>
</tr>
<tr>
<td>Adjusted Estimates</td>
<td>.617</td>
<td>.542</td>
<td>1.482</td>
<td>.227</td>
<td>1.584</td>
<td>.211</td>
</tr>
<tr>
<td>Numerical Subsection</td>
<td>13.976</td>
<td>.000*</td>
<td>.382</td>
<td>.538</td>
<td>6.424</td>
<td>.002*</td>
</tr>
<tr>
<td>Application Subsection</td>
<td>.393</td>
<td>.676</td>
<td>.561</td>
<td>.456</td>
<td>.454</td>
<td>.636</td>
</tr>
<tr>
<td>No Response Items</td>
<td>.881</td>
<td>.418</td>
<td>1.095</td>
<td>.298</td>
<td>.265</td>
<td>.768</td>
</tr>
</tbody>
</table>

*p < .01   **p < .001

Total score and multiplication subscale score provided two major sources for comparing subject’s performance in terms of general estimation ability as measured by the ACE posttest. Descriptive statistics for these measures are presented in Tables 45 and 46 in Appendix E.

A two-factor analysis of variance (ANOVA) was performed on the ACE posttest data to determine what differences existed between the three groups. The following hypotheses were tested and the results are presented in Table 47.
HO: 1 There will be no difference between the experimental and control groups on the ACE posttest in:

a. total score or
b. multiplication subscale score

HO: 2 There will be no direct effects between gender, or interactional effects between gender and instructional treatments on the ACE posttest in:

a. total score or
b. multiplication subscale score

Results from this analysis indicated that no significant differences existed between treatment group, gender, or interactions between treatment group and gender. Hypotheses HO:1 and HO:1 can not be rejected at the .001 level of significance, therefore, the groups can be considered to be similar.

The AE-M was designed to assess student's ability to estimate the results of multiplication exercises with greater specificity. Total score, multiplication subscale score, and the number of predicted initial and adjusted estimates were of major interest in comparing the three groups.

Descriptive statistics for total score, multiplication subscale, and predicted initial estimates are presented below in Tables 14, 15, and 16 respectively. Descriptive statistics for the adjusted estimates appear in Table 48.
### TABLE 14
MEANS AND STANDARD DEVIATIONS FOR THE AE-M POSTTEST TOTAL SCORE

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Gender</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
<th>Gender</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Males</td>
<td></td>
<td></td>
<td></td>
<td>Females</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>14</td>
<td>males</td>
<td>18.07</td>
<td>2.586</td>
<td>18</td>
<td>females</td>
<td>15.84</td>
<td>3.517</td>
<td>32</td>
<td>16.56</td>
<td>3.379</td>
</tr>
<tr>
<td>T2</td>
<td>16</td>
<td>males</td>
<td>14.56</td>
<td>5.341</td>
<td>13</td>
<td>females</td>
<td>18.69</td>
<td>5.528</td>
<td>29</td>
<td>16.41</td>
<td>5.723</td>
</tr>
<tr>
<td>T3</td>
<td>19</td>
<td>males</td>
<td>13.42</td>
<td>3.453</td>
<td>20</td>
<td>females</td>
<td>11.95</td>
<td>5.165</td>
<td>39</td>
<td>12.67</td>
<td>4.421</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>males</td>
<td>15.12</td>
<td>4.357</td>
<td>51</td>
<td>females</td>
<td>14.88</td>
<td>5.384</td>
<td>100</td>
<td>15.00</td>
<td>4.885</td>
</tr>
</tbody>
</table>

### TABLE 15
MEANS AND STANDARD DEVIATIONS FOR THE AE-M POSTTEST MULTIPLICATION SUBSCALE

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Gender</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
<th>Gender</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Males</td>
<td></td>
<td></td>
<td></td>
<td>Females</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>14</td>
<td>males</td>
<td>11.93</td>
<td>2.165</td>
<td>18</td>
<td>females</td>
<td>10.00</td>
<td>3.068</td>
<td>32</td>
<td>10.84</td>
<td>2.841</td>
</tr>
<tr>
<td>T2</td>
<td>16</td>
<td>males</td>
<td>9.31</td>
<td>4.094</td>
<td>13</td>
<td>females</td>
<td>12.31</td>
<td>3.250</td>
<td>29</td>
<td>10.66</td>
<td>3.976</td>
</tr>
<tr>
<td>T3</td>
<td>19</td>
<td>males</td>
<td>7.74</td>
<td>3.331</td>
<td>20</td>
<td>females</td>
<td>6.05</td>
<td>4.347</td>
<td>39</td>
<td>6.87</td>
<td>3.928</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>males</td>
<td>9.45</td>
<td>3.692</td>
<td>51</td>
<td>females</td>
<td>9.04</td>
<td>4.427</td>
<td>100</td>
<td>9.24</td>
<td>4.068</td>
</tr>
</tbody>
</table>
TABLE 16
MEANS AND STANDARD DEVIATIONS FOR THE AE–M POSTTEST PREDICTED INITIAL ESTIMATES

Gender

| Group | Males | | | Females | | | Total | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|       | N     | Mean  | SD    | N     | Mean  | SD    | N     | Mean  | SD    |
| T1    | 14    | 7.21  | 2.607 | 18    | 4.06  | 2.287 | 32    | 5.44  | 2.873 |
| T2    | 16    | 4.31  | 3.071 | 13    | 6.92  | 2.722 | 29    | 5.48  | 3.158 |
| T3    | 19    | 3.74  | 2.130 | 20    | 2.15  | 2.455 | 39    | 2.92  | 2.410 |
| Total | 49    | 4.92  | 2.950 | 51    | 4.04  | 3.072 | 100   | 4.47  | 3.030 |

A two-factor ANOVA was performed on the posttest data for these measures to test the following hypotheses:

HO: 3 There will be no difference between the experimental and control groups on the AE–M posttest in:

a. total score;
b. multiplication subscale score;
c. Number of predicted rounded responses on the multiplication subscale; or
d. Number of adjusted responses on the multiplication subscale.

HO: 4 There will be no direct effects between gender, or interactional effects between gender and instructional treatments on the AE–M posttest in:

a. total score;
b. multiplication subscale score;
c. Number of predicted rounded responses on the multiplication subscale; or
d. Number of adjusted responses on the multiplication subscale.
The results from these analyses are shown in Table 17 and 18.

**TABLE 17**

TWO-FACTOR ANALYSIS OF VARIANCE FOR THE AE-M POSTTEST TOTAL SCORE AND MULTIPLICATION SUBSCALE SCORE BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (A)</td>
<td>2</td>
<td>188.92</td>
<td>9.79</td>
<td>.000**</td>
<td>189.18</td>
<td>15.36</td>
<td>.000**</td>
</tr>
<tr>
<td>Gender (B)</td>
<td>1</td>
<td>.00</td>
<td>.00</td>
<td>.993</td>
<td>1.04</td>
<td>.085</td>
<td>.772</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>99.35</td>
<td>5.15</td>
<td>.008</td>
<td>58.61</td>
<td>4.758</td>
<td>.011</td>
</tr>
<tr>
<td>Error</td>
<td>94</td>
<td>19.29</td>
<td></td>
<td></td>
<td>12.317</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**p < .01

**TABLE 18**

TWO-FACTOR ANALYSIS OF VARIANCE FOR THE AE-M POSTTEST PREDICTED INITIAL ESTIMATES ON THE MULTIPLICATION SUBSCALE BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (A)</td>
<td>2</td>
<td>85.17</td>
<td>13.257</td>
<td>.000**</td>
</tr>
<tr>
<td>Gender (B)</td>
<td>1</td>
<td>12.35</td>
<td>1.923</td>
<td>.169</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>66.57</td>
<td>10.363</td>
<td>.000**</td>
</tr>
<tr>
<td>Error</td>
<td>94</td>
<td>6.42</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**p < .001

A review of the results shown in Tables 17 and 18 indicated that significant differences existed between the treatment groups and the control group with respect to total score, multiplication subscale score, and the number of predicted initial estimates generated for items on the multiplication subscale. Hypotheses
3a, 3b, and 3c, therefore, were rejected at the .001 level of significance. Table 49 indicates, however, that no differences existed between the three groups with respect to the number of adjusted estimates generated on the multiplication subscale. Therefore, hypothesis 3d can not be rejected at the .001 level.

Tables 17, 18, and 49 also indicated that no differences existed between gender or between the interaction of treatment group and gender for total score, multiplication subscale, or the number of adjusted estimates generated on the multiplication subscale. Therefore, 4a, 4b, and 4d can not be rejected at the .001 level of significance. Interactional differences were evidenced, however, regarding the number of Initial estimates generated on the multiplication subscale. Therefore, hypothesis 4c was rejected at the .001 level of significance for interactional effects. These results indicated that the groups did differ significantly in terms of ability to generate estimates for multiplication exercises.

Performance on the numerical and application subsections and the addition, subtraction, division, and multiple operations subscales of the ACE posttest were of additional interest as assessment of these measures helped to complete the evaluation of students' general mental computation and estimation abilities. Descriptive statistics for these measures appear in Appendix E as Tables 50 through 55.

A two-factor ANOVA was performed on the posttest data for each of the six measures to test the following hypotheses.
**HO: 5** There will be no difference between the experimental and control groups on the ACE posttest in:

- a. Numerical;
- b. Application; or
- c. Operation subscale scores.

**HO: 6** There will be no direct effects between gender, or interactional effects between gender and instructional treatments on the ACE posttest in:

- a. Numerical;
- b. Application; or
- c. Operation subscale scores.

The results of these analyses, presented in Tables 56, 57, and 58, indicated that no significant differences existed between treatment groups, gender, or interactions between treatment group and gender. Therefore hypotheses HO: 5 and HO: 6 can not be rejected at the .001 level of significance.

The numerical and application subsections and number of items that were not responded to on the AE-M posttest provided additional information critical to understanding the nature of the impact of the instructional activities. Descriptive statistics for the numerical subsection are reported below in Table 19. Descriptive statistics for the application subsection and items with no response appear in Appendix E—Tables 59 and 60 respectively.
### TABLE 19
MEANS AND STANDARD DEVIATIONS FOR THE AE-M POSTTEST
NUMERICAL SUBSCALE

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>13.21</td>
<td>2.424</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>10.56</td>
<td>3.794</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>10.90</td>
<td>3.423</td>
</tr>
</tbody>
</table>

A two-factor ANOVA was performed on the posttest data for these measures to test the following hypotheses:

**HO: 7** There will be no difference between the experimental and control groups on the AE-M posttest in:

a. Numerical subscale score;  
b. Application subscale score; or  
c. Total number of items with no response.

**HO: 8** There will be no direct effects between gender, or interactional effects between gender and instructional treatments on the AE-M posttest in:

a. Numerical subscale score;  
b. Application subscale score; or  
c. Total number of items with no response.
TABLE 20

TWO-FACTOR ANALYSIS OF VARIANCE FOR THE AE–M POSTTEST NUMERICAL SUBSECTION SCORES BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (A)</td>
<td>2</td>
<td>157.25</td>
<td>13.976</td>
<td>.000**</td>
</tr>
<tr>
<td>Gender (B)</td>
<td>1</td>
<td>4.30</td>
<td>.382</td>
<td>.538</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>72.28</td>
<td>6.424</td>
<td>.002</td>
</tr>
<tr>
<td>Error</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**p < .001

The results from the analysis of the application subsection and items with no response (see Table 61 in Appendix E) indicated that no differences existed between the treatment groups, gender, or interaction between treatment group and gender. Therefore, hypotheses 7b, 7c, 8b, and 8c can not be rejected at the .001 level. Additionally, the results shown in Table 20 indicated that no differences existed between gender or interaction between treatment group and gender on the numerical subsection. Hypothesis 8a, therefore, can not be rejected at the .001 level of significance. Differences were evidenced in Table 20 between the treatment groups on the numerical subsection. Therefore, hypothesis 7a was rejected at the .001 level.

Summary of Findings

Pretest

Two significant differences (p < .001) were revealed for two hypotheses regarding students' performance on the AE–M pretest. A significant difference was indicated between females and males for the number of adjusted estimates
generated on the AE–M multiplication subscale, and a significant difference was found to exist between the experimental and control groups for the number of items with no response on the AE–M pretest.

**Posttest**

No significant differences were found to exist between groups, gender, or interactions between group and gender on any of the subsections or subscales for the ACE posttest. Significant differences (p < .001) were revealed, however, between experimental and control groups on the AE–M posttest for total score, multiplication subscale score, and predicted initial estimates. Moreover, the analysis of AE–M posttest data also indicated a significant interaction between groups and gender with respect to the number of initial estimates generated on the multiplication subscale.

**Post Hoc Analyses**

The results of analyses performed on the posttest data revealed that the experimental groups (T1 and T2) differed significantly from the control group (T3) with respect to total score, multiplication subscale score, number of predicted initial estimates, and numerical subsection score on the AE–M test. These findings suggested that a real difference existed between the experimental and control groups as a result of influences attributable to the instructional treatment, although they provide little insight regarding how much (or how little) students’ estimating changed.

A number of researchers have noted that a real difference is not necessarily an important or meaningful difference (Gay, 1981; Kennedy, 1979). In order to
address the question of a meaningful difference, several analyses were conducted using paired t-tests to compare the pre and posttest scores of students in the three groups. Only 87 students completed both pre and posttest forms of the AE–M test. Therefore the paired t-tests were performed using data for students who had completed both forms of the AE–M test.

The results of these comparisons appear below in Table 21. The label “Mean differences” is used to refer to the mean of the differences of the pretest/posttest pairs for each individual for each measure. Following the logic from the previous discussion regarding significance levels, the .001 level of significance was selected as a more conservative standard as fifteen separate comparisons were made in evaluating the paired t-tests for the five different measures.

<table>
<thead>
<tr>
<th>Group:</th>
<th>T1 N=28</th>
<th>T2 N=25</th>
<th>T3 N=34</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Differences t   p</td>
<td>Mean Differences t   p</td>
<td>Mean Differences t   p</td>
</tr>
<tr>
<td>AE–M Measure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Score</td>
<td>5.75   5.83 .000**</td>
<td>5.28   5.06 .000**</td>
<td>1.35   1.88 .069</td>
</tr>
<tr>
<td>Multiplication Subscale</td>
<td>4.61 5.14 .000**</td>
<td>3.72 4.32 .000**</td>
<td>.65 1.15 .260</td>
</tr>
<tr>
<td>Predicted Initial</td>
<td>2.29 3.01 .005*</td>
<td>1.12 1.44 .164</td>
<td>-1.32 -2.57 .015</td>
</tr>
<tr>
<td>Numerical Subsection</td>
<td>4.25 5.37 .000**</td>
<td>3.80 3.83 .001**</td>
<td>.65 1.07 .293</td>
</tr>
<tr>
<td>No response items</td>
<td>-4.39 -3.59 .001**</td>
<td>-5.92 -4.01 .000**</td>
<td>.77 1.22 .233</td>
</tr>
</tbody>
</table>

* p < .01  ** p < .001

The results of these comparisons indicated significant improvement in the total, multiplication subscale, and numerical subsection scores for the two experimental groups (T1 and T2) between the pretest and posttest. Additionally, the
comparisons indicated a significant increase in the use of predicted initial estimates for group T1.

**Instructional Influence and Ability**

An additional post hoc analysis was conducted to determine if the instructional had a differential influence on students of differing abilities. Students in the two treatment groups were placed in one of three categories of ability (good, fair, or poor) on the basis of total score on the ACE pretest— the range for each category was poor (0-8), fair (9-16), and good (17-34). A series of nine paired t-tests were performed to compare the pretest and posttest means of these estimators for total score, multiplication subscale score, and number of initial estimates generated for the multiplication subscale.

The means and standard deviations for each group is presented in Table 22, and the paired t-test comparisons are presented in Table 23.

**TABLE 22**

**MEANS AND STANDARD DEVIATIONS FOR GOOD, FAIR, AND POOR ESTIMATORS ON SELECTED AE-M MEASURES**

<table>
<thead>
<tr>
<th></th>
<th>Poor</th>
<th>Fair</th>
<th>Good</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Total Score</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>14</td>
<td>6.21</td>
<td>5.147</td>
</tr>
<tr>
<td>Post</td>
<td>14</td>
<td>13.79</td>
<td>4.677</td>
</tr>
<tr>
<td><strong>Multiplication Subscale</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>14</td>
<td>13.14</td>
<td>3.860</td>
</tr>
<tr>
<td>Post</td>
<td>14</td>
<td>9.29</td>
<td>3.688</td>
</tr>
<tr>
<td><strong>Initial Estimates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre</td>
<td>14</td>
<td>1.57</td>
<td>2.409</td>
</tr>
<tr>
<td>Post</td>
<td>14</td>
<td>4.43</td>
<td>3.081</td>
</tr>
</tbody>
</table>
TABLE 23
A SUMMARY OF PAIRED T-TESTS FOR GOOD, FAIR, AND POOR
ESTIMATORS IN THE TWO EXPERIMENTAL GROUPS
ON SELECTED MEASURES OF THE AE-M INSTRUMENT

<table>
<thead>
<tr>
<th>Group:</th>
<th>Poor N=15</th>
<th>Fair N=18</th>
<th>Good N=17</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>t</td>
<td>p</td>
</tr>
<tr>
<td>Total Score</td>
<td>7.46</td>
<td>4.27</td>
<td>.002**</td>
</tr>
<tr>
<td>Multiplication Subscale</td>
<td>6.27</td>
<td>4.47</td>
<td>.001**</td>
</tr>
<tr>
<td>Predicted Initial Estimates</td>
<td>3.00</td>
<td>2.23</td>
<td>.050</td>
</tr>
</tbody>
</table>

* p < .002 ** p < .001

The results of these analyses revealed that a significant difference (p < .002) existed for all levels of ability for total score on the AE-M. Comparisons of performance on the multiplication subscale indicated a significant improvement between the two tests for both fair and poor estimators.

Gender Differences

Past research has consistently revealed profound differences in estimation ability between females and males across a wide range of age, grade, and ability groups. Although no salient, uniform differences were observed between females and males in this study, the results indicated significant differences (p < .01) for the interaction between treatment group and gender for three measures of the AE-M instrument, including total score, predicted initial estimates, and numerical subsection score. Graphs of pretest and posttest mean scores by group and gender appear below in Figures 14 and 15 for all three measures. In all three cases (posttest) the interactions were found to be disordinal, indicating that no
uniform difference existed between males and females. For reasons that can not be determined, estimation ability varied between groups on the basis of gender.

Figure 14  Graphs of Means by Treatment Group and Gender
Figure 15  Graphs of Means by Treatment Group and Gender for the AE-M Numerical Subsection
CHAPTER V
DISCUSSION

Introduction

This study was organized to investigate the influence of computer-based instructional activities on the acquisition and use of a narrowly defined subset of estimation strategies. The computer-based activities were designed to introduce students to three strategies for rounding both numbers in a two-factor multiplication exercise and provide opportunities for practice in the selection and use of these (or other) strategies with a range of numbers. In addition to the rounding strategies, one experimental group received additional instruction in techniques for adjusting two of the three rounding strategies.

A sample of 149 students in six intact eighth grade general mathematics classrooms was identified from two middle schools in Columbus, Ohio. A fifty-five item test, the Assessing Computational Estimation Test (ACE), was used as a measure of general estimation ability. A second thirty-four item test, the Adjusting Estimates-Multiplication Test (AE-M), was constructed to test specific skills that were addressed in the computer based estimation activities. Parallel forms of both tests were used for pretest and posttest analysis. The tests were administered in booklet form (one item per page) with each item being timed. The scores on the two tests, including the various subsections and subscales, formed the basis for comparing the three groups.

A set of computer-based estimation activities were introduced each day during a five-minute discussion that covered technical aspects of interacting with
the software and expectations for its use (e.g. number of problems to try). The computer-based activities were designed using the first two levels of the three level model of strategy acquisition—introduction, practice, and extended application. Students used the courseware for six thirty-minute periods—twice weekly for three weeks. The strategy introduction lessons included tutorial and practice activities (with tutorial option available) using numerical exercises for the three rounding strategies that were taught. Practice activities were structured to encourage and reward a broad range of strategies for generating and refining estimates, and provided a context for student's estimates which was used to provide feedback regarding the quality and directionality of estimates.

**Pretest Findings**

An analysis of the pretest data was conducted to assess existing levels of computational estimation ability and to determine what differences existed, if any, between groups prior to beginning instructional treatments. These analyses indicated that the three groups were generally equivalent across all major areas of interest and most minor ones as well, although several surprises were revealed.

**Major Interests**

The total score on the Assessing Computational Estimation (ACE) test is an indicator of general computational estimation ability. Performance on the ACE pretest was found to be poor across all groups with students responding within the preestablished range of reasonableness to an average of 13.15, or 24% of the 55 items. The analysis of the AE-M pretest yielded a similar finding. Student's
performance on this pretest was also found to be poor across all groups with students responding reasonably to 10.57, or 31% of the 34 items.

Both pretests (ACE and AE-M) are comprised of two subsections—numerical, or computation, and applied. The total scores for all groups were found to be 5.8 and 7.25 respectively for the two subsections of the ACE pretest, indicating that the applied subsection was a little easier than the numerical subsection. This finding supports a similar finding by R. Reys et al. (1980) which indicated that subjects performed better on applied items than on numerical ones.

Student's performance on the multiplication subscale of both pretests was of particular interest in this investigation as instruction focuses on strategies for estimating products. No differences were found between groups, between gender, or interactions between treatment and gender. However, the analysis of this subscale did reveal the presence of some ability to estimate the products of multiplication exercises. Responses on the this subscale accounted for approximately 41% of student's total score on the ACE and about 57% on the AE-M pretest. Additionally, more than one-half (60%) of their responses on the multiplication subscale involved the use of the appropriate rounding strategy— the strategy indicated by the numerical context of the item. A review of the eighth grade mathematics text used by the Columbus Public Schools (Addison Wesley Mathematics, 1985) revealed that seven lessons involving estimation with whole numbers and decimals were included in the text— three of which were devoted to teaching strategies for estimating the product of multiplication problems. Discussions with participating teachers indicated that both had covered all seven of the estimation lessons. It was refreshing to discover this degree of integration as past reviews have shown little or no estimation was included in most textbooks.
Minor Interests

The first surprise revealed by these analyses was that no differences existed between males and females on the ACE pretest and only one was found to exist for the AE-M pretest. Several researchers have previously reported profound gender differences on a variety of tests of estimation skill, with males outperforming their counterparts. The sole statistical difference that existed between females and males on the pretests indicated that females generated more adjusted estimates on the 17-item open-ended multiplication subscale. Table 24 indicates the mean number of adjustments and percentage of use on the multiplication subscale for females and males by treatment group. This table is presented to illustrate the low level of use evidenced despite the statistically significant finding.

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>%</th>
<th>Male</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>1.63</td>
<td>(10)</td>
<td>.50</td>
<td>(3.0)</td>
</tr>
<tr>
<td>T2</td>
<td>1.35</td>
<td>(8)</td>
<td>.25</td>
<td>(1.5)</td>
</tr>
<tr>
<td>T3</td>
<td>.61</td>
<td>(4)</td>
<td>.63</td>
<td>(4.0)</td>
</tr>
</tbody>
</table>

The notion of adjusted estimates stems from the compensation strategies that were to be introduced in the instructional activities used with treatment group T1. The two adjustment strategies were keyed to two of the three rounding strategies, which can be used effectively to generate initial estimates for
multiplication exercises. Figure 16 illustrates the strategy for generating an initial estimate (underestimate) and the associated strategy for adjusting it.

\[
\begin{align*}
42 \times 63 &= 2646 \\
2 \times 60 &= 120 \quad \text{Use part of the discarded product} \\
40 \times 60 &= 2400 \\
100 &= \text{Round the adjustment} \\
+2400 &= \text{add it to the initial estimate} \\
2500 &= \text{to get the adjusted estimate}
\end{align*}
\]

Figure 16 Example of Adjusting an Initial Estimate

The procedure that was adopted for scoring adjusted estimates included any response that fell between the predicted initial estimate and the computed answer. In the example in Figure 16 any response falling between 2400 and 2646 (but including neither). This range was adopted because adjustments made to an initial estimate could easily include 100, 120, 200, and 240—a minimum of four adjustments—which would yield adjusted estimates of 2500, 2520, 2600, and 2640. The wide range of potential adjustments allowed complicates the scoring of these items as responses that were generated using other methods, either appropriate or inappropriate, may be included in the total. Students have been observed, for example, to compute the answer (2646) and round the result to reflect an estimate (2600). This is the most likely of the potential inappropriate methods that students could have used to generate an “estimate” that would fall in the range of acceptable adjustments. Regardless of the method used, it is clear that the accounting of adjusted estimates may include results that do not reflect the use of an adjustment strategy. The rather weak level of use that was indicated, and the
A wide range of potential errors raise questions whether any practical difference existed between females and males for this variable.

A significant difference ($f = 11.53, p < .001$) was found to exist between groups on the AE-M pretest regarding the number of items for which students offered no response. Subjects in the two experimental groups were found to differ significantly from the control group on this variable. Subjects did not respond to an average of 9.43 (28%), 11.15 (32%), and 3.76 (11%) respectively for groups T1, T2, and T3. This result is difficult to understand in that there were no differences between the groups in terms of the number of reasonable responses generated as measured by other scales and subscales. The AE-M is a timed test and it is quite possible that a relaxed time interval would produce different results. Other potential sources of variance that were not investigated include affective factors such as perception of item difficulty, willingness to try (self-efficacy), self-confidence, and frustration. Differences between groups for any of these factors could be responsible for this finding.

**Summary Of Pretest Findings**

Although two significant findings were observed regarding minor hypotheses, the groups showed no evidence of differences in general estimation ability or other specific abilities measured by subscales. When significant pretest differences are found, researchers commonly pursue posttest analyses using an analysis of covariance (ANCOVA). The weak result regarding student's use of adjustment strategies, as well as other potential explanations, indicated that no special analyses were necessary based on this finding. Moreover, the finding of differences in the number of item responses had apparently no impact on other factors
studied (e.g. total score etc.). Therefore an analysis of covariance was not pur-
sued for the posttest.

**Posttest Findings**

The computer-based instructional activities were expected to influence learner's acquisition and use of the rounding strategies in several ways. These influences have been referred to previously as major and minor interests. This section will include analysis of both sets of interests, as well as a discussion of certain issues that may have had a moderating effect on the outcome of this study.

**Major Interests**

The courseware focused on the use of the rounding strategies to produce estimates for multiplication exercises. If learning did occur, then the use of those strategies should be reflected in both the multiplication subscale score and total score for both tests (Hypotheses 1–4). In addition to the rounding strategies, treatment group T1 received additional instruction in the use of strategies for adjusting initial estimates. If this instruction were effectively learned, strategy use should be reflected in the responses of student's in the T1 group on the multiplication subscale as well (Hypotheses 3d and 4d). The posttest data were scored and analyzed, and appropriate comparisons were made. The analysis indicated that no differences existed between groups, gender, or interactions for total score or multiplication subscale on the ACE posttest. However, similar investigation of the AE-M posttest data revealed significant differences between
the experimental and control groups for both total score and multiplication subscale score. An analysis of student's performance on the multiplication subscale indicated a significant increase in the use of the rounding strategies, but not in the use of the adjustment strategies. Further comparisons, using paired t-tests, were made to determine the practical significance of these findings. The results of the paired t-test comparisons indicated that students in the experimental groups scored an average of 5 points higher on the AE-M (total score) and an average of 4 points higher on the multiplication subscale. Performance on this subscale accounted for about 80% of the increase in total score for these two groups. Additionally, a comparison of the pretest/posttest means indicated a 50% increase in total scores and 100% for the multiplication subscale scores.

A less convincing result was evidenced regarding student's use of the rounding strategies on the multiplication subscale, where subjects used an averaged of about 1.7, or approximately 30%, more predicted initial estimates on the posttest. Their use of the rounding strategies accounted for about 50% of the score on this subscale. This raises the question of what alternative strategies were used for the other half of the problems. A post-hoc analysis of student's responses was conducted for the experimental groups. The results indicated that students inappropriately employed one of the other two rounding strategies to generate an estimate for 20% of the problems. The strategy that subjects used for the remaining responses judged reasonable, is partially a matter of surmise. Some responses (= 10%) were found to be better than the initial estimates, suggesting the possible use of an adjustment strategy. Errors in mental computation and responses obtained by computing the answer and rounding it to reflect an estimate are both possible explanations for the remaining 20%.
An additional post-hoc analysis was conducted to explore the influence of instructional activities for different ability groups. Students in the two experimental groups were placed in one of three categories—good, fair, or poor—based on their total score on the ACE pretest. Analyses were conducted for AE-M total score, multiplication subscale score, and number of initial estimates using paired t-tests to compare pre/post means for these factors. The results indicated that the instructional activities were most effective for poor estimators, less so for fair estimators, and least effective for good estimators across all three factors.

Poor estimators used more initial estimates which accounted for about one-half (48%) of the increase on the multiplication subscale and their multiplication subscale scores accounted for 84% of the increase in their total score on the AE-M. Fair estimators improved less dramatically than poor estimators in terms of initial estimates used, but these accounted for about 48% of the increase in their multiplication subscale scores which, in turn, accounted for 79% of the increase in their total scores. Good estimators showed significant improvement ($t = 3.74, p < .002$) in total score on the AE-M pretest, of which 53% was attributable to increased performance on the multiplication subscale. However, their use of appropriate initial estimates remained about the same on the posttest.

Although the instructional activities had greater influence with poor estimators, good estimators were still better estimators than fair estimators, and fair estimators remained better than poor estimators across all three measures. The fact that these analyses indicated that poor and fair estimators benefitted most might lead one to conclude that the instructional activities were of little benefit to good estimators. However, such a conclusion is not warranted. The courseware
sequence includes one group of activities that is structured to introduce and provide opportunities for practice with the three rounding strategies. The three remaining activities were designed to accommodate the use of a broader range of strategies as the feedback was more general—informing students of the quality and direction of their estimates with no reference to a particular strategy. Moreover, options to redo problems and to set accuracy limits were incorporated in some activities to encourage and reinforce the development and use of other strategies. If the good estimators in this study were similar to those identified by R. Reys et al. (1980), then it is likely that they possessed a broad range of strategies for generating estimates. Moreover, knowledge of and practice with such strategies may be responsible for the finding that good estimators relied less on rounding strategies on the AE–M posttest.

Although a more pronounced general effect would have been more conclusive, it must be remembered that these tests are administered with fairly strict time limits. While the time limits encourage the use of estimation strategies, they may, in fact, cause students to make more errors in selection of an appropriate rounding strategy, as well as computationally. Other factors that may have influenced these findings will be discussed below.

Minor Interests

The minor interests were pursued to provide as complete a picture of the influence of the instructional activities as is possible with the experimental methodology that was used. No significant differences were found for any comparisons on the ACE posttest for either subsection or the remaining operation
subscales. However, an analysis of the AE-M posttest data did provide two findings worthy of note.

The finding of no differences between groups for the two subsections on the ACE posttest was not surprising in light of similar findings for total and multiplication subscale scores. All groups were shown to do better on the application subsection of this test, a result consistent with the pretest findings. The analysis of the remaining operation subscales (addition, subtraction, division, and multiple operation items) was conducted to investigate the potential transfer of learning to these operations as the rounding strategies can be generalized effectively beyond multiplication to other operations. The results indicated that transfer did not occur as the pre/post means were relatively stable for all groups.

An analysis of the AE-M posttest subsections indicated that the scores of students in the two experimental groups increased significantly for the numerical subsection. Their scores on this subsection accounted for 60% of their total score. The pre/post comparisons of means reflected a 74% increase in reasonable responses, indicating that the instructional activities had a greater influence on student's ability to estimate the results of numerical items. This result is consistent with expectations that were based on the nature of the learning activities used with the two experimental groups.

The differences between experimental and control groups on the numerical subsection of the AE-M appear to be inconsistent with the finding of no difference between groups for the same subsection of the ACE posttest. However, some caution is advised, as several other factors should be weighed in interpreting the finding in this way. The AE-M multiplication subscale contains seventeen open-ended items of which only five appear in the applied section. Two of
the applied items involve the use of decimals (cents) which have been shown to be more difficult for students to estimate. Moreover, the test is timed and the timing was not varied between subsections—following the practice of the authors of the ACE test—to allow for the increased difficulty associated with reading and interpreting application items. Additionally, it is important to remember that the students in the experimental group were exposed to three different strategies, representing an increase in strategy sophistication. An increase in sophistication may, at least initially, require students to spend more time selecting an appropriate strategy. Therefore, the finding that students perform better on the numerical subsection is consistent with the hypothesis that greater performance should be found on this section of the AE-M posttest.

The analysis of the pretest data revealed that students in the two treatment groups responded to significantly fewer items on the AE-M pretest than did students in the control group. A similar analysis conducted on the AE-M posttest revealed no significant differences between the groups. A comparison of pre/post means indicated that individuals in the experimental groups responded to an average of about five more items on the posttest.

Summary of Findings

The significant results that were observed for total and multiplication sub-scale scores and the use of adjusted estimates on the AE-M posttest indicated that the computer-based estimation activities were at least somewhat effective in influencing student’s computational estimation ability. Moreover, the finding of differential effects for students of differing ability indicated that the learning
activities were sensitive to the needs of different individuals—a major goal of the software development group.

That no significant differences were observed between experimental and control groups on the ACE posttest is a source of concern. These findings raise several questions regarding the results noted above. If the estimation activities were effective, as indicated on the AE-M why weren’t those influences evidenced on similar measures for the ACE posttest? The most obvious answer to this question is that the AE-M was designed to be more sensitive to the use of rounding strategies—particularly for multiplication items. An alternative explanation might be that the AE-M is simply a harder test in terms of item difficulty, scoring procedures and length of test. It was noted in Chapter III that the scoring intervals used for judging multiplication items were relaxed for the AE-M to include the possible use of inappropriate rounding strategies, as well as possible adjustments to such inappropriate techniques.

A visual inspection of means for the total and multiplication subscale scores indicated some improvement across all groups for the ACE instrument. The source of this improvement for the experimental groups may be related to instructional influences. However, the increases were approximately the same for all three groups—indicating, perhaps, some growth in ability due to maturity or test taking skills. Continued experience in mathematics and potential fluctuations in cognitive ability might explain improved performance for all groups observed on the ACE posttest. More likely, perhaps, are effects related to the taking of two pretests which may have helped students become better at taking the posttest. Having once experienced the frustration of taking an estimation test may be enough to make a second experience easier.
The results of this investigation support similar findings observed by other researchers. Bestgen et al. (1980) found that "brief systematic" instruction in estimation strategies was beneficial for preservice elementary teachers. In a summary of findings observed by other investigators, Benton (1986) reported that instruction does have beneficial effects across a variety of grades and age ranges. The results of the present investigation support the general conclusion that instructional activities in computational estimation are worthwhile. Damarin et al. (1988) observed that computer-based activities can be used effectively to support the development of computational estimation strategies for a range of high school students. The current investigation lends further support for this finding.

Limitations

It has been suggested that science proceeds by creating experimental conditions that are favorable to the confirmation of ideas and relationships that are the object of investigation and through which "we seek to increase our certainty that [a given] proposition is true" (Krathwohl, 1985). Although this study was conducted in this spirit, several external factors conspired to reduce this investigation to a condition that was much less favorable than desired. The study was originally designed to allow for eight weeks of instruction and two weeks of testing. It was to begin in February and end in April. However, because this time frame overlapped with city-wide eighth grade assessment, the study was postponed until after testing was completed. This forced the duration of the study to be reduced from ten to five weeks, thereby weakening the instructional treatment.
In addition to these problems, a large degree of absenteeism was observed among students in the experimental groups. Approximately 65% of the subjects missed at least one lesson, 50% missed two lessons, and 30% missed three lessons. In most cases the lessons could not be made up because of the inflexible schedule; however, an attempt was made to insure that all students had had some practice with the strategy introduction activities. Additionally, the majority of one class period was interrupted by a fire drill.

It is contended that these "difficulties" served to compromise the conduct of this study, thereby weakening the instructional treatments and casting some doubt on inferences that are drawn from the findings. It is noted, however, that such difficulties do reflect the realities of schooling. Conclusions drawn from the findings of this study are further limited by the sample size and the degree of representativeness of the sample in relation to the broader population of eighth-grade general mathematics students.

The much broader issue—whether quantitative instruments are adequate enough to provide an informed perspective of the strategies and processes that students used to generate estimates—places further limits on generalization of these findings. In scoring student’s responses, it is not possible to determine what processes or strategies were used to reach the solution. One must assume that a given response that is the same as the result that would be obtained by using a rounding strategy really was generated using the rounding strategy and not the result of a computational error or use of another strategy, or for that matter, a combination of strategies.

It must be recognized that no attention was given to other related affective and cognitive factors which may have influenced students in this investigation.
Knowledge of student's mathematics abilities and attitudes toward mathematics, computer-based instruction, and estimation may provide greater insight into the nature of interrelationships between these factors and instructional influences.

Recommendations

This research represents one of a handful of studies that have been conducted to investigate the effects of instruction in the area of computational estimation. Moreover, it is one of only two undertaken to explore the use of computer-based activities. Although the instructional treatments were shown to have a positive influence on student's estimation abilities, many questions remain to be investigated including the duration of each activity, the sequence in which activities were used, the optimum number of instructional sessions, and the level of integration of activities in the classroom.

Thirty minutes of class time were allotted for each of the six instructional sessions. During the latter part of some periods, some students began to lose concentration and, presumably, interest in the activities—indicating that beyond a certain point instruction is no longer effective. Would shorter instructional sessions be more consistent with student's attention span and serve to strengthen the instructional treatments? A variation of this theme— one that deserves serious consideration—would involve allowing students to determine, within some minimum and maximum time limits, how long they want to work on a given activity. The use of computer-based instruction places students in a position of monitoring their own progress, therefore, the responsibility of making decisions regarding how many problems to attempt, the setting of time limits and
difficulty levels is a logical extension of this monitoring process. Admittedly this has important implications regarding self-regulated learning.

In Chapter II it was hypothesized that strategy acquisition should take place across three levels of instruction involving an increasing degree of contextual complexity. The strategy development activities introduced in the Round UP/Round down disk followed a different sequence because of time constraints. Those activities were used to introduce the strategies and provide opportunities to practice them in a numerical context one at a time (introduction level). Would a different sequence strengthen the instructional influences? The three rounding strategies that were taught have many similarities and the introduction and development of all three prior to practice in context may have caused confusion in selection and application in other contexts. An alternative approach would involve introducing each strategy independently, providing opportunities to practice in other low levels of contexts before introducing additional strategies.

The time restrictions placed on this investigation forced the use of fewer instructional periods of greater duration than was thought to be optimal. It was noted that more frequent sessions of shorter duration would serve to strengthen the instructional treatment. Moreover, it was suggested that a different sequence of activities be used. Both of these recommendations indicate that instruction should take place of a longer period of time. A longer time frame, in and of itself, may provide a greater opportunity for students to reflect upon, and integrate these strategies into existing structures.

A previous study of the effectiveness of computer-based estimation activities was organized to integrate instruction into the existing classroom structure (Damarin et al., 1988). This approach may have several significant benefits over
a laboratory-based approach. It is appropriate to acknowledge that there are differing degrees of integration that could be explored. Damarin et al. (1988) placed three computers in the classroom and offered the TABS-Math estimation activities as an alternative to pursuing other lessons. The teacher encouraged students to use the courseware and answered questions regarding its use, but refrained from any other involvement. An alternative approach— one that should be pursued— would involve teachers in other critical aspects of instruction including the introduction and debriefing of estimation activities. Introduction phases would involve both mechanical (how to use the software) and goal setting functions.

Allowing the teacher to become in debriefing functions may provide several benefits as well. Reviewing lessons should help to reveal gaps in student’s understanding, reveal critical processes in their thinking, and facilitate closure on those concepts vis a’ vis goals and expectations for learning.

An additional benefit of teacher involvement relates to the broader goals of estimation. An awareness of, and appreciation for computational estimation, as well as a sense of its usefulness and importance are critical affective goals. A more thorough integration of estimation activities is likely to foster these goals as students are more likely to value estimation if it is addressed with the same degree of seriousness and sensitivity as other topics in mathematics.

**Evaluation**

During the course of this investigation, two instruments were used to measure estimation abilities. The ACE test was used as a measure of general
estimation ability as it provides a balance of items for the four arithmetic operations and includes several multiple-step problems. Moreover, items require students to work with a variety of number forms including whole, decimal, percent, and fractional. The AE-M test was constructed with a focus upon measuring strategies for generating estimates to multiplication exercises. Testing consumed the better part of two class periods and was a source of much frustration for many of the participants. Frustration was an artifact of lengthy testing, students' lack of experience in estimating, and time limits imposed to discourage computation. In retrospect, there are several areas of evaluation that should be changed which would enhance information yield and reduce the strain imposed upon participating students.

A single test integrating both broad and narrow evaluation goals could be constructed by replacing some of the ACE multiplication items with similar items from the AE-M test. A substitution of items would permit quantitative exploration of strategy use in a secondary analysis of these items while still maintaining the broader index of general estimation ability. The use of a single test would also tend to reduce the probability of confounding influences associated with the use of two tests.

In addition to reducing the length of evaluation, a better approach to test administration would eliminate some of the problems observed with the use of test booklets. Many students deviated from instructions in returning to previous, uncompleted items in the booklet or by continuing beyond the current item to work on others. This practice subverts the imposed time limits and may alter the nature of the data collected. Placing the presentation of items in an over-head
format would eliminate "page flipping" and related problems such as losing track of item order and, therefore, increase the certainty that responses were estimated within the allowed time frame and were not computed.

The data from the AE-M test were analyzed from a quantitative perspective to determine specific strategy use (e.g. predicted initial estimates and adjusted estimates). Quantitative exploration of strategy use is based on the assumption that a given strategy is associated with a unique result. However, students may derive the same estimate using a variety of methods. Interviews have been used by some researchers to identify strategies and processes that estimators use in the solution of problems. Conducting interviews with a cross-section of participants would shed additional light on instructional influences in terms of the strategies that students actually use to solve estimation problems during an interview. Ostensibly, data gathered during interviews should facilitate a broader understanding of strategy development and use and motivation for strategy selection, as well as other factors such as errors in selection and mental computation, and understanding of, and sensitivity to limitations of accuracy.

**Evaluating Affective Factors**

Learning is a complex phenomenon that involves both external conditions such as the structure and presentation of content, and internal cognitive and affective conditions of the learner. This inquiry has focused upon cognition in relation to learning strategies for estimating, but has paid little attention to the powerful influences that are associated with learner's internal affective conditions. In this study, previous experiences notwithstanding, there were several affective factors that may have influenced both test performance and learning
including attitudes toward mathematics, estimation, and computers, and aspects of self-development such as locus of control and self-efficacy. An investigation of these important factors should yield a greater understanding of the breadth of instructional influences and may help to explain differential instructional effects based on ability.
APPENDIX A

ESTIMATION COURSEWARE

Introduction

The software used in this study includes the TABS-Math Estimation software and a fifth disk, Adjusting Estimates. The Adjusting Estimates disk was created to provide students with experiences in refining initial estimates. This fifth set of activities was used in conjunction with the TABS-Math Estimation software to create two instructional sequences.

TABS-Math Estimation Software

The TABS-Math Estimation courseware was developed by the TABS-Math project under a federal grant at The Ohio State University. The software was published by the Encyclopaedia Britannica Educational Corporation (1985). The Estimation software was created for use on Apple II series computers and consists of four disks: 1) Round Up/Round down; 2) Bull’s-Eye; 3) At The Races; and 4) Estimation Invasion. The software provides a highly interactive easy to use menu-driven environment which is consistent across the four sets of activities. General instructional features include the randomization of numbers in a “problem,” or exercise, practice estimating in both numerical and application (limited-context) environments, colorful and aesthetically designed screen displays, aspects of learner control (e.g. pace, sequence of instruction, quantity of exercises approached, selection of feedback type, and levels of difficulty), and
timed practice. These features vary in their nature and availability in any given program. For example, learners are allowed to control the sequence of activity, pace of instruction, and quantity of exercises in each of the Round Up/Round down practice activities.

Development of the TABS-Math estimation activities “was guided by a point of view concerning the place and value of computers in the classroom” and was based on “the principle of active learning which suggests that students must be actively involved in their own learning” (Damarin, 1984). The “key elements” of the philosophy that guided the TABS project development team included:

- Courseware materials should be viewed, not as curriculum, but as learning activities that can support and enhance the mathematics curriculum as it evolves in our schools. The materials should lend themselves to integration with other instructional materials according to the plans and style of the teacher and, wherever possible, should be adaptable to local curricula and concerns.

- Courseware development should be based on an analysis of the educational literature, including both “common wisdom” concerning children’s learning of mathematics and research studies.

- Computer-based experiences should help young students understand the relationships between “real world” experience and abstract mathematics.

- Children can and should interact with computers in many ways, but the children should always be in control of the computer.

- Finally, the capabilities of the computer, the nature of the mathematical content, and the computer/learner interactions should shape the materials. Rather than beginning with existing materials and asking how they could be adapted to the computer, the project staff should begin with considerations of content and objective and ask how the computer can bring new understanding to this content. (Damarin, 1984; p. 63-64)
Prerequisite Skills

The major emphasis of these estimation activities is on the conceptual development of the process of selecting the appropriate management strategy and automaticity of mental computation. However, learners are expected to begin to judge the reasonableness of their own estimates based on comparisons with the computed product and the appropriate estimate for each exercise.

In order for a learner to be able to succeed in these activities a few minimum skills are necessary. These skills include facility with arithmetic operations (e.g. addition subtraction, multiplication, and division) and place value concepts. Several supportive prerequisites, including positive attitudes toward mathematics and computers, confidence in mathematical skills, tolerance for inaccuracy, and willingness to estimate, are also important factors contributing to success with these materials.

Content Analysis

A post-hoc analysis of the content of the Round UP/Round down diskette yielded instructional goals that indicated learners should be able to:

1) identify the three categories of numerical context which can occur with two two-digit numbers in a multiplication exercise. The numerical contexts are:

   a. round both numbers down
      \[31 \times 63 \rightarrow 30 \times 60,\]
   
   b. round both numbers up
      \[56 \times 48 \rightarrow 60 \times 40,\] and

   c. round one number up and one number down
      \[28 \times 72 \rightarrow 30 \times 70.\]
2) select the appropriate rounding (management) strategy for a given exercise.

3) complete the estimate by mentally computing the product of the two rounded factors.

4) identify which of two two-factor "exercises" yields the largest product.

5) understand/judge the effects of the rounding process in terms of directionality and accuracy.

6) feel comfortable in handling the various steps which lead to the generation of an estimate.

7) begin to feel comfortable in the learning environment provided by the software.

8) begin to feel comfortable with the limitations of accuracy when estimating.

9) experience some success with respect to objectives 1 through 4.

Task Analysis

In each of the practice activities, the learner is presented with a two-factor exercise (39 X 62, for example) with either multiplication or division as the operation. In order to complete the exercise, the learner must perform the following cognitive operations:

1.1) Identify the numerical operation
1.2) Identify the numerical context and select the strategy to be used.
2.1) Select the first number
2.2) determine the direction for rounding
2.3) round the first number (39 --> 40)
2.4) and store 40;

3.1) Select the second number
3.2) determine the direction for rounding
3.3) round the second number (62 → 60)
3.4) hold 60 in short-term memory;

4.1) Recall the first number (40)
4.2) organize the numbers for ease of multiplication (possible)
4.3) complete the multiplication and hold the factor in short-term memory

5.1) assess the reasonableness (possible)

6.1) type in the estimate

7.1) Analyze the feedback
7.2) identify the appropriate strategy (depicted as thrifty, plenty, or Up/down)
7.3) identify one's own strategy (optional order)
7.4) compare one's own strategy to the appropriate strategy

8.1) assess the outcome (reasonableness).

This analysis suggests that the learner is faced with a great deal of complexity in terms of the number of cognitive "moves" that are required in the selection of the appropriate strategy, mental computation of the estimate, and assessment of the reasonableness and accuracy of the estimate. Moreover, it indicates the need for a set of intellectual skills that are well enough developed to approach such complexity. In order to generate an estimate the learner must possess the ability to understand defined concepts such as round down, and apply single-step rules for rounding a number down, and multiple-step rules for multiplying two rounded factors.
Similarly, in analyzing the reasonableness and accuracy of an estimate the learner must be able, to retrace the steps used to generate the estimate, compare the selected strategy to the preferred strategy for similarity and differences, and draw conclusions as to which strategy yields the best (most accurate) result. If learning to generate an estimate based on the numerical context is to take place, the learner must be able to apply the results of the feedback in subsequent “exercises” to reinforce and refine the selection process. Finally, in interpreting the feedback, the learner must be able to identify instances where the selected strategy deviates from the appropriate strategy, and differentiate between instances where errors in selection, rounding, or computation have occurred.

Round Up/Round down

The Round Up/Round down diskette was developed to provide students with experiences in the rounding of two digit numbers in a two factor multiplication exercise. These activities represent the only source of explicit instruction among the four diskettes. This diskette includes the following activities:

1) Rounding Up - 3 activities: Introduction - Tutorial - Practice
2) Rounding down - 3 activities: Introduction - Tutorial - Practice
4) Rounding Up and down - 3 activities: Introduction - Tutorial - Practice

The activities which help the learner to deal with strategy selection (RU, Rd, and RU/Rd) reflect consideration of the factors outlined above. The exercise screens were designed to present the exercise and information regarding the entry of the estimate in a clear and concise manner. The cursor was used to give
a clue as to which strategy was appropriate for the numerical context. A blinking up arrow was used to indicate contexts which should be rounded up, a down arrow for down, and an alternating up and down arrow blinked to suggest rounding up and down.

**Immediate Feedback**

The feedback screens were designed to "restate" the problem and present a comparison between the estimate resulting from the appropriate strategy and the learner's estimate. Two types of screens were included for the strategy activities, a graphics and a text screen. The feedback screens were designed to include three essential components, the exercise and it's product, the estimated product (based on the appropriate strategy being taught for the numerical context), and the learner's estimate.

On the graphics screen, the appropriate estimate and the learner's estimate were depicted graphically and the learner's attention was drawn first to the appropriate estimate and then to her/his estimate. When the two estimates matched, additional graphics were used to indicate that a match between the two strategies existed. When the learner's estimate failed to match the appropriate estimate, the feedback highlighted the appropriate estimate only and provided clues to the directionality and quality of the estimate generated. On the text screen, similar techniques were employed to draw the learner's attention to first the appropriate estimate and then the learner's estimate. When no match occurred, text was used to describe the estimate, e.g. "you did not use the Thrifty strategy," or "Your estimate was better than the Plenty strategy."
Two other features of the immediate feedback are worth noting as they represent attempts to help learners to grapple with difficulties that they encountered. An option for sound was included in the feedback design, and when it was operational it provided additional clues to the nature of the learner's estimate. An option to review the appropriate strategy was offered to the student at the end of any exercise where the learner's estimate failed to match that of the appropriate strategy. This "mini-tutorial" allowed learners who were unsure of themselves to review the techniques appropriate for that exercise.

**Summative Feedback**

Summative feedback was an option in each of the three strategy oriented activities discussed above. The learner had the option to exit exercises at the beginning or end of an exercise. Upon exit the learner is branched to an activity menu that includes the option to review one's overall performance. This feedback was organized to relate the number of exercises attempted, and the number and percentage of estimates that were either above, the same as, or below the targeted strategy. The learner was allowed to restart the given activity with a clean slate at any time. This option was included to allow learners an increased sense of success once they got the feel of the activity. Moreover, the design group hoped that this would also encourage learners to experiment with different strategies in the three activities.

**Practice Activities**

The estimation courseware included *Bull's-Eye*, *At the Races*, and *Estimation Invasion*, all of which were designed to provide opportunities to practice
generating estimates for multiplication exercises and some division as well. Feedback in these activities was designed to reward and encourage the use of a broad range of strategies. Learners were expected to develop a greater sense of reasonableness as a result of features that were incorporated to reinforce strategies for refining estimates, although such strategies were not explicitly taught.

**Bull's-Eye**

Bull's-Eye is a limited-context practice activity which places the learner in the familiar context of a dart board. Like the *Round Up/Round down* activities, the learner is required to select the appropriate management strategy, round the two factors, mentally compute the estimate, and judge the reasonableness. The accuracy of an estimate is represented through closeness to the bull's-eye. Generally, an estimate which falls within the bull's-eye is very accurate (very good), the first ring accurate (good) and so on.

Bull's-Eye provides three instructional features which are notable extensions of previous experiences with the *Round Up/Round down* activities. First, the learner is allowed to choose the number of digits for each factor in an exercise, thus allowing her/him to transfer skills with two-digit numbers to those required with larger numbers. The skills required for handling larger numbers place a greater emphasis on applying place value concepts. Second, the learner is allowed, and encouraged, to choose division as well as multiplication exercises. Again the student is expected to transfer the rounding skills, this time to a different operation. Practice estimating quotients was included here as a prerequisite for proceeding to the *At The Races* activities, where division constitutes approximately one third of the practice exercises. Third, this activity
provides the learner with an opportunity to explore the limits of accuracy. The learner can control the “level” of accuracy required to reach the bull’s-eye and other rings by specifying the percent difference between the estimate and the computed product or quotient. A default accuracy setting was included to reduce the initial complexity of the activity. Learner’s were encouraged to explore this feature as their sophistication increased.

The setting of accuracy criteria is itself an activity requiring somewhat sophisticated mathematical skill. The learner must have some understanding of percentages and be able to apply those to notions of accuracy which are expressed in the activity as differences from the bull’s-eye, or computed result (product or quotient). On the default setting, for example, a learner’s estimate must be within 10% of the product or quotient to fall in the bull’s-eye.

The Bull’s-Eye activity was structured to provide learners with an opportunity to practice skills as well as to provide experiences which would help to extend the learners understanding of notions of accuracy. For this reason it is approachable by both unsophisticated and sophisticated estimators alike. The Bull’s-Eye activity places the learner in an intuitive context for exploring concepts related to accuracy.

At The Races

The At The Races diskette contains two activities, Swimmer and Lemans? These activities present the learner with distance, rate, and time exercises. Two numbers in the triad are randomly generated and the student is required to supply an estimate of the remaining product or quotient. Like Bull’s-Eye, the learner is placed in a limited-context environment that facilitates the process of
interpreting the accuracy of estimates generated. Learners are given the option to set the level of accuracy that they feel is appropriate to provide a continued challenge. This helps to make the activity accessible to both able and less-able learners, and provides a constant challenge to students who are in the process of refining their estimation skills. A redo option was included in these activities to help learners to adjust to the increased demands related to the identification of arithmetic operations and greater numerical complexity. This option was also viewed as important to providing students with opportunities for refining their estimates and recovering from errors. The refinement process alluded to here, is viewed as an iterative process (successive approximation), one that reinforces experimentation with different techniques for generating estimates by removing the “penalty” of having to accept the results of a single estimate. This differs from the use of the word refine in other contexts that involve numerical adjustments.

Prerequisite skills include previous experience with distance-rate-time problems, whole number and decimal representations, and place value skills. The thoughtful use of the accuracy option also requires prerequisite knowledge of and skill with concepts of percent. The activities include estimation of both products and quotients similar to those addressed in Bull’s-Eye.

Swimmer

Swimmer is an activity that forces learners to address exercises with smaller numbers, including decimals. Estimates of rate are generated in whole numbers, decimals, or mixed numbers. Estimates of time are generated in seconds, or minutes and seconds. And estimates of distance, based on a twenty-five
meter pool, are generated in meters using whole numbers. The problem and feedback screens depict a twenty-five meter pool with seven lanes in a vertical arrangement. The middle lane is occupied by the “local champ,” a concept used to represent a “good” estimate. Overestimates place the learner’s swimmer in one of the top 3 lanes with a good estimate falling in the top lane nearest the “local champ.” Similarly, underestimates place the student’s swimmer in one of the lower three lanes. A very good estimate results in the student’s swimmer beating the “local champ.” While a good estimate results in a tie.

Lemans?

The context for Lemans? is a twenty-four hour automobile race. A race track is depicted on both the exercise and feedback screens. The track is a relative referent as it symbolizes the distance traveled for a given exercise. The race begins when the student enters her/his estimate and the car travels around the course. A gross overestimate results in a crash, while an underestimate results in the car travelling only part way around the track—a percentage of the total distance, derived from the student’s estimate as it compares with the computed result. The student’s car travels completely around the track and receives a checkered flag for a “good” estimate. A very good estimate is depicted by the car traversing the track, receiving the checkered flag, and continuing into the “winner’s circle.”

Estimation Invasion

The two activities, Estimation Invaders and Magic Garden provide learner's with extended practice with numbers ranging from small to large.
These activities are presented in a fast-paced "arcade style" game environment with a fantasy context. Both activities provide timed practice with two-factor multiplication exercises. Each game has four different levels of difficulty that are based on the size of the numbers that must be estimated. Learner's are promoted to the next, more difficult level, by compiling enough points to successfully complete the current level. Points and bonuses increase on successive levels as do the requirements for promotion. These activities are both stimulating and challenging, and provide learner's with opportunities for both success and enjoyment.

*Estimation Invaders*, a take-off of the ever-popular *Space Invaders* game, pits the learner against evil aliens who are attacking the planet Estima. A successful estimate results in the destruction of a percentage of the attackers which is proportionate to the accuracy of the estimate— the greater the accuracy the greater number zapped. Successful over-estimates cause the ship to attack the aliens from above while under-estimates result in attack from below.

*Magic Garden* was developed as a non-violent parallel version of *Estimation Invaders*. It differs in both theme and visual effects, but retains other characteristics such as leveling, points, and timing. In this game the estimator is the protector of the magic garden. A successful estimate transforms some of the evil "Weed Seeds" from weeds into beautiful flowers that grow in the garden. One is promoted to the next level of difficulty upon completion of a garden.

In both games, the first digit of the multiplicand is visible and the remaining digits are covered as each exercise begins. The remaining digits are exposed as the invaders move across and then down the screen. This feature was adopted to highlight the relative importance of the digits in each numeral. The first is the
most significant digit, with the second used to determine the direction for rounding. The remaining digits are used to judge accuracy and make a final adjustment to the rounded estimate (if one is attempted). With the time element imposed, the learner was forced to focus on the front-end of the multiplicand in order to easily complete the estimate. This was designed to reinforce the idea that estimates are meant to be generated quickly and that one need only consider the leading digits to accomplish this with some degree of accuracy.

Each level of these activities was constructed to increase the accuracy required for an estimate to be considered acceptable. The initial level was set at fifteen percent- an estimate had to be within fifteen percent to be considered reasonable. Each level became more difficult in terms of the size of numbers and accuracy. The accuracy element was included to reinforce the need to consider techniques for refining estimates.

Adjusting Estimates

The Adjusting Estimates diskette contains three activities that were created to help students learn to generate more accurate estimates. The learner's attention is drawn, in previous activities, to setting limits of accuracy. In the adjusting activities the focus shifts to constructive strategies for producing estimates of greater accuracy. These strategies are based on the rounding methods that were learned and practiced in previous activities.

In the Finding Parts activity, the student learns that there are two significant residual "parts" that remain after one rounds the two factors. Consider, for example, 74 X 92.
\[ 74 \times 92 = 70 \times 90 = 6300 \]

After 74 is rounded to 70 and 92 is rounded to 90, two quantities remain:

\[ 4 \times 90 \text{ and } 2 \times 70 \]

In an initial estimate the product of the rounded factors is used and these amounts are disregarded. In this activity the learner is asked to identify the larger of the two residual "parts."

In the Using Parts activity the student learns how to make use of the larger "part" to help generate an estimate that is more accurate than an initial estimate would be. To continue the example, the \( 4 \times 90 \) part would be chosen as the largest part, and the multiplication would result in a product of 360. To continue in the spirit of estimating, the learner is asked to round this "part" to 300 and use it to adjust the initial estimate. The product of the initial estimate (6300) is an under-estimate of the actual product (6808). Therefore, the adjustment should be added to the initial estimate to create a more accurate estimate (6600). Similarly, the adjustment for an over-estimate would be subtracted from the initial estimate.

The process of adjusting an initial estimate requires six additional cognitive steps to complete a partially refined estimate. These are:

1. Identify the two significant "parts" \( 4 \times 90 \text{ and } 2 \times 70 \]
2. Identify and select the largest "part" \( 4 \times 90 \]
3. Calculate the product of the largest "part" \( 360 \]
4. Truncate the adjustment \( 300 \]
5. Decide whether to add or subtract the adjustment \( + \]
6. Complete the final estimate \( 6300 + 300 = 6600 \)
This process, although somewhat complex, was adopted because it epitomizes the nature of computational estimation in the sense that an estimate should be quick and fairly easy to generate while maintaining some degree of accuracy.
APPENDIX B

THE ASSESSING COMPUTATIONAL ESTIMATION TEST (ACE)

SAMPLE ITEMS
Pretest Item

3

78,328
81,876
80,137
72,893

+ 87,213

Acceptable Interval 380,000 — 410,000
Computed Value 400,447 Time Allowed 17 sec.
Desrimination Index ? Percent Correct ?

Posttest Item

3

87,419
92,765
90,045
81,974

+ 98,102

Acceptable Interval 430,000 — 460,000
Computed Value 450,305 Time Allowed 17 sec.
Desrimination Index ? Percent Correct ?
Pretest Item

\[ 1 \frac{7}{8} \times 1.19 \times 4 \]

Acceptable Interval 8 - 10
Computed Value 8.925  Time Allowed 17 sec.
Descrimination Index ?  Percent Correct ?

Posttest Item

\[ 2 \frac{5}{6} \times 2.18 \times 3 \]

Acceptable Interval 15 - 20
Computed Value 18.53  Time Allowed 17 sec.
Descrimination Index ?  Percent Correct ?
### Pretest Item 43

<table>
<thead>
<tr>
<th>You owe...</th>
<th>About how much change will I get?</th>
</tr>
</thead>
<tbody>
<tr>
<td>$17.23</td>
<td></td>
</tr>
<tr>
<td>$20 $20</td>
<td>Twenty Dollars</td>
</tr>
</tbody>
</table>

Acceptable Interval $2.50 — $3.00

Computed Value $2.77

Time Allowed 12 sec.

Descrimination Index

Percent Correct

### Posttest Item 43

<table>
<thead>
<tr>
<th>You owe...</th>
<th>About how much change will I get?</th>
</tr>
</thead>
<tbody>
<tr>
<td>$16.34</td>
<td></td>
</tr>
<tr>
<td>$20 $20</td>
<td>Twenty Dollars</td>
</tr>
</tbody>
</table>

Acceptable Interval $3.50 — $4.00

Computed Value $3.66

Time Allowed 12 sec.

Descrimination Index

Percent Correct
Pretest Item

49

I need $3\frac{3}{4}$ yards.

About how much will it cost?

Acceptable Interval $15.00 - 17.50$

Computed Value $16.39$

Time Allowed 12 sec.

Discrimination Index ?

Percent Correct ?

Posttest Item

49

I need $4\frac{3}{4}$ yards.

About how much will it cost?

Acceptable Interval $14.00 - 17.00$

Computed Value $16.44$

Time Allowed 12 sec.

Discrimination Index ?

Percent Correct ?
APPENDIX C

ADJUSTING ESTIMATES–MULTIPLICATION TEST

The 34-item AE–M test was created to assess students’ abilities to estimate the results of multiplication exercises. Two parallel forms were constructed by the investigator and evaluated by a panel of mathematics educators. Both forms were piloted with a sample of seventh and eighth grade mathematics students and the results were used to refine test items, as well as to establish appropriate time limits for the different types of items on the test.

Item Analysis

The StatPack Item Analysis program, developed by the Center for Measurement and Evaluation at the Ohio State University, was used to establish the internal consistency, or reliability, for the AE–M test for both pre and posttest forms using data generated in this study. The ItemA program was used to generate a Kuder-Richardson (KR-20) reliability estimate for each of the two forms. This analysis revealed a KR-20 of .79 for the pretest and .75 for the posttest. A mean item difficulty of .58 was found for the pretest and .54 for the posttest. Mean item discrimination was .383 for the pretest and .356 for the posttest.

Subsections and Subscales

As indicated in Chapter III, the AE–M test consists of two major subsections—numerical and application. A multiplication subscale, consisting of 150
all open-ended multiplication items, was used for an in-depth analysis of student's use of strategies for estimating the results of multiplication exercises. Figure C.1 illustrates the different subsections and subscales that were used.

Test Instructions

Figure C.2 contains a complete listing of the instructions that were used during the administration of the AE–M tests. These were adapted from instructions created by the TABS–Math Lab for use with the parallel forms of the ACE test.

The AE–M Forms

The pages following figure C.2 include a complete listing of both forms of the AE–M test. Acceptable intervals, computed answer, item discrimination factor, and percentage of correct responses is included for each item.
Figure 17 Adjusting Estimates Multiplication Test Organization
Test Instructions for the AE-M Instrument

* Clear your desk except for the test booklet, answer sheet and pen or pencil.

* Fill out the information at the top of your answer sheet.

Instructions

This test is very much like the test that you took yesterday. You are being asked to estimate the answer to several computation exercises. Each exercise will appear on a separate page in the test booklet. You will have a limited amount of time to determine about how much each answer is so it is important that you estimate rather than try to calculate an exact answer.

The first 20 exercises of this estimation test are straight computation problems. You will see addition, subtraction, multiplication, and division problems. Record your estimate on your answer sheet. Be sure to put an X on your answer sheet for any problem you can not complete. DO NOT MAKE ANY MARKS ON YOUR TEST BOOKLET OR ANSWER SHEET.

You will be given anywhere from 12 - 17 seconds to make your estimate depending on the difficulty of the problem. You will have more time to complete harder problems. At the of the allotted time I will say "TIME" after which you will have between 2 and 3 seconds to record your estimate or an X if you cannot make an estimate. After 2 - 3 seconds I will ask you to turn to the next problem. Move promptly to this exercise so that you will not fall behind.
Pretest Item

1

42 X 64

Acceptable Interval 2400 — 3500
Computed Value 2688 Time Allowed 12 sec.
Discrimination Index .61 Percent Correct 68

Posttest Item

1

62 X 44

Acceptable Interval 2400 — 3500
Computed Value 2728 Time Allowed 12 sec.
Discrimination Index .65 Percent Correct 75
### Pretest Item

- **27 x 79**

<table>
<thead>
<tr>
<th>Acceptable Interval</th>
<th>1400 — 2400</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computed Value</td>
<td>2133</td>
</tr>
<tr>
<td>Time Allowed</td>
<td>12 sec.</td>
</tr>
<tr>
<td>Discrimination Index</td>
<td>.790</td>
</tr>
<tr>
<td>Percent Correct</td>
<td>64</td>
</tr>
</tbody>
</table>

### Posttest Item

- **39 x 67**

<table>
<thead>
<tr>
<th>Acceptable Interval</th>
<th>1800 — 2800</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computed Value</td>
<td>2613</td>
</tr>
<tr>
<td>Time Allowed</td>
<td>12 sec.</td>
</tr>
<tr>
<td>Discrimination Index</td>
<td>.600</td>
</tr>
<tr>
<td>Percent Correct</td>
<td>78</td>
</tr>
</tbody>
</table>
### Pretest Item

- **Computed Value**: 259
- **Time Allowed**: 12 sec.
- **Percent Correct**: 57

### Posttest Item

- **Computed Value**: 260
- **Time Allowed**: 12 sec.
- **Percent Correct**: 53
Pretest Item

\[ 84 \times 46 \]

Acceptable Interval \[ 3200 - 4000 \]

Computed Value \[ 3864 \]

Time Allowed \[ 12 \text{ sec.} \]

Discrimination Index \[ .73 \]

Percent Correct \[ 61 \%

Posttest Item

\[ 76 \times 54 \]

Acceptable Interval \[ 3500 - 4800 \]

Computed Value \[ 4104 \]

Time Allowed \[ 12 \text{ sec.} \]

Discrimination Index \[ .70 \]

Percent Correct \[ 70 \% \]
Pretest Item

\[
\begin{array}{c}
5 \\
31 \times 441 \\
31 \times 441 \\
\hline
\text{Acceptable Interval} \quad 12000 - 16000 \\
\text{Computed Value} \quad 13671 \\
\text{Time Allowed} \quad 12 \text{ sec.} \\
\text{Discrimination Index} \quad .40 \\
\text{Percent Correct} \quad 59
\end{array}
\]

Posttest Item

\[
\begin{array}{c}
5 \\
41 \times 332 \\
41 \times 332 \\
\hline
\text{Acceptable Interval} \quad 12000 - 16000 \\
\text{Computed Value} \quad 13612 \\
\text{Time Allowed} \quad 12 \text{ sec.} \\
\text{Discrimination Index} \quad .59 \\
\text{Percent Correct} \quad 61
\end{array}
\]
Pretest Item

\[
47783 - 8678
\]

Acceptable Interval \[36000 - 40000\]

Computed Value \[39105\]

Time Allowed 12 sec.

Discrimination Index .11

Percent Correct 25

Posttest Item

\[
55689 - 7885
\]

Acceptable Interval \[47000 - 50000\]

Computed Value \[47804\]

Time Allowed 12 sec.

Discrimination Index .33

Percent Correct 27
Pretest Item

Which Product is larger?

70 X 70 or 68 X 71

A. ___________  B. ___________

Acceptable Interval ___________

Computed Value ___________ Time Allowed ___________

Discrimination Index ___________ Percent Correct ___________

Posttest Item

Which product is larger?

60 X 50 or 58 X 51

A. ___________  B. ___________

Acceptable Interval ___________

Computed Value ___________ Time Allowed ___________

Discrimination Index ___________ Percent Correct ___________
Pretest Item

8

68 X 189

Acceptable Interval 12000 — 14000
Computed Value 12852
Time Allowed 12 sec.
Discrimination Index .63
Percent Correct 47

Posttest Item

8

77 X 293

Acceptable Interval 14000 — 24000
Computed Value 22561
Time Allowed 12 sec.
Discrimination Index .64
Percent Correct 66
Pretest Item

9

\[ 95 \times 823 \]

Acceptable Interval \[ 72000 - 82300 \]

Computed Value \[ 78185 \] Time Allowed 12 sec.

Discrimination Index \[ .87 \] Percent Correct \[ 41 \]

Posttest Item

9

\[ 85 \times 716 \]

Acceptable Interval \[ 56000 - 72000 \]

Computed Value \[ 60860 \] Time Allowed 12 sec.

Discrimination Index \[ .72 \] Percent Correct \[ 63 \]
Pretest Item

Which product is larger

187 X 34 or 190 X 30

A. B.

Acceptable Interval

Computed Value

Time Allowed 12 sec.

Discrimination Index 0.25

Percent Correct 39

Posttest Item

Which product is larger?

179 X 43 or 180 X 40

A. B.

Acceptable Interval

Computed Value

Time Allowed 12 sec.

Discrimination Index 0.22

Percent Correct 54
Pretest Item

11

153 \times 307

Acceptable Interval \quad 45000 \quad \text{to} \quad 60000

Computed Value \quad 78185 \quad \text{Time Allowed} \quad 12 \text{ sec.}

Desrimination Index \quad 0.67 \quad \text{Percent Correct} \quad 53

Posttest Item

11

162 \times 409

Acceptable Interval \quad 40000 \quad \text{to} \quad 100000

Computed Value \quad 66258 \quad \text{Time Allowed} \quad 12 \text{ sec.}

Desrimination Index \quad 0.70 \quad \text{Percent Correct} \quad 67
Pretest Item

12

486 X 695

Acceptable Interval 280,000 — 350,000
Computed Value 337,700
Time Allowed 12 sec.
Desrimination Index .40
Percent Correct 36

Posttest Item

12

392 X 791

Acceptable Interval 210,000 — 320,000
Computed Value 310.072
Time Allowed 12 sec.
Desrimination Index .74
Percent Correct 51
Pretest Item

Which product is larger?

230 X 460 or 228 X 455

A. B.

Acceptable Interval

Computed Value

Time Allowed 12 sec.

Discrimination Index .46 Percent Correct 41

Posttest Item

Which product is larger?

618 X 84 or 620 X 80

A. B.

Acceptable Interval

Computed Value

Time Allowed 12 sec.

Discrimination Index .19 Percent Correct 77
Pretest Item

14

6979 \times 52

Acceptable Interval \hspace{1cm} \underline{350,000 - 420,000}

Computed Value \hspace{1cm} 362908 \hspace{1cm} Time Allowed \hspace{1cm} 12 \text{ sec.}

Descrimination Index \hspace{1cm} 0.59 \hspace{1cm} Percent Correct \hspace{1cm} 45

Posttest Item

14

7984 \times 63

Acceptable Interval \hspace{1cm} \underline{420,000 - 560,000}

Computed Value \hspace{1cm} 502,992 \hspace{1cm} Time Allowed \hspace{1cm} 12 \text{ sec.}

Descrimination Index \hspace{1cm} 0.67 \hspace{1cm} Percent Correct \hspace{1cm} 64
Pretest Item

15

73 X 3011

Acceptable Interval 210,000 — 240,000
Computed Value 219803 Time Allowed 12 sec.
Descrimination Index .70 Percent Correct 41

Posttest Item

15

64 X 4035

Acceptable Interval 240,000 — 280,000
Computed Value 258240 Time Allowed 12 sec.
Descrimination Index .62 Percent Correct 63
Pretest Item

16

194 X 419

Acceptable Interval 80,000 — 90,000
Computed Value 81286
Time Allowed 12 sec.
Discrimination Index .66
Percent Correct 45

Posttest Item

16

295 X 512

Acceptable Interval 100,000 — 180,000
Computed Value 151,040
Time Allowed 12 sec.
Discrimination Index .65
Percent Correct 63
Pretest Item

\[
248 \times 12 = 7
\]

Acceptable Interval: 360 — 500
Computed Value: 425.143
Time Allowed: 17 sec.
Discrimination Index: -.04
Percent Correct: 17

Posttest Item

\[
197 \times 23 = 6
\]

Acceptable Interval: 750 — 1000
Computed Value: 755.17
Time Allowed: 17 sec.
Discrimination Index: .05
Percent Correct: 4
Pretest Item

18

$4392 \times 78$

Acceptable Interval $32,000 - 40,000$

Computed Value $34,257$  Time Allowed $12$ sec.

Discrimination Index $-0.04$  Percent Correct $11$

Posttest Item

18

$3921 \times 87$

Acceptable Interval $240,000 - 360,000$

Computed Value $286,317$  Time Allowed $12$ sec.

Discrimination Index $0.59$  Percent Correct $66$
Pretest Item

Which product is larger?

\[ 618 \times 84 \quad \text{or} \quad 620 \times 80 \]

A. \hspace{1cm} B.

Acceptable Interval \hspace{1cm} 72000 \quad 82300

Computed Value \hspace{1cm} 78185 \hspace{1cm} Time Allowed \hspace{1cm} 12 \text{ sec.}

Descrimination Index \hspace{1cm} .87 \hspace{1cm} Percent Correct \hspace{1cm} 41

Posttest Item

Which product is larger?

\[ 657 \times 93 \quad \text{or} \quad 660 \times 90 \]

A. \hspace{1cm} B.

Acceptable Interval

Computed Value \hspace{1cm} \hspace{1cm} Time Allowed \hspace{1cm} 12 \text{ sec.}

Descrimination Index \hspace{1cm} .08 \hspace{1cm} Percent Correct \hspace{1cm} 41
Pretest Item

20

3519 / 68

Acceptable Interval 50 — 60
Computed Value 51.75 Time Allowed 17 sec.
Discrimination Index .49 Percent Correct 16

Posttest Item

20

4937 / 72

Acceptable Interval 60 — 80
Computed Value 68.57 Time Allowed 17 sec.
Discrimination Index .23 Percent Correct 21
**Pretest Item**

21

$6.75

$14.48

$127.98

How much for all of these?

Acceptable Interval $145 — $150

Computed Value $149.21

Time Allowed 12 sec.

Discrimination Index .42

Percent Correct 30

**Posttest Item**

21

$3.88

$11.63

$146.92

How much for all of these?

Acceptable Interval $155 — $165

Computed Value 162.43

Time Allowed 12 sec.

Discrimination Index .35

Percent Correct 36
Pretest Item

12 lbs. of onions

37¢ per lb.

About how much for onions?

Acceptable Interval $4.00 — $6.00
Computed Value $4.44 Time Allowed 12 sec.
Descrimination Index .56 Percent Correct 34

Posttest Item

21 lbs. of onions

28¢ per lb.

About how much for onions?

Acceptable Interval $72000 — $82300
Computed Value $78185 Time Allowed 12 sec.
Descrimination Index .87 Percent Correct 41
Pretest Item

30 pencils for $1.76

How much for each one?

Acceptable Interval $0.05 - $0.06

Computed Value $0.0587

Time Allowed 17 sec.

Desrimination Index .41

Percent Correct 23

Posttest Item

30 pencils for $1.55

How much for each one?

Acceptable Interval $0.04 - $0.06

Computed Value $0.052

Time Allowed 12 sec.

Desrimination Index .05

Percent Correct 24
Pretest Item

25 Popsicles per box.

6¢ for each popsicle

About how much for a box of Popsicles?

Acceptable Interval $2.00 — $2.50
Computed Value $2.00

Time Allowed 12 sec.

Discrimination Index .48
Percent Correct 45

Posttest Item

24 Popsicles per box.

9¢ for each popsicle

About how much for a box of Popsicles?

Acceptable Interval $2.00 — $3.00
Computed Value $2.16

Time Allowed 12 sec.

Discrimination Index .05
Percent Correct 55
### Pretest Item

**25**

**About how much was the average speed?**

19 hrs. 13 min.

965 miles

Acceptable Interval 45 m.p.h. — 55 m.p.h.

Computed Value 50.22 m.p.h.  
Time Allowed 17 sec.

Descrimination Index 0  
Percent Correct 0

### Posttest Item

**25**

**About how much was the average speed?**

21 hrs. 16 min.

1032 miles

Acceptable Interval 40 m.p.h. — 50 m.p.h.

Computed Value 48.53 m.p.h.  
Time Allowed 17 sec.

Descrimination Index 15  
Percent Correct 27
Pretest Item

26

$19.97

$158.88

$7.39

How much for all of these?

Acceptable Interval $170 — $188

Computed Value $186.24

Time Allowed 12 sec.

Descrimination Index .54

Percent Correct 49

Posttest Item

26

$29.83

$128.78

$8.88

How much for all of these?

Acceptable Interval $140 — $170

Computed Value $167.49

Time Allowed 12 sec.

Descrimination Index .42

Percent Correct 68
Pretest Item

27

About how many miles traveled?

---

less than 3000 or more than 3000

127 mph avg. speed

23 hrs. 43 min.

Acceptable Interval Greater than 3000

Computed Value 3012.0167 Time Allowed 12 sec.

Descrimination Index 0 Percent Correct 0

Posttest Item

27

About how many miles traveled?

---

less than 3000 or more than 3000

135 mph avg. speed

22 hrs. 47 min.

Acceptable Interval greater than 3000

Computed Value 3005.75 Time Allowed 12 sec.

Descrimination Index 0 Percent Correct 0
Pretest Item

28

$7,528.00

$16,983.00

About how much is the difference in price?

Acceptable Interval $8,000 — $10,000

Computed Value $9,455

Time Allowed 12 sec.

Descrimination Index .12

Percent Correct 65

Posttest Item

28

$9,534.00

$18,872.00

About how much is the difference in price?

Acceptable Interval $9,000 — $10,000

Computed Value $9,338

Time Allowed 12 sec.

Descrimination Index .40

Percent Correct 42
Pretest Item

215 nuts per jar!

About how many nuts in all the jars?

Acceptable Interval _______ 2000 — 4000

Computed Value _______ 3010

Time Allowed _______ 12 sec.

Descrimination Index _______ .66

Percent Correct _______ 56

Posttest Item

412 nuts per jar!

About how many nuts in all the jars?

Acceptable Interval _______ 4000 — 8000

Computed Value _______ 5768

Time Allowed _______ 12 sec.

Descrimination Index _______ .37

Percent Correct _______ 43
Pretest Item

30

18 hrs 27 min

average speed
87 m.p.h.

About how many miles?

Acceptable Interval 1600 miles — 1800 miles

Computed Value 1605.15 miles

Time Allowed 12 sec.

Descrimination Index .11

Percent Correct 5

Posttest Item

30

22 hrs 14 min

average speed
137 m.p.h.

About how many miles?

Acceptable Interval 2200 miles — 3200 miles

Computed Value 3045.97 miles

Time Allowed 12 sec.

Descrimination Index .05

Percent Correct 24
Pretest Item

200 meters
This sprinter can run
9.3 meters per second!

How long will it take?

Acceptable Interval 20 sec — 25 sec
Computed Value 21.505 sec Time Allowed 17 sec.
Discrimination Index .50 Percent Correct 22%

Posttest Item

400 meters
This sprinter can run
9.5 meters per second!

How long will it take?

Acceptable Interval 40 sec — 45 sec
Computed Value 42.11 sec Time Allowed 17 sec.
Discrimination Index .41 Percent Correct 24%
Pretest Item

32

36 lbs.

Sunflower seeds

3150 seeds per lb.

About how many seeds in all?

Acceptable Interval 90,000 — 120,000

Computed Value 119,700 Time Allowed 12 sec.

Discrimination Index .66 Percent Correct 32

Posttest Item

32

46 lbs.

Sunflower seeds

1875 seeds per lb.

About how many seeds in all?

Acceptable Interval 80,000 — 100,000

Computed Value 86,250 Time Allowed 12 sec.

Discrimination Index .02 Percent Correct 5
**Pretest Item**

33

2360 miles
96 gallons of gas used.

How many miles per gallon of gasoline?

Acceptable Interval 23 m.p.g. — 27 m.p.g.

Computed Value 24.79 m.p.g. Time Allowed 17 sec.

Descrimination Index -.02 Percent Correct 1

**Posttest Item**

33

3186 miles
108 gallons of gas used.

How many miles per gallon of gasoline?

Acceptable Interval 28 m.p.g. — 32 m.p.g.

Computed Value 29.5 m.p.g. Time Allowed 17 sec.

Descrimination Index -.05 Percent Correct 5
Pretest Item

34

$47,500

About how much difference in price?

$113,861

Acceptable Interval $60,000 — $70,000

Computed Value $66,361

Time Allowed 12 sec.

Discrimination Index .04

Percent Correct 17

Posttest Item

34

$67,450

About how much difference in price?

$133,732

Acceptable Interval $60,000 — $70,000

Computed Value $66,282

Time Allowed 12 sec.

Discrimination Index -.10

Percent Correct 9
TABLE 25
MEANS AND STANDARD DEVIATIONS FOR THE ACE PRETEST TOTAL SCORE

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>16.38</td>
<td>7.963</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>14.20</td>
<td>7.734</td>
</tr>
</tbody>
</table>

TABLE 26
MEANS AND STANDARD DEVIATIONS FOR THE ACE PRETEST MULTIPLICATION SUBSCALE

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>5.73</td>
<td>3.220</td>
</tr>
<tr>
<td>T2</td>
<td>13</td>
<td>5.46</td>
<td>2.504</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>5.68</td>
<td>3.735</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>5.64</td>
<td>3.194</td>
</tr>
</tbody>
</table>
### TABLE 27
TWO-FACTOR ANALYSIS OF VARIANCE FOR THE ACE PRETEST TOTAL SCORE AND MULTIPLICATION SUBSCALE SCORE BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (A)</td>
<td>2</td>
<td>50.66</td>
<td>.793</td>
<td>.455</td>
<td>2.28</td>
<td>.195</td>
<td>.823</td>
</tr>
<tr>
<td>Gender (B)</td>
<td>1</td>
<td>2.39</td>
<td>.037</td>
<td>.847</td>
<td>.92</td>
<td>.079</td>
<td>.780</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>78.64</td>
<td>1231</td>
<td>296</td>
<td>4.99</td>
<td>.427</td>
<td>.654</td>
</tr>
<tr>
<td>Error</td>
<td>103</td>
<td>63.88</td>
<td></td>
<td></td>
<td>11.69</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 28
MEANS AND STANDARD DEVIATIONS FOR THE AE-M PRETEST TOTAL SCORE

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>14</td>
<td>10.08</td>
<td>18</td>
</tr>
<tr>
<td>T2</td>
<td>16</td>
<td>10.30</td>
<td>13</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>11.43</td>
<td>21</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>10.66</td>
<td>52</td>
</tr>
</tbody>
</table>
TABLE 29
MEANS AND STANDARD DEVIATIONS FOR THE AE-M PRETEST MULTIPLICATION SUBSCALE

<p>| Group | Males | | | | | | Females | | | | | | Total | | |
|-------|-------|----------------|----------------|----------------|----------------|----------------|-------|----------------|----------------|----------------|----------------|----------------|-------|----------------|</p>
<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>14</td>
<td>5.57</td>
<td>4.052</td>
<td>18</td>
<td>7.17</td>
<td>3.714</td>
<td>32</td>
<td>6.47</td>
<td>3.885</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>16</td>
<td>5.94</td>
<td>4.057</td>
<td>13</td>
<td>7.92</td>
<td>4.291</td>
<td>29</td>
<td>6.83</td>
<td>4.210</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>6.55</td>
<td>3.502</td>
<td>21</td>
<td>5.14</td>
<td>3.718</td>
<td>39</td>
<td>5.79</td>
<td>3.643</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>6.06</td>
<td>3.795</td>
<td>52</td>
<td>6.54</td>
<td>3.973</td>
<td>100</td>
<td>6.31</td>
<td>3.876</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 30
TWO-FACTOR ANALYSIS OF VARIANCE FOR THE AE-M PRETEST TOTAL SCORE AND MULTIPLICATION SUBSCALE SCORE BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (A)</td>
<td>2</td>
<td>1.50</td>
<td>.059</td>
<td>.942</td>
<td>9.67</td>
<td>.649</td>
<td>.525</td>
</tr>
<tr>
<td>Gender (B)</td>
<td>1</td>
<td>15.17</td>
<td>.599</td>
<td>.441</td>
<td>12.74</td>
<td>.855</td>
<td>.358</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>25.84</td>
<td>1.021</td>
<td>.364</td>
<td>30.36</td>
<td>2.037</td>
<td>.136</td>
</tr>
<tr>
<td>Error</td>
<td>94</td>
<td>25.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 31
MEANS AND STANDARD DEVIATIONS FOR THE AE-M PRETEST PREDICTED INITIAL ESTIMATES ON THE MULTIPLICATION SUBSCALE

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>14</td>
<td>3.51</td>
<td>3.150</td>
</tr>
<tr>
<td>T2</td>
<td>16</td>
<td>3.30</td>
<td>3.694</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>4.71</td>
<td>3.175</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>3.89</td>
<td>3.341</td>
</tr>
</tbody>
</table>

### TABLE 32
TWO-FACTOR ANALYSIS OF VARIANCE FOR THE AE-M PRETEST PREDICTED INITIAL ESTIMATES ON THE MULTIPLICATION SUBSCALE SCORE BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Predicted Initial Estimates</th>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment (A)</td>
<td>2</td>
<td>7.71</td>
<td>0.714</td>
<td>.493</td>
</tr>
<tr>
<td></td>
<td>Gender (B)</td>
<td>1</td>
<td>.84</td>
<td>0.078</td>
<td>.781</td>
</tr>
<tr>
<td></td>
<td>AB</td>
<td>2</td>
<td>30.61</td>
<td>2.834</td>
<td>.064</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>94</td>
<td>10.80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 33
MEANS AND STANDARD DEVIATIONS FOR THE ACE PRETEST NUMERICAL SUBSECTION

<table>
<thead>
<tr>
<th>Gender</th>
<th>N</th>
<th>Males Mean</th>
<th>SD</th>
<th>Females Mean</th>
<th>SD</th>
<th>Total Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>109</td>
<td>6.19</td>
<td>3.609</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>7.83</td>
<td>3.728</td>
<td>20</td>
<td>5.87</td>
<td>2.221</td>
<td>40</td>
</tr>
<tr>
<td>T2</td>
<td>13</td>
<td>5.62</td>
<td>3.664</td>
<td>17</td>
<td>7.16</td>
<td>4.237</td>
<td>30</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>5.21</td>
<td>3.296</td>
<td>21</td>
<td>5.33</td>
<td>3.916</td>
<td>39</td>
</tr>
</tbody>
</table>

### TABLE 34
MEANS AND STANDARD DEVIATIONS FOR THE ACE PRETEST APPLICATION SUBSECTION SCORE

<table>
<thead>
<tr>
<th>Gender</th>
<th>N</th>
<th>Males Mean</th>
<th>SD</th>
<th>Females Mean</th>
<th>SD</th>
<th>Total Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>109</td>
<td>7.70</td>
<td>5.228</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>8.55</td>
<td>4.846</td>
<td>20</td>
<td>5.87</td>
<td>2.221</td>
<td>40</td>
</tr>
<tr>
<td>T2</td>
<td>13</td>
<td>7.62</td>
<td>4.538</td>
<td>17</td>
<td>8.52</td>
<td>6.755</td>
<td>30</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>7.26</td>
<td>5.363</td>
<td>21</td>
<td>7.29</td>
<td>5.917</td>
<td>39</td>
</tr>
</tbody>
</table>

Total  | 51  | 7.86 | 4.897 | 58  | 7.55 | 5.542 | 109 | 7.70 | 5.228 |
TABLE 35

TWO-FACTOR ANALYSIS OF VARIANCE FOR THE ACE PRETEST NUMERICAL AND APPLICATION SUBSCALE SCORES BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2</td>
<td>25.71</td>
<td>2.041</td>
<td>.135</td>
<td>5.57</td>
<td>.197</td>
<td>.821</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>.24</td>
<td>.019</td>
<td>.891</td>
<td>1.12</td>
<td>.040</td>
<td>.843</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>27.07</td>
<td>2.149</td>
<td>.122</td>
<td>13.46</td>
<td>.476</td>
<td>.623</td>
</tr>
<tr>
<td>Error</td>
<td>103</td>
<td>12.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 36

MEANS AND STANDARD DEVIATIONS FOR THE ACE PRETEST ADDITION SUBSCALE

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Females</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td></td>
<td>20</td>
<td>3.29</td>
<td>1.943</td>
<td>20</td>
<td>2.67</td>
<td>1.325</td>
<td>40</td>
<td>2.98</td>
<td>1.671</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td></td>
<td>13</td>
<td>3.00</td>
<td>1.633</td>
<td>17</td>
<td>3.47</td>
<td>2.023</td>
<td>30</td>
<td>3.27</td>
<td>1.849</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td></td>
<td>18</td>
<td>2.28</td>
<td>1.553</td>
<td>21</td>
<td>2.19</td>
<td>1.601</td>
<td>39</td>
<td>2.23</td>
<td>1.559</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>51</td>
<td>2.86</td>
<td>1.759</td>
<td>58</td>
<td>2.73</td>
<td>1.705</td>
<td>109</td>
<td>2.79</td>
<td>1.723</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 37

MEANS AND STANDARD DEVIATIONS FOR THE ACE PRETEST SUBTRACTION SUBSCALE

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>4.27</td>
<td>2.770</td>
</tr>
<tr>
<td>T2</td>
<td>13</td>
<td>2.62</td>
<td>2.103</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>2.42</td>
<td>1.970</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>3.20</td>
<td>2.463</td>
</tr>
</tbody>
</table>

### TABLE 38

TWO-FACTOR ANALYSIS OF VARIANCE FOR THE ACE PRETEST ADDITION AND SUBTRACTION SUBSCALE SCORES BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2</td>
<td>9.62</td>
<td>3.359</td>
<td>.039</td>
<td>8.10</td>
<td>1.556</td>
<td>.216</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>.16</td>
<td>.056</td>
<td>.814</td>
<td>1.53</td>
<td>.293</td>
<td>.590</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>2.53</td>
<td>.884</td>
<td>.416</td>
<td>13.19</td>
<td>2.533</td>
<td>.084</td>
</tr>
<tr>
<td>Error</td>
<td>103</td>
<td>2.86</td>
<td>5.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 39
MEANS AND STANDARD DEVIATIONS FOR THE ACE PRETEST
DIVISION SUBSCALE

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>2.52</td>
<td>1.544</td>
</tr>
<tr>
<td>T2</td>
<td>13</td>
<td>1.31</td>
<td>1.316</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>1.60</td>
<td>1.363</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>1.89</td>
<td>1.493</td>
</tr>
</tbody>
</table>

### TABLE 40
MEANS AND STANDARD DEVIATIONS FOR THE ACE PRETEST
MULTIPLE OPERATIONS SUBSCALE

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>.57</td>
<td>.793</td>
</tr>
<tr>
<td>T2</td>
<td>13</td>
<td>.85</td>
<td>.899</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>.49</td>
<td>.677</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>.61</td>
<td>.780</td>
</tr>
</tbody>
</table>
### TABLE 41

**TWO-FACTOR ANALYSIS OF VARIANCE FOR THE ACE PRETEST DIVISION AND MULTIPLE OPERATIONS SUBSCALE SCORES BY GROUP AND GENDER**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2</td>
<td>2.32</td>
<td>1.254</td>
<td>.290</td>
<td>.81</td>
<td>1.400</td>
<td>.251</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>1.00</td>
<td>.539</td>
<td>.465</td>
<td>.02</td>
<td>.028</td>
<td>.868</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>5.42</td>
<td>2.929</td>
<td>.058</td>
<td>.17</td>
<td>.289</td>
<td>.749</td>
</tr>
<tr>
<td>Error</td>
<td>103</td>
<td>1.85</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 42

**MEANS AND STANDARD DEVIATIONS FOR THE AE-M PRETEST NUMERICAL SUBSECTION**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>16</td>
<td>7.79</td>
<td>3.429</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>8.15</td>
<td>3.855</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>7.86</td>
<td>3.522</td>
</tr>
</tbody>
</table>
### TABLE 43
MEANS AND STANDARD DEVIATIONS FOR THE AE-M PRETEST APPLICATION SUBSECTION

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>14</td>
<td>2.53</td>
<td>2.816</td>
</tr>
<tr>
<td>T2</td>
<td>16</td>
<td>2.50</td>
<td>2.385</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>3.28</td>
<td>2.287</td>
</tr>
<tr>
<td>Total</td>
<td>48</td>
<td>2.80</td>
<td>2.458</td>
</tr>
</tbody>
</table>

### TABLE 44
TWO-FACTOR ANALYSIS OF VARIANCE FOR THE AE-M PRETEST NUMERICAL AND APPLICATION SUBSECTION SCORES BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (A)</td>
<td>2</td>
<td>5.92</td>
<td>.462</td>
<td>.632</td>
<td>1.85</td>
<td>.392</td>
<td>.677</td>
</tr>
<tr>
<td>Gender (B)</td>
<td>1</td>
<td>19.13</td>
<td>1.492</td>
<td>.225</td>
<td>.23</td>
<td>.049</td>
<td>.826</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>14.57</td>
<td>1.136</td>
<td>.325</td>
<td>1.70</td>
<td>.36</td>
<td>.699</td>
</tr>
<tr>
<td>Error</td>
<td>94</td>
<td>12.82</td>
<td>.472</td>
<td>.698</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## APPENDIX E

### POSTTEST STATISTICAL TABLES

**TABLE 45**

**MEANS AND STANDARD DEVIATIONS FOR THE ACE POSTTEST TOTAL SCORE**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th></th>
<th>Females</th>
<th></th>
<th>Total</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>21.70</td>
<td>10.678</td>
<td>20</td>
<td>18.05</td>
<td>7.857</td>
</tr>
<tr>
<td>T2</td>
<td>13</td>
<td>14.85</td>
<td>8.562</td>
<td>17</td>
<td>19.71</td>
<td>10.629</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>17.38</td>
<td>7.315</td>
<td>21</td>
<td>20.32</td>
<td>6.282</td>
</tr>
</tbody>
</table>

**TABLE 46**

**MEANS AND STANDARD DEVIATIONS FOR THE ACE POSTTEST MULTIPLICATION SUBSCALE SCORE**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th></th>
<th>Females</th>
<th></th>
<th>Total</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>8.40</td>
<td>4.751</td>
<td>20</td>
<td>8.15</td>
<td>3.167</td>
</tr>
<tr>
<td>T2</td>
<td>13</td>
<td>6.69</td>
<td>0.70</td>
<td>17</td>
<td>8.88</td>
<td>4.372</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>6.93</td>
<td>2.834</td>
<td>21</td>
<td>7.75</td>
<td>3.096</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>7.45</td>
<td>3.986</td>
<td>58</td>
<td>8.22</td>
<td>3.505</td>
</tr>
</tbody>
</table>

198
TABLE 47
TWO-FACTOR ANALYSIS OF VARIANCE FOR THE ACE POSTTEST TOTAL SCORE AND MULTIPLICATION SUBSCALE SCORE BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (A)</td>
<td>2</td>
<td>57.38</td>
<td>.765</td>
<td>.468</td>
<td>8.56</td>
<td>.607</td>
<td>.547</td>
</tr>
<tr>
<td>Gender (B)</td>
<td>1</td>
<td>50.96</td>
<td>.679</td>
<td>.412</td>
<td>22.46</td>
<td>1.593</td>
<td>.210</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>181.33</td>
<td>2.418</td>
<td>.094</td>
<td>12.63</td>
<td>.896</td>
<td>.411</td>
</tr>
<tr>
<td>Error</td>
<td>103</td>
<td>75.00</td>
<td></td>
<td></td>
<td>14.10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 48
MEANS AND STANDARD DEVIATIONS FOR THE AE-M POSTTEST ADJUSTED ESTIMATES

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th>Females</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>T1</td>
<td>14</td>
<td>.43</td>
<td>.938</td>
</tr>
<tr>
<td>T2</td>
<td>16</td>
<td>.75</td>
<td>.931</td>
</tr>
<tr>
<td>T3</td>
<td>19</td>
<td>.47</td>
<td>.697</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>.55</td>
<td>.843</td>
</tr>
</tbody>
</table>
### TABLE 49

**TWO-FACTOR ANALYSIS OF VARIANCE FOR THE AE-M POSTTEST ADJUSTED ESTIMATES ON THE MULTIPLICATION SUBSCALE BY GROUP AND GENDER**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (A)</td>
<td>2</td>
<td>.60</td>
<td>.617</td>
<td>.542</td>
</tr>
<tr>
<td>Gender (B)</td>
<td>1</td>
<td>1.44</td>
<td>1.482</td>
<td>.227</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>1.54</td>
<td>1.584</td>
<td>.211</td>
</tr>
<tr>
<td>Error</td>
<td>94</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 50

**MEANS AND STANDARD DEVIATIONS FOR THE ACE POSTTEST NUMERICAL SUBSECTION**

<table>
<thead>
<tr>
<th>Gender</th>
<th>N</th>
<th>Males Mean</th>
<th>SD</th>
<th>Females Mean</th>
<th>SD</th>
<th>Total Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td></td>
<td>N</td>
<td></td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Males</td>
<td></td>
<td></td>
<td>Females</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>9.80</td>
<td>5.435</td>
<td>20</td>
<td>8.35</td>
<td>3.407</td>
<td>40</td>
</tr>
<tr>
<td>T2</td>
<td>13</td>
<td>5.69</td>
<td>4.535</td>
<td>17</td>
<td>9.35</td>
<td>5.086</td>
<td>30</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>7.86</td>
<td>3.947</td>
<td>21</td>
<td>8.88</td>
<td>3.225</td>
<td>39</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>8.07</td>
<td>4.913</td>
<td>58</td>
<td>8.84</td>
<td>3.866</td>
<td>109</td>
</tr>
</tbody>
</table>
TABLE 51
MEANS AND STANDARD DEVIATIONS FOR THE ACE POSTTEST APPLICATION SUBSECTION

<table>
<thead>
<tr>
<th>Gender</th>
<th>N</th>
<th>Males Mean SD</th>
<th>Females Mean SD</th>
<th>Total Mean SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td>Males</td>
<td>Females</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N   Mean SD</td>
<td>N   Mean SD</td>
<td>N   Mean SD</td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>11.90 7.040</td>
<td>20  9.70 5.090</td>
<td>40  10.80 6.165</td>
</tr>
<tr>
<td>T2</td>
<td>13</td>
<td>9.15 5.145</td>
<td>17 10.35 6.451</td>
<td>30  9.83 5.855</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>9.52 4.022</td>
<td>21 11.45 4.220</td>
<td>39 10.56 4.190</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>10.36 5.681</td>
<td>58 10.52 5.208</td>
<td>109 10.45 5.410</td>
</tr>
</tbody>
</table>

TABLE 52
MEANS AND STANDARD DEVIATIONS FOR THE ACE POSTTEST ADDITION SUBSCALE SCORE

<table>
<thead>
<tr>
<th>Gender</th>
<th>N</th>
<th>Males Mean SD</th>
<th>Females Mean SD</th>
<th>Total Mean SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td>Males</td>
<td>Females</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N   Mean SD</td>
<td>N   Mean SD</td>
<td>N   Mean SD</td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>3.85 1.663</td>
<td>20  3.25 1.482</td>
<td>40  3.55 1.584</td>
</tr>
<tr>
<td>T2</td>
<td>13</td>
<td>2.39 1.805</td>
<td>17  3.53 2.294</td>
<td>30  3.03 2.141</td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>3.24 1.516</td>
<td>21  3.64 1.237</td>
<td>39  3.46 1.369</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>3.26 1.718</td>
<td>58  3.47 1.666</td>
<td>109 3.37 1.686</td>
</tr>
</tbody>
</table>
### TABLE 53

**MEANS AND STANDARD DEVIATIONS FOR THE ACE POSTTEST SUBTRACTION SUBSCALE SCORE**

<table>
<thead>
<tr>
<th>Gender</th>
<th>N</th>
<th>Males</th>
<th>Mean</th>
<th>SD</th>
<th>Females</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>4.70</td>
<td>2.975</td>
<td>20</td>
<td>3.80</td>
<td>2.238</td>
<td>40</td>
<td>4.25</td>
<td>2.638</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>13</td>
<td>3.08</td>
<td>2.139</td>
<td>17</td>
<td>3.82</td>
<td>3.087</td>
<td>30</td>
<td>3.5</td>
<td>2.701</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>4.12</td>
<td>2.518</td>
<td>21</td>
<td>4.77</td>
<td>2.232</td>
<td>39</td>
<td>4.47</td>
<td>2.359</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>4.08</td>
<td>2.652</td>
<td>58</td>
<td>4.16</td>
<td>2.512</td>
<td>109</td>
<td>4.12</td>
<td>2.567</td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 54

**MEANS AND STANDARD DEVIATIONS FOR THE ACE POSTTEST DIVISION SUBSCALE SCORE**

<table>
<thead>
<tr>
<th>Gender</th>
<th>N</th>
<th>Males</th>
<th>Mean</th>
<th>SD</th>
<th>Females</th>
<th>Mean</th>
<th>SD</th>
<th>Total</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>20</td>
<td>3.45</td>
<td>2.282</td>
<td>20</td>
<td>1.90</td>
<td>1.804</td>
<td>40</td>
<td>2.68</td>
<td>2.177</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>13</td>
<td>1.85</td>
<td>1.819</td>
<td>17</td>
<td>2.35</td>
<td>1.656</td>
<td>30</td>
<td>2.13</td>
<td>1.717</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>18</td>
<td>1.91</td>
<td>1.373</td>
<td>21</td>
<td>2.83</td>
<td>1.427</td>
<td>39</td>
<td>2.41</td>
<td>1.459</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>2.50</td>
<td>2.003</td>
<td>58</td>
<td>2.37</td>
<td>1.650</td>
<td>109</td>
<td>2.43</td>
<td>1.816</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 55

MEANS AND STANDARD DEVIATIONS FOR THE ACE POSTTEST MULTIPLE OPERATIONS SUBSCALE SCORE

<table>
<thead>
<tr>
<th>Gender</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>T1</td>
<td>20</td>
<td>1.30</td>
<td>.979</td>
<td>20</td>
<td>.95</td>
<td>.945</td>
<td>40</td>
<td>1.13</td>
<td>.966</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>13</td>
<td>.85</td>
<td>.801</td>
<td>17</td>
<td>1.12</td>
<td>1.111</td>
<td>30</td>
<td>1.00</td>
<td>.983</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>18</td>
<td>1.17</td>
<td>.785</td>
<td>21</td>
<td>1.34</td>
<td>.854</td>
<td>39</td>
<td>1.26</td>
<td>.817</td>
</tr>
</tbody>
</table>

### TABLE 56

TWO-FACTOR ANALYSIS OF VARIANCE FOR THE ACE POSTTEST NUMERICAL AND APPLICATION SUBSCALE SCORES BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2</td>
<td>20.47</td>
<td>1.101</td>
<td>.336</td>
<td>9.50</td>
<td>.321</td>
<td>.726</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>30.80</td>
<td>1.656</td>
<td>.201</td>
<td>2.52</td>
<td>.085</td>
<td>.771</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>55.68</td>
<td>2.994</td>
<td>.054</td>
<td>46.91</td>
<td>1.585</td>
<td>.210</td>
</tr>
<tr>
<td>Error</td>
<td>103</td>
<td>18.59</td>
<td></td>
<td></td>
<td></td>
<td>29.60</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 57

TWO-FACTOR ANALYSIS OF VARIANCE FOR THE ACE POSTTEST ADDITION AND SUBTRACTION SUBSCALE SCORES BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>2</td>
<td>3.23</td>
<td>1.158</td>
<td>.318</td>
<td>8.93</td>
<td>1.356</td>
<td>.262</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>2.60</td>
<td>.932</td>
<td>.337</td>
<td>.73</td>
<td>.111</td>
<td>.740</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>6.66</td>
<td>2.387</td>
<td>.097</td>
<td>8.01</td>
<td>1.216</td>
<td>.301</td>
</tr>
<tr>
<td>Error</td>
<td>103</td>
<td>2.79</td>
<td>6.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 58

TWO-FACTOR ANALYSIS OF VARIANCE FOR THE ACE POSTTEST DIVISION AND MULTIPLE OPERATIONS SUBSCALES BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Division</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>2</td>
<td>2.86</td>
<td>.927</td>
<td>.399</td>
<td>.63</td>
<td>.745</td>
<td>.477</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>.05</td>
<td>.015</td>
<td>.902</td>
<td>.02</td>
<td>.026</td>
<td>.872</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>16.88</td>
<td>5.485</td>
<td>.006</td>
<td>1.02</td>
<td>1.194</td>
<td>.307</td>
</tr>
<tr>
<td>Error</td>
<td>103</td>
<td>3.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 59
MEANS AND STANDARD DEVIATIONS FOR THE AE-M POSTTEST APPLICATION SUBSECTION

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th></th>
<th></th>
<th>Females</th>
<th></th>
<th></th>
<th>Total</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>14</td>
<td>4.86</td>
<td>2.033</td>
<td>18</td>
<td>3.83</td>
<td>1.978</td>
<td>32</td>
<td>4.28</td>
<td>2.036</td>
</tr>
<tr>
<td>T2</td>
<td>16</td>
<td>4.00</td>
<td>3.011</td>
<td>13</td>
<td>4.15</td>
<td>3.313</td>
<td>29</td>
<td>4.07</td>
<td>3.093</td>
</tr>
<tr>
<td>T3</td>
<td>19</td>
<td>3.95</td>
<td>2.223</td>
<td>20</td>
<td>3.70</td>
<td>2.179</td>
<td>39</td>
<td>3.82</td>
<td>2.175</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>4.22</td>
<td>2.443</td>
<td>51</td>
<td>3.86</td>
<td>2.408</td>
<td>100</td>
<td>4.04</td>
<td>2.420</td>
</tr>
</tbody>
</table>

### TABLE 60
MEANS AND STANDARD DEVIATIONS FOR THE AE-M POSTTEST ITEMS WITH NO RESPONSE

<table>
<thead>
<tr>
<th>Gender</th>
<th>Males</th>
<th></th>
<th></th>
<th>Females</th>
<th></th>
<th></th>
<th>Total</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>14</td>
<td>4.29</td>
<td>3.791</td>
<td>18</td>
<td>4.17</td>
<td>4.048</td>
<td>32</td>
<td>4.22</td>
<td>3.875</td>
</tr>
<tr>
<td>T2</td>
<td>16</td>
<td>6.50</td>
<td>5.060</td>
<td>13</td>
<td>4.62</td>
<td>3.969</td>
<td>29</td>
<td>5.66</td>
<td>4.624</td>
</tr>
<tr>
<td>T3</td>
<td>19</td>
<td>4.63</td>
<td>6.448</td>
<td>20</td>
<td>3.65</td>
<td>3.884</td>
<td>39</td>
<td>4.13</td>
<td>5.242</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>5.14</td>
<td>5.331</td>
<td>51</td>
<td>4.08</td>
<td>3.903</td>
<td>100</td>
<td>4.60</td>
<td>4.665</td>
</tr>
</tbody>
</table>
TABLE 61

TWO-FACTOR ANALYSIS OF VARIANCE FOR THE AB-M POSTTEST APPLICATION SUBSECTION SCORE AND ITEMS WITH NO RESPONSE BY GROUP AND GENDER

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>P</th>
<th>F</th>
<th>P</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment (A)</td>
<td>2</td>
<td>2.37</td>
<td>.393</td>
<td>.676</td>
<td></td>
<td>19.44</td>
<td>.881</td>
<td>.418</td>
</tr>
<tr>
<td>Gender (B)</td>
<td>1</td>
<td>3.83</td>
<td>.561</td>
<td>.456</td>
<td></td>
<td>24.15</td>
<td>1.095</td>
<td>.298</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>2.74</td>
<td>.454</td>
<td>.636</td>
<td></td>
<td>5.85</td>
<td>.265</td>
<td>.768</td>
</tr>
<tr>
<td>Error</td>
<td>106</td>
<td>6.03</td>
<td></td>
<td></td>
<td></td>
<td>22.06</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
LIST OF REFERENCES


Reys, Robert E., Bestgen, Barbara J., Rybolt, James F., and Wyatt, J. Wendell
The identification and characterization of computational estimation pro­
cesses used by in-school pupils and out-of-school adults. Final report, Con­
tact No. NIE-G-79-0088, Washington, D. C. National Institute of Educa­

Rubenstein, Rheta N. (1985) Computational estimation and related mathemati­
cal skills. Journal for Research in Mathematics Education, 1985, 16(2),
pp. 106-119.


Threadgill-Sowder, Judith Computational estimation procedures of school

Timnick, Lois Electronic bullies. Psychology Today, February 1982, pp. 10-
15.

Trafton, Paul R. Estimation and mental arithmetic: Important components of
computation. In Marilyn N. Suydam and Robert E. Reys (Eds.), Devel­
op ing Computational Skills, 1979 Yearbook, Reston VA: National Coun­
cil of Teachers of Mathematics, pp. 196-213.

Trafton, Paul R. Teaching computational estimation: Establishing an estimation
mind-set. In Harold L. Schoen and Marilyn J. Zweng (Eds.), Estimation
and Mental Computation, 1986 Yearbook, Reston VA: National Coun­
cil of Teachers of Mathematics, pp. 16-30.

Usiskin, Zalman. (1986). Reasons for Estimating. In Harold L. Schoen and
Marilyn J. Zweng (Eds.), Estimation and Mental Computation, 1986
1-15.

Warner Books.

Wyatt, James Wendell (1985). A case study survey of computational estimation
processes and notions of reasonableness among ninth grade students.