THE EFFECTS OF COMPUTER ALGEBRA SYSTEMS ON STUDENTS’ ACHIEVEMENT IN MATHEMATICS

A dissertation submitted to the Kent State University College and Graduate School of Education, Health and Human Services in partial fulfillment of the requirements for the degree of Doctor of Philosophy

By
Christopher L. Tokpah

August, 2008
A dissertation written by

Christopher L. Tokpah

B. Sc., University of Liberia, 1989

M.B.A., Kent State University, 2002

Ph.D., Kent State University, 2008

Approved by

Dr. Shawn M. Fitzgerald
Director, Doctoral Dissertation Committee

Dr. Tricia M. Niesz
Member, Doctoral Dissertation Committee

Dr. Karen Swan
Member, Doctoral Dissertation Committee

Accepted by

Dr. Melody Tankersley
Interim Chairperson, Department of Educational Foundations and Special Services

Dr. Daniel F. Mahony
Dean, College and Graduate School of Education, Health, and Human Services
This meta-analysis sought to investigate the overall effectiveness of computer algebra systems (CAS) instruction, in comparison to non-CAS instruction, on students’ achievement in mathematics at pre-college and post-secondary institutions. The study utilized meta-analysis on 31 primary studies (102 effect sizes, N= 7,342) that were retrieved from online research databases and search engines, and explored the extent to which the overall effectiveness of CAS was moderated by various study characteristics.

The overall effect size, 0.38, was significantly different from zero. The mean effect size suggested that a typical student at the 50th percentile of a group taught using non-CAS instruction could experience an increase in performance to the 65th percentile, if that student was taught using CAS instruction. The fail-safe N, Nfs, hinted that 11,749 additional studies with nonsignificant results would be needed to reverse the current finding. Three independent variables (design type, evaluation method, and time) were found to significantly moderate the effect of CAS.

The current results do not predict future trends on the effectiveness of CAS; however, these findings suggest that CAS have the potential to improve learning in the classroom. Regardless of how CAS were used, the current study found that they contributed to a significant increase in students’ performance.
ACKNOWLEDGEMENTS

But for the overwhelming blessings of the Almighty God, and the tremendous support of a group of self-less and dedicated individuals, the completion of this dissertation would not have been possible.

I would like to extend my appreciation to my dissertation advisor, Dr. Shawn Fitzgerald, and also to the other members of my dissertation committee: Dr. Tricia Niesz and Dr. Karen Swan. They all provided me with professional guidance as I wrote this paper.

I express my appreciation to Diane Munson, Veronica Love, Mary Yost, and the staff of the Academic Success Center at Kent State University. Besides editorial suggestions, they inspired and encouraged me to keep moving forward.

I sincerely appreciate the efforts of my brother (Samson), his wife (Helena), Abraham Samukai, Richmond Draper, Peter Karmei, and Dr. Gary Padak for providing me with the unique opportunity to pursue my graduate education at Kent State University. I remain eternally grateful.

I express deepest gratitude to my wife, Sheila, who allowed me to encroach on our quality time to get this task done. Thanks for your patience.

My friend and brother, Mosoka Fallah, members of the Christian Life Center in Kent, and all of my other true friends: thanks for your continuous encouragement and prayers.
Finally, I dedicate this work to my dad, the late Samson S. Tokpah, Sr., and my mom, Mrs. Angeline Tokpah. They taught me at an early age to “learn to sacrifice today’s pleasures for tomorrow’s treasures.”
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
</tr>
<tr>
<td>CHAPTER</td>
</tr>
<tr>
<td>I INTRODUCTION</td>
</tr>
<tr>
<td>Concerns about Students’ Performance in Mathematics</td>
</tr>
<tr>
<td>Technology goes to School</td>
</tr>
<tr>
<td>CAS</td>
</tr>
<tr>
<td>Rationale for the Study</td>
</tr>
<tr>
<td>Purpose of the Study</td>
</tr>
<tr>
<td>Research Questions</td>
</tr>
<tr>
<td>Null Hypotheses</td>
</tr>
<tr>
<td>Limitations</td>
</tr>
<tr>
<td>Delimitations</td>
</tr>
<tr>
<td>Operational Definitions</td>
</tr>
<tr>
<td>Summary</td>
</tr>
<tr>
<td>II LITERATURE REVIEW</td>
</tr>
<tr>
<td>Introduction</td>
</tr>
<tr>
<td>The Math Wars</td>
</tr>
<tr>
<td>The Traditionalists</td>
</tr>
<tr>
<td>The Constructivists</td>
</tr>
<tr>
<td>Teaching Philosophy and the Use of Technology</td>
</tr>
<tr>
<td>Primary Studies on the Effectiveness of CAS</td>
</tr>
<tr>
<td>Limitations of Primary Studies</td>
</tr>
<tr>
<td>Traditional Syntheses of Research</td>
</tr>
<tr>
<td>Narrative Reviews</td>
</tr>
<tr>
<td>Vote Counting</td>
</tr>
<tr>
<td>Combined Significance Tests</td>
</tr>
<tr>
<td>Overview of Meta-analysis</td>
</tr>
<tr>
<td>Advantages of Meta-analysis</td>
</tr>
</tbody>
</table>
Overall Effectiveness of CAS .............................................................. 86
Moderator Variables ........................................................................ 87
Limitations ....................................................................................... 90
Implications ..................................................................................... 92
Conclusion ....................................................................................... 95

APPENDICES ..................................................................................... 96
APPENDIX A .................................................................................. 97
APPENDIX B .................................................................................. 103
APPENDIX C .................................................................................. 106
REFERENCES .................................................................................. 113
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Stem-and-Leaf plot of all effect sizes</td>
<td>67</td>
</tr>
<tr>
<td>2.</td>
<td>Box and whiskers plot of all effect sizes</td>
<td>68</td>
</tr>
<tr>
<td>3.</td>
<td>Box and whiskers plots of effect sizes after removing all extreme values</td>
<td>69</td>
</tr>
<tr>
<td>4.</td>
<td>Overall effect size comparison for CAS and Non-CAS instruction</td>
<td>72</td>
</tr>
<tr>
<td>5.</td>
<td>Box and whiskers plots of effect sizes by type of publication</td>
<td>73</td>
</tr>
<tr>
<td>6.</td>
<td>Box and whiskers plots of effect sizes by usage</td>
<td>74</td>
</tr>
<tr>
<td>7.</td>
<td>Box and whiskers plots of effect sizes by location</td>
<td>75</td>
</tr>
<tr>
<td>8.</td>
<td>Box and whiskers plots of effect sizes by time</td>
<td>77</td>
</tr>
<tr>
<td>9.</td>
<td>Box and whiskers plots of effect sizes by educational level</td>
<td>78</td>
</tr>
<tr>
<td>10.</td>
<td>Box and whiskers plots of effect sizes by study design</td>
<td>80</td>
</tr>
<tr>
<td>11.</td>
<td>Box and whiskers plots of effect sizes evaluation method</td>
<td>81</td>
</tr>
<tr>
<td>12.</td>
<td>Box and whiskers plots of effect sizes by course type</td>
<td>83</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Search Results by Research Database</td>
</tr>
<tr>
<td>2</td>
<td>Study Features by Categories</td>
</tr>
<tr>
<td>3</td>
<td>Number of Effect Sizes per Category of Moderator Variable</td>
</tr>
<tr>
<td>4</td>
<td>Number and Range of Effect Sizes for Independent Studies</td>
</tr>
<tr>
<td>5</td>
<td>Studies Coded by Publication Type, Type of Usage, and Location</td>
</tr>
<tr>
<td>6</td>
<td>Studies Coded by Time, Educational Level, and Course Type</td>
</tr>
<tr>
<td>7</td>
<td>Studies Coded by Design and Evaluation Method</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

This chapter begins with a general description of the concerns of mathematics professionals and policy makers about the performance, or lack thereof, of students in mathematics. It is followed by issues surrounding the use of technology, in general, and Computer Algebra Systems (CAS), in particular, in mathematics classrooms. Subsequent sections discuss the rationale, purpose, research questions, null hypotheses, limitations delimitations, and operational definitions of key words used in the study.

Concerns about Students’ Performance in Mathematics

American students’ performance in mathematics is an issue of paramount concern for educators and policy makers. Recent statistics released by American College Testing (ACT) on students’ math-readiness for college is alarming, at best. ACT (the company which administers ACT exams) defines college math readiness as the level of preparation a student needs to enroll and succeed, without remediation, in a credit bearing math course at a two-year or four-year institution (ACT, 2005).

The testing company found that a student with a math ACT test score of 22 or higher was expected to have at least a 75% chance of passing the first college level math course, with a grade of C or better, on the first attempt (ACT, 2005). Of the 1.2 million students in the United States (US) who took the ACT exam in 2006, only 42% met this benchmark and the percentage was even lower for minorities, as only 11% of African Americans, 23% of Native Americans, and 25% of Hispanics scored at least 22 (ACT, 2006).
Reports from other researchers on mathematics performance are equally appalling. Haycock and Huang (2001) reported that “only about 1 in 12 of all 17-year olds can comfortably do multi-step problem-solving and elementary algebra- a finding that may surprise those who know that 91% of those students took at least one algebra course” (p. 5).

One of the requirements of the No Child Left Behind Act is that the National Center for Educational Statistics (NCES) conduct ongoing assessment of students’ performances in various subjects. In their 2005 report on the National Assessment of Educational Progress (NAEP), the NCES declared that only 36% of 4th graders, 30% of 8th graders, and less than 25% of 12th graders in the US were proficient in math. As would be expected, these rates were lower for minority groups.

In a 2004 study conducted by the United Negro College Fund, 62 students in a high school drop-out recovery program were asked what made them quit high school. Almost all of the respondents replied that the reason was “math” (Viadero, 2005).

The problem of math proficiency in American high schools is felt by the higher educational system as well. Many students who leave high school are required to take at least one remedial math course in college. For example, 56% of African American students, 40% of Hispanic students, and 30% of white students who graduated from high school in the State of Ohio in 2003 were required to do remediation in math (Ohio Board of Regents, 2005).

While students’ math performance in the US is worse than that of many developed countries, these countries are equally concerned about this troubling situation.
Using trends from the NCES results and International Educational Achievement results, Phillips (2007) found that the percentage of eighth grade American students proficient in math was slightly better than in England and Italy, but worse than Australia and Finland.

Even in countries with relatively high math proficiency levels, such as China, researchers are still searching for ways to improve students’ math skills. Approximately 61% of eighth graders in China were proficient in math in 2003 (Phillips, 2007), yet Chang (1996) noted that one of China’s major concerns was “the lack of qualified teachers of mathematics and computer science, the increasing number of high-risk students in the junior high schools and the students’ lack of interest in studying mathematics” (p. 6).

Technology goes to School

The situations described above suggest that there is an ongoing local and international public unease about the decline in students’ mathematical skills. Adding to this worrisome problem is the lack of agreement among math teachers and researchers on what is the best way to deliver the course content. As the technology literacy of youths continues to increase, there are suggestions that technology could be used as a tool to minimize difficulties that students incur in mathematics classrooms.

In their suggested principles and standards, the National Council for the Teachers of Mathematics (NCTM) - an umbrella organization for more than 110,000 mathematics teachers - in the US advocated:

Every school mathematics program should provide students and teachers with access to tools of instructional technology, including appropriate calculators, computers with mathematical software, Internet connectivity, handheld data
collection devices, and sensing probes… They [technology tools] also enhance computational power and provide convenient, accurate, and dynamic drawing, graphing, and computational tools… With such devices, students can extend the range and quality of their mathematical investigations and encounter mathematical ideas in more realistic settings (NCTM, 2003, p. 2).

The NCTM’s call for the use of technology in the classroom is based on the assumption that technology will help improve the learning of mathematical skills without compromising computational skills that, many argue, students need to acquire.

Other organizations also have joined the call for the inclusion of technology in mathematics classrooms, and this call is vibrant in the corridors of institutions that provide training for mathematics instructors. A technical committee for the Association for Mathematics Teachers Education (AMTE) recommended that teacher training programs need to focus on preparing pre-service teachers to incorporate technology in their classrooms to facilitate students’ learning (AMTE, 2005).

Along the same line of reasoning, Niess (2006) proposed that teacher training programs need to:

- allow teacher candidates to explore and learn mathematics using technology in ways that build confidence and understanding of technology and mathematics,
- model appropriate uses of a variety of established and new applications of technology tools to develop a deep understanding of mathematics in varied contexts, help teacher candidates make informed decisions about appropriate and effective uses of technology in the teaching and learning of mathematics, and, provide opportunities for teacher candidates to develop and practice teaching lessons that take advantage of the ability of technology to enrich and enhance the learning of mathematics (pp. 197-198).

It has been declared that computers could be used in classrooms to: (a) improve students’ acquisition of basic skills in specific subject areas, (b) reduce the drudgery of learning by blending text with multimedia, (c) broaden curriculum objectives through the
use of simulations to aid in problem-based and collaborative learning, (d) enable teachers
to strengthen their mode of content delivery, and (e) prepare technology literate citizens
for the workplace (Abdullah, 2005; Pierce & Stacey, 2004; Adym, 2002; Heid &
Edwards, 2001; Kutzler, 2000; Norton, McRobbie & Cooper, 2000; Heid, 1997; Phillips,
1995).

Some researchers have used generic terms such as computer-assisted instruction
(CAI), computer-based education (CBE), computer-managed instruction (CMI), and
computer-based instruction (CBI) to describe the use of technology to enhance learning.
In many cases, they use these words interchangeably blurring any difference in meaning.

Others categorized computer applications based on their usage. For example,
Atkinson (1969) and Watson (1972) identified 4 uses of computers in an educational
environment: (a) drill-and-practice: the lesson is presented as in a traditional classroom,
but the computer provides follow-up exercises for reinforcement, (b) tutorial: the
computer presents new lessons and follow-up exercises for reinforcement, (c) dialogue:
the computer presents new lessons, follow-up exercises but also allows the student to
respond in everyday language, and (d) computer-managed instruction: the computer
adapts to the student by selecting appropriate instructional resources and maintaining
record of academic progress.

Taylor (1980) identified three ways computers are used in educational settings: as
a tutor, a tutee and/or a tool. When used as a tutor, the computer adapts to the student by
selecting appropriate instructional resources and maintaining record of academic progress
(similar to Atkinson and Watson’s definition of computer-managed instruction). As a
tutee, the computer receives and executes instructions (in the form of programming language) from the student. As a tool, the computer is used to aid in calculations and graphical displays.

**CAS**

Since the advent of calculators, mathematics professionals continue to debate how calculators should be used (or if they should be used at all) in mathematics classrooms. On one extreme of the debate are those who believe that calculators have no place in the classroom. On the other extreme are those who believe that calculators should be used for all computational problems.

The birth of CAS in the 1970’s, and their subsequent introduction into the classrooms in the 1980’s, have further exacerbated this on-going feud. CAS are computer software packages that perform calculations, symbolic manipulations, and also offer graphical representations (Pierce & Stacey, 2001). CAS can be installed on desktop/laptop computers or are available as portable CAS (PCAS) devices such as graphing calculators.

CAS are being used to provide individualized instruction, offer drill and practice to students, broaden teachers’ presentations through the use of simulations and graphical representations, give students immediate feedback, provide examples to students on how to solve equations, and make mathematics fun and interesting (Aydm, 2005; Arikawe, 1989; Noddings, 1989; Silver, 1987).

Those who recommend the use of technology in mathematics classrooms have hinted that CAS: (a) allow students to do more complicated problems than they would
have otherwise been able to do by hand (Kutzler, 1994; Abdullah, 2006; Allen et al, 1999), (b) ease the amount of symbolic, arithmetic manipulations and routine processes that students have to do and empower them to concentrate on learning concepts and problem solving strategies (Abdullah, 2006; Gleason, 2001; Allen et al, 1999; Heid, 1997), (c) enable students to generate and manipulate symbolic expressions that are too time-consuming and complicated to do by hand, (d) assist students to develop generalized rules for problem-solving, (e) help students develop confidence in problem-solving by reducing computational anxiety, (f) improve students attitude towards mathematics (Heid & Edwards, 2001; Chambers & Sprecher, 1980), (g) make mathematics enjoyable (h) provide immediate feedback to students (Heid, 2002; Heid, 1997), (i) prepare students with skills that are required in the workplace (Gleason, 2001; Allen et al,1999), and (j) aid students to view mathematical concepts concretely (Allen et al, 1999).

It can be contended that Taylor’s (1980) categorization of instructional roles of computers (tutor, tutee, and tool) can be reduced to two roles where CAS are concerned since CAS are generally user-friendly and, in most cases, require no (or very little) programming language knowledge.

Academic Systems software provides a good example of CAS being used in a tutorial mode. When a student logs onto the system to study a particular topic, he/she is provided with an introduction which is followed by a pretest. If the pretest score was high enough (around 80%), the student is allowed to skip to the next lesson. However, if the student does not demonstrate mastery on the pretest, the software provides a video lesson which includes examples of solved problems. The lesson is climaxed by a lesson
summary and a posttest with problems similar to the pretest. The system maintains record of students’ performance and also supplies homework problems. As tools, CAS are simply used to aid in calculations and provide graphical displays. All CAS in the current study were categorized either as tutors or tools.

While the bulk of the research on the use of CAS has been conducted by proponents of technology, some strong counter-arguments have been made that are worthy of consideration. Gleason (2001) contended that there are too many different types of CAS available and, in most cases, systems used in high schools are vastly different from those used at the collegiate level. This lack of congruence often leads to professors at the collegiate level expending a considerable portion of their scarce time training students how to use a new system.

Some researchers (e.g. Gleason, 2001; Allen et al, 1999) further warned that CAS promotes dependence on machines at the expense of developing computational skills; students who use CAS in classrooms are often at a disadvantage when they have to sit for standardized exams which forbid the use of calculators and other technological crutches. Earlier, Hiebert and Lefevre (1986) acknowledged that students need to combine both conceptual and procedural knowledge in order to do well in mathematics.

Russell and Hanley (1997) cautioned that CAS underestimates students’ achievement since the software does not provide partial credit and also does not grade all equivalent forms of short answer questions. Nguyen, Hsieh, and Allen (2006) and Allen et al (1999) were concerned about computer network problems such as slow bandwidth or network jam that could inhibit students’ learning.
Allen et al (1999) also suggested the following disadvantages of CAS: “greater time needed for class preparation, lack of administrative recognition of increased teaching load, [instructors] lack of familiarity with the computer and CAS; fear of making syntactical error in class, CAS syntax is an unreasonable burden on students, and, learning curve is too steep; subtracts time from the learning of mathematics” (p. 4).

Clark (1983) argued that technology and other media assist in the process of instruction but do not directly influence instruction. Clark insisted, “The best current evidence is that media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes a change in our nutrition” (p. 445).

**Rationale for the Study**

Perhaps the brouhaha about the use of CAS in mathematics classrooms would have been settled if there was an agreement on how they affect students’ achievement. On the contrary, there appears to be a lack of consensus among researchers on whether CAS have positive impact on students’ achievement. While one group of researchers (e.g., Hamide, 2001; Tiwari, 1999; Runde, 1998; Hembre & Dessart, 1992; Portis, 1991; Schrock, 1989) found that students exposed to CAS performed better than students exposed to non-CAS instruction, others (e.g., White, 1998; Bosché, 1997; Melin-Conejeros, 1992; Hamm, 1989; Gesshel-Green, 1986; Hawker, 1986) discovered no significant difference in the performance of students who receive CAS instruction and students who receive non-CAS instruction. Additionally, some researchers (e.g., Keepers,
1995; Haller, Child, & Walberg, 1988) even implied that students taught using non-CAS instruction performed better than students taught using CAS instruction.

It has been suggested that whenever there are studies with contradicting findings, there is a need to draw general conclusions that take into account characteristics of each study (Glass, 1976). This can be achieved by conducting a meta-analysis of the primary studies. “Meta-analysis offers a way of integrating the findings from a large number of studies of an issue in a more objective fashion without limiting the creative input of the reviewer” (Houston, Peter, & Sawyer, 1983, p. 497). A meta-analyst finds the effect size for each study by dividing the difference of the means of the two comparison groups by either the standard deviation of the control group or the pooled standard deviation of both groups. Effect sizes of individual studies are then combined to find the overall effect of the treatment.

While a search of the literature revealed several meta-analyses that investigated the effect of technology on education in general (e.g., Lou, Phillips & d'Apollonia, 2001; Christmann & Badgett, 2000; Yaakub, 1998; Azvedo & Bernard, 1995; Liao, 1992; Kulik, Kulik, & Shwalb, 1986), a meta-analytic study that integrated findings on the effect of CAS on students’ achievement was not found. This study sought to fill that void.

Additionally, there is an exorbitant cost associated with purchasing technology equipment, providing training for faculty and staff, hiring technology support staff, and maintaining and replacing electronic hardware and software. Kleiman (2006) reported that schools in the US spent approximately 6.9 billion in 1999 for technology-related expenses. In this era of scarcity and accountability, it is prudent to question whether this
amount has been spent wisely. Findings from the current study could provide some
guidance for policy makers.

Purpose of the Study

The purpose of this study was to investigate the overall effectiveness of CAS
instruction, in comparison to non-CAS instruction, on students’ achievement in
mathematics at pre-college and post-secondary institutions. The current study utilized
meta-analysis on a group of primary studies that individually investigated the
effectiveness of CAS on students’ mathematics achievement and explored the extent to
which this overall effectiveness of CAS was moderated by various study characteristics.

CAS instruction was operationally defined as any form of instruction in which
CAS were used to supplement or replace the classroom learning experience. Non-CAS
instruction was operationally defined as any form of instruction in which CAS were not
used during content delivery. Within the context of this research, it included one of the
following situations: (a) the control group was taught using traditional chalk-and-talk
method, (b) the control group was taught using a blend of the traditional method and
cooperative group activities, and (c) the control group was not allowed to use CAS.

Research Questions

1. What is the overall effectiveness of CAS instruction on students’ achievement in
mathematics, in comparison to non-CAS instruction, at pre-college and post-
secondary institutions?
2. Does the effectiveness of CAS on students’ achievement in mathematics differ by type of publication (published or unpublished), type of usage (tutor or tool), location (US or international), time (1980-1989, 1990-1999, or 2000-2007), educational level (pre-college or postsecondary), study design (studies that controlled for the effect of teacher or studies that did not control for the effect of teacher), evaluation method (studies in which the experimental group used CAS during evaluation or studies in which the experimental group did not use CAS during evaluation), or course-type (algebra, calculus, or other)?

Null Hypotheses

The research questions were scrutinized via the following null hypotheses:

1. There is no significant difference between the effectiveness of CAS instruction and the effectiveness of non-CAS instruction in mathematics at pre-college and post-secondary institutions.

2. The effectiveness of CAS on students’ achievement does not differ by type of publication (published or unpublished).

3. The effectiveness of CAS on students’ achievement does not differ by type of usage (tutor or tool).

4. The effectiveness of CAS on students’ achievement does not differ by location (US or international).

6. The effectiveness of CAS on students’ achievement does not differ by educational level (pre-college or postsecondary).

7. The effectiveness of CAS on students’ achievement does not differ by study design (studies that controlled for the effect of teacher vs. studies that did not control for the effect of teacher).

8. The effectiveness of CAS on students’ achievement does not differ by evaluation method (studies in which the experimental group used CAS during evaluation or studies in which the experimental group did not use CAS during evaluation).

9. The effectiveness of CAS on students’ achievement does not differ by course-type (algebra, calculus, or other mathematics courses).

Limitations

Many researchers in the field of social science are constrained to use quasi-experimental design, since in most cases; neither random selection nor random assignment is feasible. Findings from the current study depended on results from primary studies that generally measured the effect of CAS on achievement through the use of quasi-experimental design. Any flaws inherent in the primary studies may have influenced the overall result of the meta-analysis. This influence, however, was minimized when I coded studies into categories (based on type of design and method of evaluation) to investigate whether well-designed studies were more likely to produce positive results than poorly-designed studies.
Delimitations

This meta-analysis integrated studies on the effectiveness of CAS on achievement from online sources. Only studies that were available online and/or were cited by online databases were included. I also set other requirements (e.g., year of publication, type of design, etc.) for selecting primary studies. Studies that did not meet the selection criteria and/or studies with insufficient data (e.g., means, standard deviations) to calculate effect sizes were excluded. Results from the current meta-analysis should only be generalized to studies that met the entry requirements.

Operational Definitions

*Computer Algebra Systems (CAS).* Computer software packages that perform calculations, symbolic manipulations, and offer graphical representations (Pierce and Stacey, 2001).

*Non-CAS Instruction.* Any form of instruction in which CAS were not used during content delivery.

*Control Group.* Group of students who were not allowed to use CAS to supplement or replace their classroom learning experience.

*Experimental Group.* Group of students who were allowed to use some form of CAS to supplement or replace their classroom learning experience.

*Computer-assisted instruction (CAI)/Computer-based instruction (CBI).* The use of computer or computer technology to provide course content in the form of drill, practice, tutorial, and/or simulations (Adym, 2005).
Computer-managed Instruction (CMI). A form of instruction in which a computer program adapts to the student by selecting appropriate instructional resources and maintaining record of academic progress.

Computer-based CAS. CAS delivered through the use of laptop or desktop computers.

Portable CAS. CAS delivered through the use of a portable device such as graphing calculators.

Computer Algebra System as a tutor. A situation in which a CAS adapts to a learner by selecting, presenting, and evaluating appropriate instructional resources based on the needs of the learner. The system also maintains a record of academic progress.

Computer Algebra System as a tool. A situation in which a CAS is used to execute instructions (such as performing calculations and displaying graphs).

CAS used. The experimental group used CAS during evaluation

CAS not used. The experimental group did not use CAS during evaluation.

Same instructor. The experimental group and the control group were taught by the same instructor

Different instructors. The experimental group and the control group were taught by different instructors.

Pre-college. Kindergarten to grade 12.

Postsecondary institution. Any institution that caters to students who have completed the 12th grade.

Mathematics. All math courses taught at pre-college and post-secondary levels (excluding statistics).
International. Any country besides the United States.

Achievement. The dependent variable in this study includes students’ mathematics performance outcomes as measured by quizzes, posttests, delayed posttests, final grades, or final examination scores.

Summary

Chapter 1 provided general information about the concerns of mathematics professionals and policy makers regarding the performance, or lack thereof, of students in mathematics classrooms. It highlighted issues surrounding the use of technology, in general, and CAS, in particular, in mathematics classrooms. The chapter concluded with the rationale, purpose, research questions, null hypotheses, limitations, delimitations, and operational definitions used in the current study.
CHAPTER II
LITERATURE REVIEW

Introduction

This chapter provides a synopsis of two different and contending philosophies on the teaching of mathematics. The two schools of thought differ on what should be taught and how teaching should be conducted in mathematics classrooms. The chapter explores what teachers from different philosophies believe should be the role of technology in the classroom. A brief discussion of the findings of primary studies on the impact of CAS is presented along with an overview of meta-analysis. The chapter concludes with a summary of meta-analytic reviews on the effectiveness of technology.

The Math Wars

The math wars can be traced to two competing philosophies--classical (traditional) education and progressive (constructivist) education. Classical education has been linked to Plato who suggested that “education for a just society requires reinforcement of the rational over the instinctive and emotional aspects of human nature” (Klien, 2007 p.22). The Platonists subscribe to the “global understanding of mathematics as a consistent connected and objective structure” (Ernest, 1988, p. 250). Progressive education (linked to Jean Jacques Rousseau, John Dewey, and William Kilpatrick) emphasizes student-centered naturalistic instruction (Klien, 2007).

A detailed description of these competing philosophies is beyond the scope of this paper. The discussion is further mired by the fact that even within each philosophy, there are groups who do not quite agree with their peers on key issues. These math wars are
discussed elsewhere (e.g. Klien, 2007; Klien, 2005; Salomon, 1998; Quirk, 1997; Heid, 1997; Greg, 1995); however, a brief overview of the conflict is presented to provide a context for the use of technology in mathematics classrooms.

*The Traditionalists*

Klien (2003) reduced the major contention between the competing philosophies to two key issues: what to teach in a math classroom and how it should be taught. Classical education is often linked with behaviorism. Behaviorism has its genesis in Skinner’s theory of stimulus and response. Skinner maintained that learners can be conditioned to respond based on an existing stimulus and further suggested that all learners are able to gain the same understanding and learning, given the proper environmental influences (Faryadi, 2007; Klien, 2003).

Some (e.g., Straits & Wilkes, 2007; Olech, 1999) have described learning in a traditional classroom as a process of knowledge transmission from outside to within the learner; lecture is the primary mode of instruction and the teacher plays the role of passing down facts, that have been established by others, to the students. Stated differently, the teacher is responsible for providing materials and creating the environment where students learn, that is, the teacher strives to modify students’ behavior. Students are expected to acquire knowledge through drill and practice or through practice and reinforcement. Learning occurs when students are able to demonstrate a new observable behavior (Faryadi, 2007; Adym, 2005; Olech, 1999).

Classical mathematicians (traditionalists) believe that math is abstract only in the minds of humans and that it can be learned in the same way by different individuals.
They further argue that when math is taught, its purpose should include: (a) preparing citizens for everyday life, (b) developing the mind to reason both in logical terms and in abstract terms, (c) preparing individuals for careers that might demand math knowledge, and (d) creating an atmosphere where students can systematically develop their memory skills (Quirk, 1997).

Traditional mathematicians also claim, “Learning math is a process of building a personal knowledge that can be stored in the brain” (Quirk, 1997, p.1). Most traditionalists would agree that students need to develop basic skills in the lower grades and that calculators and other computing devices should not be used for simple computations that can be done by hand. Hunsaker (1997) even cautioned that “calculators prevent students from seeing [the] inherent structure and beauty of math” (p. 20).

In a traditional mathematics classroom, the teacher plays the key role of presenting facts, demonstrating math skills, orienting students to new topics, solving examples, providing questions for drill and practice, continuously asking questions to gauge students’ understanding, and providing constant feedback. Students acquire skills and algorithms through solving problems that are similar to examples that have been worked by the teacher and are encouraged to explain their answers in proper mathematical language (Abdullah, 2006; Quirk, 1997).

The entire process of learning in a traditional classroom begins at a slow pace with students having very little or no mathematical knowledge and moves on to a faster pace as they expand their knowledge base. As the process continues, new knowledge
helps in clarifying the students’ initial thoughts and understanding evolves. Teachers are expected to assist students to learn math skills and facts (Quirk, 1997).

The Constructivists

Progressive math educators (constructivists) generally subscribe to constructivism as a teaching philosophy. Constructivists describe learning as a process in which students construct their own meaning and create their own interpretation through interaction with the environment. For learning to occur, information is constructed in the mind of the learner through the process of assimilation and accommodation. During assimilation, a learner takes new information and blends it with existing knowledge without making any changes in the knowledge structure. Accommodation involves the learner taking new information and restructuring the existing knowledge base to fit in the new information (Straits & Wilkes, 2007; Rakes, Fields & Cox, 2006; Gales & Yan, 2001; Reinking, Labbo, & McKenna, 2000).

Constructivists in mathematics education generally agree: “all knowledge is constructed. Mathematics knowledge is constructed, at least in part, through the process of reflective abstraction. There exist cognitive structures that are activated in the process of construction…Cognitive structures are under continuous development” (Salomon, 1998, p. 10).

Phillips (1995) even argued: “these days we do not believe that individuals come into the world with their cognitive ‘data banks’ already pre-stocked with empirical knowledge…Nor do we believe that most of our knowledge is acquired, ready-formed by some sort of direct perception or absorption” (p. 5).
Constructivists describe their mathematics classroom as “learner-centered environments,” where students are “actively engaged” in the process of “discovery,” with the teacher serving as a “facilitator.” Teachers who ascribe to a constructivist philosophy claim to organize their information around concepts that affect real life situations. They also advocate for providing an atmosphere for self-regulated learning (Gales & Yan, 2001).

Some constructivists (e.g., Faryadi, 2007; Hung, 2001) argue for mathematics to be more relevant to students’ everyday lives and have called for less emphasis on the development of basic skills and more emphasis on broader concepts. Davis, Maher, and Noddings (1990) opposed the learning of mathematics in a cumulative fashion and contended that “the acquisition of rote skills in no way ensures that learners will be able to use these skills intelligently in mathematical settings” (p. 187). Davis et al. further proposed that teachers ought to give students the tools to think by emphasizing concepts, metaphors and other heuristics that would allow students to create their own understanding. Elsewhere, Juniu (2006) suggested constructivists generally agree that students need to work in small collaborative groups to enhance their learning.

In summary, while a traditionalist might perceive learning as a process where knowledge is transferred from the teacher to the students, a constructivist sees learning as a process where knowledge is created by students as they (students) impose meaning on their experiences.
Teaching Philosophy and the Use of Technology

Much has been said about what constructivists and traditionalists believe should be the role of technology in mathematics classrooms and how technology can be used to support learning. There appears to be tremendous support for technology use from the NCTM, the AMTE and other powerful teacher organizations who have advocated for a learner-centered approach to teaching mathematics.

Starting with NCTM’s Agenda for Action, released in 1980, the organization has been very vocal in calling on teachers to “decrease emphasis on such activities as …performing paper and pencil calculation with numbers of more than two digits…all students should have access to calculators and increasingly to computers throughout their school mathematics programs” (p. 8).

Technology seems to have found a natural ally in constructivists’ classrooms. For instance, Heid and Edwards (2001) declared: “the introduction of computer-based manipulation utilities into secondary school classrooms opened the possibility of a shift from emphasis on traditional algebraic tasks such as equation solving and simplification of algebraic expressions to the development of deeper conceptual understanding and the ability to apply algebra to real word settings” (p. 128).

Elsewhere, Heid (1997) stated, “student-centered education is valuable…technology is a powerful way to make education more student-centered…Giving a student the experience of being a mathematician is valued and technology is thought to provide the opportunities for these experiences” (p. 8).
Most traditionalists do not entirely reject the use of technology in the classroom; however, they argue that the main goal of elementary mathematics is for students to develop their thinking about numbers and to learn basic arithmetic. Traditionalists further contend that if students do not develop basic skills (such as learning how to add, subtract, multiply, and divide single digit numbers) and memorize basic math facts, they (students) would have a difficult time with mathematics beyond elementary school (Klien, 2005). Stated simply, traditionalists tend to agree that students need to develop their basic skills before they are allowed to use calculators and other technology devices.

Traditionalists who use mathematical software in the classroom are more apt to choose programs that emphasize drill-and-practice. Many of the large meta-analyses on computer-based education (e.g., Kulik & Kulik, 1980; Kulik & Kulik, 1986; Roblyer, 1989; Kulik & Kulik, 1991) involved primary studies in which the experimental group was taught using drill-and-practice programs.

Several researchers (Juniu, 2006; Hung, 2001; Norton et al., 2000; Olech, 1999) discovered a linkage between teachers’ resistance to the use of technology in the classroom and their educational philosophy; traditionalists were less likely to use technology in the classroom than teachers with a learner-centered focus. Hung (2001) also added that when traditionalists used technology in their math classrooms, they were more likely to choose programs that emphasize the importance of drill-and-practice.

Primary Studies on the Effectiveness of CAS

A plethora of independent research articles have been written about the effectiveness of technology, in general, and CAS, in particular. While some researchers
compared the effectiveness of CAS with traditional instruction (e.g., Batchelder & Rachal, 2000; Campbell, 1996; Cooley, 1995; Cooley, 1997; Fletcher, Hawley, & Piele, 1990; Hagerty & Smith, 2005; Hamide, 2001; Hollard & Norwood, 2005), others (e.g. Brown, 2007; Ford & Klicka, 1994; Gesshel-Green, 1986; Hawker, 1986; Keepers, 1995; Melin-Conejeros, 1992; Runde, 1997; Runde, 1998) did not clearly indicate what form of instruction was used on the control group (students in the control were only described as not been allowed to use CAS). Few others (e.g., Powers, Allison, & Grassl, 2005, Kramarski & Hirsch, 2003) reported a blend of traditional instruction and cooperative group activities for students in the control group. Findings from studies on the effectiveness of CAS have been inconclusive.

One group of researchers (e.g., Fletcher, et al., 1990; Levin, Glass, & Meister, 1987; Jamison, Fletcher, Suppes, & Atkinson, 1976) approached the issue of the effectiveness of CAS from a cost/benefit analysis. Jamison et al.(1976) conducted a study on providing academic support to disadvantaged students and reported a cost of $14 monthly (per student) for grade placement gain in arithmetic computation skills (as measured by the Stanford Achievement Test) when using technology. Levin et al. (1987) arrived at a cost of $143 annually to provide a student with 10 minutes of computer-assisted mathematics instruction but argued that this was less costly than tutoring by adults for 20 minutes per day.

Fletcher et al. (1990) provided even more details. They reported: “In grade 3, the cost per month of grade placement gain in total mathematics was about $20 per month per student for microcomputer assisted instruction and $33 per student for conventional
instruction. These costs were about $17 and $27, respectively, in grade 5” (p. 783). The problem with these cost/benefit analyses is that in all of these studies, students received traditional instruction part of the time, and the researchers did not isolate the effect of traditional instruction from the overall gain in students’ achievement.

Palmiter (1991) and Palmiter (1986) sought to measure the effect of MACSYMA, a CAS, on the academic performance of college students. Working from the premise that CAS reduces the amount of instructional time, Palmiter (1991) exposed students in the experimental group to five weeks of CAS instruction while students in the control group received 10 weeks of traditional instruction in both studies. All classes covered the same course content.

Palmiter (1991) found that students in the experimental group were faster, more accurate, and did significantly better on solving computational problems than students in the control group. On the other hand, Palmiter (1986) found no significant differences on the common final exam that required paper and pencil computation (students in the control group did slightly better) but reported that students in the experimental group did significantly better in a subsequent math course.

Some studies (Powers, et al., 2005; Tab, 2001; Graham & Thomas, 2000; Hollard & Norwood, 1999; Runde, 1998; Runde, 1997) examined the effectiveness of PCAS on the achievement of students. Powers et al. let students in the control group to blend traditional lecture with cooperative group activities; Tab allowed students in the control group to receive traditional instruction. Graham and Thomas (2000), Hollard and
Norwood (1999), Runde (1997), and Runde (1998) simply stated that the control group was not allowed to use CAS.

Tab (2001) did not find significant differences between students taught using PCAS and students taught using traditional instruction; however, Graham & Thomas (2000), Hollard & Norwood (1999), Runde (1998), and Runde (1997) found that students taught with PCAS performed significantly better than their peers using non-CAS instruction. Also, Runde (1998) reported that, at the end of the semester when PCAS were removed, no significant differences were found between the comparison groups on posttest.

Kramarski and Hirsch (2003), Shaw, Jean, and Peck (1997), Keepers (1995), and Melin-Conjeros (1992) used the software, DERIVE, to investigate the effectiveness of CAS in mathematics classes. Melin-Conjeros concentrated his experiment on the use of CAS to do calculus homework while the rest of the studies focused on classroom instruction.

Kramarski and Hirsch (2003) allowed students in the control group to blend traditional lecture with cooperative group activities; Shaw et al. exposed students in the control group to traditional instruction, while Keepers (1995) and Melin-Conjeros (1992) indicated that the control group was not allowed to use CAS. Two studies (Keepers, 1995; Melin-Conjeros, 1992) reported no significant differences between the experimental group and the control group on overall achievement. The remaining studies (Kramarski & Hirsch, 2003; Shaw et al, 1997) found significant achievement differences in favor of the experimental group.
Few others (Hamide, 2001; Tiwari, 1999; Cooley, 1997) used MATHEMATICA as CAS in their college classrooms. Hamide (2001), Tiwari (1999), and Cooley (1997) treated students in the control group with traditional instruction. These studies reported that students in the experimental group did significantly better than their peers in the control group; however, Hamide found that when linear algebra computation problems required the use of procedures, students in the traditional group did slightly better than their peers in the experimental group. Tiwari cautioned readers to be careful in generalizing his findings because of small sample size (29 students were in the experimental group and 29 students were in the control group).

Additionally, Trout (1993) used Mathematics Exploration Toolkit (MET) while Hawker (1986) used muMaATH/muSIMP-83. Both systems were used for drill-and-practice in undergraduate math courses. Trout and Hawker indicated that students in the control group received non-CAS instruction, and that the comparison groups received instruction for an entire semester.

Using pretest as a covariate, Trout found that students in the CAS group did significantly better than students in the non-CAS group, but were not more or less likely to choose another math course. On the other hand, Hawker found no significant differences between the comparison groups in achievement, attitudes towards mathematics, and drop-out rates.

Limitations of Primary Studies

Researchers have spent a portion of their time discussing the limitations of primary research. Davies (2003) warned: “Single studies, even if they are randomized
controlled trials or other types of experimental inquiry, have limitations of time-, sample-
and context-specificity which can undermine their applicability, relevance and usefulness
in other contexts” (p. 366). Cook et al. (1992) emphasized that it is difficult to generalize
the findings of a single study as single studies tend to “illuminate only one part of a large
explanatory puzzle” (p. 3). Green and Hall (1984) stated, “A single study is never
definitive no matter how memorable and newsworthy it may be” (p. 38).

As the volume of independent studies in a field grows over time, the average
reader experiences a tremendous amount of trepidation trying to make sense out of a
group of studies that have been conducted in different settings and under different
experimental conditions. Andrews and Harlen (2006) concluded that it is impossible for
any individual to be able to read and formulate a coherent opinion from a vast body of
literature on a particular topic.

Investigators agree that the field of research is further enhanced when
independent studies are synthesized to identify more causal factors, generalizable trends
and underlying principles. To summarize this belief, McMillan and Schumacher (1984)
maintained:

The weakest argument is based on what the researchers say, not what they do or
show… A stronger argument is made when it can be demonstrated that certain
methods and techniques were used to control threats to internal validity, and still
stronger arguments are made when replications by other researchers verify the
results. The strongest type of argument is made when reasoning, methods and
replications are combined (p. 367).

The findings from independent researchers suggest a lack of consensus on the
effectiveness of CAS. While one group of researchers reported positive findings in favor
of CAS, others reported negative or neutral findings. Moreover, the findings were
sometimes heterogeneous even for studies that used the same type of CAS. Inconsistencies in the results from primary studies provide a logical reason for combining these findings through a quantitative synthesis.

Traditional Syntheses of Research

Hale and Dillard (1991) stated that the two goals of most research syntheses are to do a general summary of what is known about a phenomenon of interest or to offer a critique on the theory that underlies a body of independent work. Each research synthesis is done with the hope that the researcher will strive to minimize biasness. Up until the mid-1970’s, combination of primary research was done almost exclusively by traditional reviews. Fitzgerald and Rumrill (2003) identified three broad traditional review methodologies: narrative review, vote counting, and combined significance tests.

Narrative Reviews

“Narrative reviews present verbal descriptions of past studies focusing on theories and framework, elementary factor and their roles (predictor, moderator, or mediator) and/or research outcomes (e.g., supported vs. unsupported) regarding a hypothesized relationship” (King & He, 2005, p. 667). When conducting a conventional narrative review, the researcher collects a group of independent studies on a topic and summarizes their findings. Johnson (1989) described the process as “stepping through the studies one-by-one as though each were a case study, attempting to form conclusions from [the] intuitive sum of the separate findings” (p.2).
While a conventional narrative review could be somewhat informative for policy makers, the process is embedded with several problems that make it susceptible to high levels of biasness. Some narrative reviewers often do a summary on a group of studies without clearly describing their methodology. This makes it extremely difficult, if not impossible, for other researchers to replicate their study and/or examine the validity of their findings (Petrosino & Lavenberg, 2007; King & He, 2005; Glass, McGaw, & Smith, 1981). It is not uncommon for researchers working from the same set of studies to arrive at different conclusions.

Some reviewers also do not clearly state their standard for determining what studies should be included in a narrative review. Studies are selected for inclusion at the discretion of the researcher who chooses studies that are considered to have high internal validity. Additionally, information, in most cases, is not provided about the reason for including or excluding studies. Since the definition of “high internal validity” might differ from one researcher to another, narrative reviews could be strongly influenced by the biases of individual reviewers (Wolf, 1986; Houston et al., 1983; Wolf, 1981). “At its worst, a reviewer advocating a position could selectively include only those studies favoring that viewpoint” (Petrosino & Lavenberg, 2007, p. 6). Moreover, unpublished studies are sometimes excluded from narrative reviews.

As the number of studies on a phenomenon of interest increases, narrative reviewers have a difficult task of combining these findings in the absence of any standardized approach. Houston et al. likened the process to one in which reviewers seeks to make sense out of “many data points that result from a single research project.”
Extracting meaning from the many studies in narrative fashion is not unlike an attempt to extract meaning from a raw data matrix prior to reduction” (p. 497).

Narrative reviewers are further criticized for failing to critically examine the characteristics, design, effect, and conclusions drawn from independent studies and for overlooking the impact of moderating variables. Narrative reviews are not very effective in examining results from quantitative studies (Fitzgerald & Rumrill, 2003; Wolf, 1986; Houston et al., & Sawyer, 1983).

This is not to say that narrative reviews have no use. Some narrative reviewers strive to minimize these biases by clearly indicating search criteria, including unpublished studies and delineating study boundaries. Even when this is done, narrative reviews still do not provide a measure of the effect size.

Vote Counting

In a bid to improve the quality of research reviews, and to address the demands for quantitative synthesis, the vote counting method was introduced. The role of the reviewer in vote counting is analogous to the role of an election officer who tallies the votes from two opposition parties (with extreme left and right views) and a moderate independent candidate (Fitzgerald and Rumrill, 2003; Houston et al., 1983; Wolf, 1981; Wolf, 1986; Glass, 1976).

Vote counting “uses the outcomes of hypothesis reported in individual studies such as probabilities, p-levels, or results falling into three categories: significantly positive effect, significantly negative effect, and non-significant effect” (King & He,
The number of studies in each category is counted and the category with the most tallies is considered as being more effective.

Vote counting has received a fair share of literary beating from other researchers. In addition to the criticisms of selection bias, failing to include unpublished studies, ignoring the design and characteristics of primary studies, and overlooking the effects of moderating variables, Rhea (2004) argued:

…the p value can be very misleading, especially when studies are performed on small samples, which limit the statistical power and increases the likelihood of the researcher making a type II error (failing to reject the null hypothesis when in fact it is false). A significant p value may also be misleading (Type I error) if the actual magnitude of the difference is so small as to be of little consequence. Thus, the p value offers no measure of the actual magnitude or direction of the treatment effect (p. 921).

Given that statistical significance is generally a function of sample size; studies with large samples are more likely to find significant results than studies with smaller samples. Further, all significantly positive effects (or significantly negative effects) are given equal weights regardless of the magnitude (Wolf, 1986; Rosenthal, 1984; Glass, 1976). Hedges (1986) added: “Vote counting not only has low power to detect effects under the conditions in which it is actually used, but the power may actually decrease (tending to zero) as the number of studies increases” (p. 356).

Since reviewers using the vote counting method simply group studies based on their positive significance, or lack thereof, one might assume that all comparisons were tested in the same way (directional or nondirectional) and/or used the same level of type I error. This generally was not the case (Jackson, 1980). Additionally, the vote counting process does not provide any measure of effect size.
Combined Significance Tests

The combined significance tests were introduced to minimize the criticisms of conventional review methods. In combined significance tests, “probabilities related to effects or relationships of primary studies [are] combined to assess the overall [average] effect for a group of studies” (Fitzgerald and Rumrill, 2003, p. 99). Stated differently, a combined significance test is used to test the overall significance of a set of independent tests, all of which focus on the same null hypothesis.

The most widely used combined significance tests is the Fisher’s (1932) combined probability test. Fisher (1932) proposed:

When a number of quite independent tests of significance have been made, it sometimes happens that although few or more can be claimed individually as significant, yet the aggregate gives an impression that the probabilities are on the whole lower than would often have been obtained by chance. It is sometimes desired, taking account only of these probabilities, and not of the detailed composition of the data from which they are derived, which may be of very different kinds, to obtain a single test of the significance of the aggregate (p. 99).

Sometimes referred to as omnibus or nonparametric test because of its lack of dependence on the type of data or the underlying statistical distribution, Fisher’s combined significance test assumes that probability (p) values are uniformly distributed between 0 and 1 and that the results of n independent tests, all of which address the same null hypotheses, can be combined by multiplying their individual probabilities.

Fisher demonstrated that for a given hypothesis, \(-2\log p\) has a chi-square distribution with 2 degrees of freedom. This line of reasoning was extended to suggest that if the null hypothesis is true, then

\[-2\log (p_1p_2p_3...p_n) = -2\log p_1 -2\log p_2 \ldots -2\log p_n,\]
where $p_1 \ldots p_n$ are the probabilities obtained from $n$ independent tests. Consequently, the null hypothesis is rejected if $P = -2 \sum \log p_i$ (for $i = 1 \ldots n) > C$, where $C$ is the critical value obtained from the upper tail of the chi-square distribution with $2n$ degrees of freedom (Hedges, 1992).

The second most widely used combined significance method was proposed by Tippett (1931). Using the assumption about the uniformity of $p$ over the interval $[0, 1]$, Tippett argued that since the probabilities associated with individual tests were independent, a test for the significance of the overall null hypothesis ($H_0$) at a given significance level ($\alpha$) could be obtained. Tippett showed that if $p_s$ was the minimum of $p_1, p_2 \ldots p_n$ obtained for $n$ null hypotheses ($H_{o1}, H_{o2} \ldots H_{on}$), the decision rule is to reject $H_0$ if $p_s < 1 - (1 - \alpha)^{1/n}$.

While combined significance tests enable researchers to aggregate probabilities from independent tests that investigate the same null hypotheses, the process has been criticized because “tallies of statistical significance or insufficiency tell little about the strength or importance of the relationship” between comparisons groups (Glass et al., 1981, p. 95). Put simply, a combined significance test does not provide a measure of the effect of an intervention.

**Overview of Meta-analysis**

The inadequacies of the earlier review processes make a strong case for an unbiased analytic review procedure. Glass (1976) proposed three levels of data analysis: primary analysis, secondary and meta-analysis. He defined primary analysis as “the original analysis of the data in a research study” and secondary analysis as “the re-
Glass coined the term meta-analysis, “the analysis of analysis,” to refer to “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings” (Glass, 1976; p. 3).

Glass (1976) used the term meta-analysis to stress a philosophical approach as opposed to a statistical technique. Emphasizing the inadequacies of the methods used to combine studies at the time, Glass felt there was a need to standardize and systematize the process of combining independent evaluations so that other researchers could replicate the process.

**Advantages of Meta-analysis**

Meta-analytic studies generally tend to produce more reliable results than traditional reviews because the biasness of the investigator is minimized. An investigator conducting a meta-analysis must carefully specify the criteria used for including (or excluding) primary studies. The researcher often does not selectively exclude studies because they lack “high internal validity” but includes all studies and could test whether studies that are methodically flawed are likely to produce different results. Meta-analysts tend to include both published and unpublished work. Further, because of the rigid structure of meta-analysis, the researcher is less likely to suffer from “cognitive overload.” (Fitzgerald & Rumrill; 2003; Rosenthal, 1984; Houston et al; 1983; Johnson, 1981; Cook & Leviton, 1980; Glass, 1976).
Moreover, the result from a meta-analytic study is more robust than results from primary studies because the meta-analyst combines several independent studies, conducted under different conditions, to reach a conclusion. As opposed to using a traditional review, the researcher finds the effect size for each study by dividing the difference of the means of the two comparison groups by either the standard deviation of the control group or the pooled standard deviation of both groups. Effect sizes of individual studies are then combined to find the overall effect of the treatment. Effect sizes have a distinct advantage over significance tests because they are less dependent on sample size (Cook & Leviton, 1980; Glass, 1976).

King and He (2005) added:

Meta-analysis enables the combining of various results, taking into account the relative sample sizes, thereby permitting studies showing insignificant results to be analyzed with studies showing significant effects. The overall result may be either significant or insignificant, but it is undoubtedly more accurate and more credible because of the overarching span of such analysis (p. 671).

Additionally, meta-analysis allows the researcher to include results from studies whose samples were so small that statistical significance was not obtained. This methodology also enables an investigator to test hypotheses and ask new questions which may not have been built into the original studies. The meta-analyst often tries to identify moderator variables that could explain why certain results are obtained from the meta-analysis. (Rosenthal, 1984; Houston et al., 1983; Johnson, 1981; Glass, 1976).

Criticisms of Meta-analysis

Like many analytic procedures, meta-analysis also has seen a fair share of criticisms. Glass et al. (1981) identified 4 major criticisms of meta-analysis: (a) meta-
analysts often combine studies that have been done under different conditions or aggregate studies in which the dependent variable may not have been measured in the same way. “The meta-analysis approach to research integration mixes apples and oranges. It makes no sense to integrate the findings of different studies” (p. 218), (b) when a researcher includes all studies that meet a priori selection criteria, “the meta-analysis approach advocates low standards of judgment of the quality of studies” (p. 20), (c) by combining the results of several studies, some of which have reported several measures of effect size, “this renders the data nonindependent and gives one a mistaken impression of the reliability of the results” (p. 229), and (d) publishers are more likely to publish studies in which the findings are significant than studies in which the findings are nonsignificant; “there are systematic differences among the results of research that appear in journals and books versus theses versus unpublished papers” (p. 226).

The File–Drawer Problem

The last criticism that stresses the inclusion of published studies in favor of unpublished studies is what Rosenthal (1979) referred to as the “file drawer problem.” Rosenthal (1979) argued: “For any given research, one cannot tell how many studies have been conducted but never reported. The extreme view of the ‘file drawer’ problem is that journals are filled with the 5% of the studies that show type I errors while the file drawers are filled with 95% of the studies that show nonsignificant results” (p. 368).

It has been suggested (e.g., Wolf, 1989; Kupfersmidt, 1988; Rosenthal, 1979; Greenwald, 1975) that the file drawer problem originates from two sources: the reluctance of researchers to submit findings from nonsignificant studies to editors for
publication and the tendency of editors to generally publish only studies that demonstrate significant differences between the comparison groups.

To demonstrate the latter point, Atkinson, Furlong and Wampold (1982) submitted three versions of the same paper to more than one hundred consulting editors of psychological journals. The only difference among the versions was the level of statistical significance. Atkinson et al. discovered that the statistically significant version of their manuscript was more than three times less likely to be rejected than the nonsignificant and marginally significant versions.

Responding to the Criticisms of Meta-analysis

Wolf (1986) proposed the following measures to address the criticisms of meta-analysis: (a) researchers are advised to code the characteristics of each study and test whether differences observed among study characteristics influence the results of the meta-analysis, (b) researchers are cautioned to code the quality of the design of studies to determine whether “the results differ for poorly designed studies and well designed studies,” and (c) it is strongly recommended that researchers categorize studies into published and unpublished studies to investigate the existence of publication bias. Alternatively, researchers are advised to “estimate the number of additional studies with nonsignificant results that would be necessary to reverse a conclusion drawn from the meta-analysis…” (p. 15).
Meta-analyses on Technology Effectiveness

Several meta-analytic studies have been conducted to determine the impact of technology on education, in general. Kulik, Kulik, and Cohen (1980) synthesized the findings from 59 independent evaluations of CBI carried out between 1967 and 1978 and reported that “CBI made small but significant contributions to the course achievement of college students” (p. 525). The primary studies in this meta-analysis described four applications of computers in a learning environment: tutoring, computer-managed teaching, simulation and programming the computer to solve problems. These researchers found an overall effect size of 0.25 and inferred that a typical student at the 50th percentile of the control group could experience an increase in performance to the 60th percentile, if that student was exposed to CBI. Their findings did not specify the effect of CBI on individual subject areas.

Having quantitatively combined the results from 51 individual studies on the impact of CBI students’ learning at the middle school and high school level, Kulik (1983) noted that CBI increased the performance of an average student in the experimental group to the 62nd percentile, when compared to an average student (at the 50th percentile) in the control group. This finding did not indicate how students performed in different content areas.

Elsewhere, Kulik, Kulik and Bangert-Drowns (1985) combined the results from 32 studies that compared traditional and computer-based education in elementary schools. These researchers also divided the use of computers in academic settings into two broad categories: CAI (drill-and-practice and tutorial) and computer-managed
instruction (CMI). When used for CAI, “the computer provided practice exercises but not
the original lesson on a topic” or “presented the lessons as well as practice exercises on
the material.” When used for CMI, “the computer evaluated student performance, guided
students to appropriate instructional resources, and kept record of student progress” (p.
64).

Kulik et al. (1985) found CAI (average effect size = 0.47) to be significantly more
effective than CMI (average effect size = 0.07); 28 studies used CAI while the remaining
4 studies used CMI. Stated differently, 68% of the students using CAI performed better
than the average student exposed to traditional instruction. In contrast, only 53% of the
students exposed to CMI performed better than the average student exposed to traditional
instruction. While this study spanned several subject areas, the effect of computer
education on mathematics was not reported.

Similarly, Kulik, et al. (1986) quantitatively fused together 23 independent studies
on adult learners in basic adult education or technical training. The average effect size,
0.42, suggested that if a regular student exposed to CBI was compared with a regular
student in a traditional classroom on the same exam, the former would score at the 66th
percentile while the latter would score at the 50th percentile. Their finding did not include
CBI’s effect on mathematics achievement.

Also in 1986, Kulik and Kulik meta-analyzed findings from 99 studies that
focused on the performance of college students. The overall effect size was 0.26 and this
meant that the use of CBI increased the mean of those exposed to the technology by 10
percentile points, when compared to the control group. The effect of CBI on mathematics was not reported.

Roblyer (1989) conducted a meta-analysis that included 38 published/unpublished studies and 44 dissertations and found that computer applications significantly improved students’ performance. Roblyer further reported that while the effectiveness of technology was not significantly different for mathematics and reading/language, computer applications had a slightly greater effect for math than language. Roblyer, however, did not report the effect sizes for each content area.

Swan, Guerrero, Mitrani, and Schoener (1990) examined the efficacy of 13 CBI programs in 26 elementary and secondary schools in New York City and found overall effect sizes for mathematics and reading to be 0.8 and 0.9, respectively. They reported effect sizes for elementary school students (1.1-reading, 1.2-mathematics), junior high school students (0.7-reading, 0.4-mathematics) and high school students (0.3-reading, no effect size reported for math); however, subjects of the study were educationally disadvantaged children. While indicating that CBI improved the performance of students in the experimental group, these results seem to suggest the effect of CBI on disadvantage children decreases with educational level. This study concentrated on a small fraction of the student population (educationally disadvantaged students) and did not specifically focus on the effects of CAS on students’ achievement in mathematics.

Kulik and Kulik (1991) combined results from 254 studies that spanned different content areas including math, science, social science, reading and language. The primary studies included in their meta-analysis were dated between 1967 and 1986 and covered
elementary, secondary and post-secondary educational levels. Kulik and Kulik (1991) reported that CBI improved students’ attitudes towards mathematics and increased the performance of students in the experimental group to the 62nd percentile, when compared to the 50th percentile of the control group. Their results, however, did not indicate how CBI affected performance for individual subjects.

Along the same lines, Liao (1992) combined the results from 31 individual studies on the effectiveness of CAI that were published between 1968 and 1989. The individual studies assessed the relationship between CAI and cognitive skills such as planning skills, problem-solving skills, and thinking skills and spanned different content areas. Liao determined that students exposed to CAI scored 18 percentile points higher on various cognitive ability tests than students who were not exposed to CAI. These findings, however, did not indicate how CAI affected each content area.

In another study, Kulik (1994) analyzed more than 500 individual research studies on the effect of CBI and reported that: (a) students learn more in less time when they are exposed to CBI, (b) students develop more positive attitudes towards learning, and (c) CBI increased the performance of students in the experimental group to the 64th percentile when compared to 50th percentile of the control group. Kulik (1994) also declared that the effect of CBI was not the same for all content areas but gave no measure of effect size for each content area.

Khalili and Shashaani (1994), having synthesized the findings from 36 independent studies published between 1988 and 1992, reported that computer applications improved the performance of students enrolled at elementary, high school,
and college levels. More specifically, a regular student exposed to computer application was expected to see a gain in achievement from the 50th percentile to the 65th percentile of the control group. Khalili and Shashaani also computed the mean effect size for mathematics (0.52) from 18 primary studies, suggesting an impressive increase of 20 percentile points. Their findings, though, included students who used CAI and students who used programming languages such as BASIC and Pascal.

To determine the effectiveness of computer-presented feedback (CPF) on learning, Azvedo and Bernard (1995) quantitatively compiled the results from 22 studies that used posttest and 9 studies that used delayed posttests. They found an overall effect size of 0.8 for studies that used posttest and 0.35 for studies that used delayed posttest, suggesting that students exposed to CPF were expected to perform better than students exposed to traditional feedback. Additionally, students exposed to CPF were more likely, than their traditional peers, to retain the knowledge they acquired. Participants in the primary studies were enrolled at all educational levels. Their study spanned several content areas but did not report the effectiveness of CPF for any particular content area.

Christmann et al. (1997) quantitatively reviewed 26 independent studies dated between 1984 and 1995. Intervention in these studies was performed on students in high school. These researchers found that “students receiving traditional instruction supplemented with CAI attained higher achievement than did 57.2% of those receiving only traditional instruction” (p. 325). Christmann et al. also reported smaller mean effect size for studies conducted between 1984 and 1995, suggesting that the effect of CAI reduced with time.
Like others, Yakuub (1998) sought to determine the effectiveness of CAI by synthesizing findings from 21 primary studies. The overall effect size, 0.35, indicated that if a regular student exposed to CBI was compared with a regular student in a traditional classroom on the same exam, the former would score at the 64th percentile while the latter would score at the 50th percentile. Yakuub’s study concentrated on learners enrolled in secondary education, postsecondary education, and adult military training and highlighted the uses of CAI in technical education and training. The results did not include the effect of CAI on mathematics.

Additionally, having combined 26 independent evaluations, Christmann and Badgett (2001) found a mean effect size of 0.127 and inferred that students exposed to CAI were likely to perform better than 55% of the students receiving only traditional instruction. Christmann and Badgett also reported mean effect sizes for several content areas: aviation (0.77), English (0.612), athletic training (0.246), education (0.222), business (0.210), science (0.173), reading (0.082), mathematics (–0.031), and music (–0.428). Of the 26 studies in the original analysis, only 2 involved mathematics. Meta-analytic results from 2 studies make it difficult to draw a firm conclusion about the effectiveness of CAI in mathematics classrooms.

Bayraktar (2001) sought to determine the effectiveness of CAI on science education by quantitatively combining the results from 42 primary studies conducted in the US between 1970 and 1999. The studies involved students enrolled at secondary and college levels. She found that, when exposed to CAI, the average student’s performance increased from the 50th percentile to the 62nd percentile. The majority of the individual
studies in this meta-analysis were from chemistry and biology. This finding makes it somewhat difficult to extrapolate as to how CAI might impact students’ learning in a mathematics classroom.

Schmidt, Weinstein, Niemic, and Walberg (2001) quantitatively combined findings from 18 independent evaluations carried out between 1975 and 1984. The individual evaluations were conducted on students from grade 4 to grade 12. Their overall effect size was 0.52. This suggests that a typical student in a computer-based class and a typical student in a conventional class were expected to score at the 70th percentile and 50th percentile, respectively, if they were both given the same exam covering the same course content. These findings did not include treatment effects for different subject areas.

In a recent study, Timmerman and Kruepke (2006) meta-analyzed findings from 118 primary studies published between 1985 and 2004 to determine the effectiveness of CAI. This study spanned a variety of content areas but focused on students in post-secondary education. The overall effect size was 0.12, suggesting that CAI improved the performance of a typical student in a traditional classroom from the 50th percentile to the 55th percentile. CAI was found to be more effective in the social sciences than in the physical sciences, life sciences or language/humanities. Like Christmann et al. (1997), Timmerman and Kruepke reported smaller mean effect size for studies conducted between 1995 and 2004, suggesting a diminishing effect of CAI over the last few years.

In 2006, Vogel et al. combined 32 individual studies to determine the effectiveness of computer gaming and interactive simulations on achievement. They
found significantly higher cognitive gains for students exposed to the treatment and declared that 1,465 studies with opposing results were needed to change their conclusion. Vogel et al. did not indicate the years in which the primary studies were conducted or the content areas covered.

Briefly stated, while several meta-analytic studies showed that technology increased students’ performance, these studies (with two exceptions) did not document the overall effect of technology on mathematics achievement. Khalili and Shashaani (1994) reported a positive effect for CAI on mathematics; however, their finding included studies in which students used programming languages. Additionally, Christmann and Badgett (2001) suggested a negative effect of CAI on mathematics achievement but this conclusion is minimized by the fact that only two of the primary studies in the meta-analysis involved the use of technology in a mathematics classroom.

Summary

This chapter focused on the known and unknown issues about the effectiveness of technology as an educational tool. Beginning with the contending philosophies on the teaching of mathematics, the chapter also discussed the impact of technology on learning and argued for the need for a meta-analysis to synthesize findings from individual studies. While detailing a series of meta-analytic studies on the effectiveness of technology in general, it was argued that these studies did not focus on the area of mathematics or the use of CAS, in particular.
CHAPTER III

METHOD

Introduction

This chapter describes the methodology applied to answer the following research questions:

1. What is the overall effectiveness of CAS instruction on students’ achievement in mathematics, in comparison to non-CAS instruction, at pre-college and post-secondary institutions?

2. Does the effectiveness of CAS on students’ achievement in mathematics differ by type of publication (published or unpublished), type of usage (tutor or tool), location (US or international), time (1980-1989, 1990-1999, or 2000-2007), educational level (pre-college or postsecondary), study design (studies that controlled for the effect of teacher or studies that did not control for the effect of teacher), evaluation method (studies in which the experimental group used CAS during evaluation or studies in which the experimental group did not use CAS during evaluation), or course-type (algebra, calculus, or other)?

Specific topics included in the chapter are the meta-analytic search procedure, the criteria for selecting primary studies, study characteristics, and a discussion on the calculation of quantitative measures of interest.

Procedure

The purpose of this study was to investigate the overall effectiveness of CAS instruction, in comparison to non-CAS instruction, on students’ achievement in
mathematics at pre-college and post-secondary institutions. The study utilized meta-analysis on a group of primary studies that individually investigated the effectiveness of CAS on students’ mathematics achievement and explored the extent to which this overall effectiveness of CAS was moderated by various studies’ characteristics.

While there are different ways in which meta-analysts combine their variables, researchers tend to have a general framework for how a meta-analysis should be conducted. A meta-analysis proceeds in the following manner: (a) the researcher decides on the issue to be investigated and defines the relationship to be examined, (b) the researcher identifies the dependent variables and the study characteristics to be investigated, (c) the meta-analyst operationally defines all variables in the study, (d) the investigator makes a decision on the sampling criteria (sampling criteria could include, but is not limited to, when the primary studies were conducted, the type of study design, and the type of subjects), (e) the researcher locates studies that meet the predetermined inclusion criteria, (f) the investigator codes study characteristics, (g) the meta-analyst converts statistics from the primary studies into standardized effect sizes, (h) the researcher calculates the average effect size and tests whether this value is significantly different from zero, (i) the investigator examines the independent studies to determine if the magnitude and direction of the effect size is consistent across studies, and (j) since the homogeneity of effect size rarely occurs across a large group of studies, the meta-analyst tries to explain sources of heterogeneity by examining study characteristics (Wolf, 1986; Johnson, 1981; Glass, 1976; Rosenthal, 1984; Houston et al., 1983).
The current study utilized the meta-analytic procedure advocated by Glass (1976), Cohen and Dacanay (1992), Kulik et al. (1980) and Azvedo and Bernard (1995). This approach requires that the researcher objectively locate studies, define study features, code studies’ characteristics, calculate effect sizes, find the mean effect size and explain sources of error variance.

Meta-analytic researchers often choose one of two methods to combine effect sizes from primary studies. Using the first approach, a researcher combines all effect sizes from an individual study and only uses one value per study in the analysis. This approach has been used by Hedges and Stock (1983), and Slavin (1984).

The second method involves extracting multiple effect sizes from each study (based on the number of comparisons of interest) and using each effect size in the meta-analysis. Those who subscribe to the second approach (e.g. Kulik and Kulik, 1991; Glass, 1982; Glass et al., 1981; Ahmad and Shashaani, 1994) have argued that too much information is lost when effect sizes are combined using the first method. The second method was used in the present study.

Criteria for Selecting Studies

The following sampling criteria were applied to select independent evaluations to investigate the research questions:

1. The studies must have been conducted between 1980 and 2007.
2. The studies compared CAS instruction with non-CAS instruction.
3. The studies included sufficient summary statistics (e.g. means, standard deviation, F-value, t-value, p-value) to allow for the calculation of effect size.
4. The researcher(s) used a between-group design (there were experimental and control groups) and compared the comparison groups on some outcome after an intervention had been introduced.

5. The researcher(s) demonstrated that the comparison groups were equivalent at the beginning of the study, used random assignment, or used some form of statistical procedure (e.g. Analysis of Covariance) to adjust for pre-existing differences between/among the groups.

6. Subjects were enrolled in a mathematics (not including statistics) class at the pre-college or post-secondary level.

**Locating Studies**

Studies included in the current meta-analysis were found using only online search engines, electronic journals and research databases. More specifically, research databases (JSTOR, EBSCO, Academic Premier, Electronic Journal Center (EJC), Educational Resources Information Center (ERIC), Dissertation Abstract International (DAI), and Google) were queried using combinations of the following key words: “computer algebra system,” “computer-assisted instruction,” “math and technology,” “computer calculus,” “computer-based instruction and math,” and “computer math.”

Primary studies were located through a combination of electronic search and the ancestry approach. According to Johnson (1981), the ancestry approach involves examining the reference list of located articles for the purpose of finding other studies for inclusion. I queried databases to also search for additional studies that were listed in the
reference list of located articles. The search, however, was limited to studies that were available through online sources or through inter-library loan.

A total of 1,605 citations were identified in the initial search. These studies were further scrutinized to determine if retrieval or inter-library loan request were necessary. Many of them were qualitative or were just descriptions of different types of CAS. After initial scrutiny, 261 articles/dissertations were downloaded or requested for a vigorous inspection.

Of the 261 studies, the most common reason for exclusion was that a study did not have sufficient summary statistics to allow for the calculation of effect sizes. The final search yielded 31 studies which generated 107 effect sizes (most of the studies had multiple effect sizes). Table 1 provides details about the research databases queried,

<table>
<thead>
<tr>
<th>Source</th>
<th>Identified</th>
<th>Reviewed</th>
<th>Included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Premier</td>
<td>206</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>ERIC</td>
<td>238</td>
<td>40</td>
<td>5</td>
</tr>
<tr>
<td>EJC</td>
<td>230</td>
<td>25</td>
<td>2</td>
</tr>
<tr>
<td>Google</td>
<td>490</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>ERC</td>
<td>203</td>
<td>34</td>
<td>5</td>
</tr>
<tr>
<td>DAI</td>
<td>677</td>
<td>69</td>
<td>12</td>
</tr>
<tr>
<td>JSTOR</td>
<td>468</td>
<td>38</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>1605</td>
<td>261</td>
<td>31</td>
</tr>
</tbody>
</table>
the number of studies identified, reviewed and selected.

Studies Characteristics

As suggested by Wolf (1986), the primary studies in this meta-analysis were coded into categories (Table 2) to explain the various sources of error variance. An in-depth review of the primary studies revealed the following: (a) only one team of investigators used random assignment of participants, (b) one researcher used CAS to completely replace traditional instruction (the rest of the studies used CAS to supplement traditional instruction), (c) seventeen studies indicated that the control group received traditional instruction, 2 studies reported that the control group received a blend of traditional instruction and cooperative group activities while the remaining 12 studies stated that the control group received non-CAS instruction, (c) all of the studies used quasi-experimental design, and (d) the duration of treatment in the majority of the studies was for several classes.

Because of insufficient number of studies in the “subject assignment,” “type of intervention,” and “duration of treatment” categories, these variables were eliminated from the analysis. Hasselbring and Goin (1988) cautioned that the effectiveness of a technology should be examined based on how it is used in an education setting. Using this thought, I eliminated the variable, type of CAS, and concentrated on how the CAS was used to provide instruction.

The remaining eight independent variables (type of publication, location, type of usage, time, study design, course type, evaluation method, and educational level) were chosen to describe treatment, study design, setting, and course type:
Table 2

*Study Features by Categories*

<table>
<thead>
<tr>
<th>Study Feature</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication Type</td>
<td>published</td>
</tr>
<tr>
<td></td>
<td>unpublished</td>
</tr>
<tr>
<td>Intervention Type</td>
<td>supplemental</td>
</tr>
<tr>
<td></td>
<td>replaced</td>
</tr>
<tr>
<td>Type of Usage</td>
<td>tutor</td>
</tr>
<tr>
<td></td>
<td>tool</td>
</tr>
<tr>
<td>Subject Assignment</td>
<td>random assignment</td>
</tr>
<tr>
<td></td>
<td>nonrandom assignment</td>
</tr>
<tr>
<td>Evaluation</td>
<td>CAS used</td>
</tr>
<tr>
<td></td>
<td>CAS not used</td>
</tr>
<tr>
<td>Treatment Duration</td>
<td>one class</td>
</tr>
<tr>
<td></td>
<td>several classes</td>
</tr>
<tr>
<td></td>
<td>entire semester</td>
</tr>
<tr>
<td>Educational level</td>
<td>pre-college</td>
</tr>
<tr>
<td></td>
<td>postsecondary</td>
</tr>
<tr>
<td>Study design</td>
<td>same instructor</td>
</tr>
<tr>
<td></td>
<td>different instruction</td>
</tr>
<tr>
<td>Study Feature</td>
<td>Categories</td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
</tr>
<tr>
<td>Location</td>
<td>US</td>
</tr>
<tr>
<td></td>
<td>international</td>
</tr>
<tr>
<td>Type of CAS</td>
<td>computer-based CAS</td>
</tr>
<tr>
<td></td>
<td>personal CAS</td>
</tr>
<tr>
<td>Course type</td>
<td>algebra</td>
</tr>
<tr>
<td></td>
<td>calculus</td>
</tr>
<tr>
<td></td>
<td>other</td>
</tr>
<tr>
<td>Time</td>
<td>1980-1989</td>
</tr>
<tr>
<td></td>
<td>1990-1999</td>
</tr>
<tr>
<td></td>
<td>2000-2007</td>
</tr>
</tbody>
</table>

1. Type of Publication- Studies were coded by type of publication to determine whether published studies were more likely to report a positive effect size (in favor of the treatment group) than unpublished studies.

2. Type of Usage- Taylor (1980) identified three general roles of instructional computers: tutor, tutee and tool. When used as a tutor, the computer adapts to the student by selecting appropriate instructional resources and maintaining record of academic progress. As a tutee, the computer receives and executes instructions (in the form of programming language) from the student. As a tool, the computer
is used to aid in calculation and graphical display. It can be argued that Taylor’s categorization of instructional roles of computers (tutor, tutee and tool) could be reduced to two roles where CAS are concerned since CAS are generally user-friendly and, in most cases, require no (or very little) programming language knowledge. Even in instances where programming is involved, the technology is still used as a tool to aid in the learning process. The usage of CAS in primary studies included in this meta-analysis was categorized into two groups: tutor or tool. This study sought to examine which mode of usage was more effective.

3. Location – It has been suggested that students in other developed countries were more proficient in mathematics than their peers in the US (Phillips, 2007). This study sought to examine the extent to which this proficiency could have influenced the effectiveness of CAS.

4. Educational Level - While several meta-analytic studies revealed that CAI has a positive effect on students’ performance (e.g., Liao, 1992; Kulik & Kulik, 1991; Vogel et al., 2006), those studies did not indicate the extent to which students’ performance at different academic levels could be affected by CAS. The current study attempted to fill that void.

5. Time - Some researchers reported that the effects of CBI differed across time (e.g., Timmerman & Kruepke, 2006; Christmann et al., 1997). This study sought to probe the extent to which these findings were applicable to CAS.

6. Study Design - Many researchers (e.g., Kulik and Kulik, 1991; Glass, 1982; Glass et al., 1981; Hedges, & Stock, 1983; Slavin, 1984) have cautioned about the need
to code study by design type to determine whether poorly designed studies and well-designed studies produce similar results. Independent evaluations were coded to determine if studies in which the researcher taught both the CAS group and the non-CAS group were more likely to produce favorable results than studies in which the comparison groups were taught by different instructors.

7. Evaluation Method – It was hypothesized that the experimental group had a distinct advantage over the control group in situations where this group was allowed to use technology during evaluation (i.e., quizzes, homework, test, and exam). Studies were coded by evaluation method to test the validity of this hypothesis.

8. Course Type – Given the variation in mathematics courses, the current study sought to examine whether the effect of CAS was the same across course type.

I and another doctoral candidate, with a graduate degree in Education (with concentration in program evaluation and measurement) independently coded the primary studies on each of the eight variables of interest. The inter-rater reliability from the initial coding was 93.2%. We met to review our code sheets and discuss discrepancies. Where discrepancies existed, we jointly reexamined our procedure, resolved the problems, and achieved 100% agreement.

Data Analysis Using DSTAT

Huedo-Medina, Sánchez-Mecca, and Marin-Martínez (2006) stated three general goals of data-analyses when conducting a meta-analysis: (a) to obtain an index that measures the overall effect size for a group of studies, (b) to determine whether the
studies are homogeneous, and (c) to identify sources of heterogeneity if the studies are found not to be homogeneous.

Researchers compared CAS instruction and non-CAS instruction in all studies included in this meta-analysis. Researchers also provided relevant data (means and standard deviations of the groups, t-value or F-value) that allowed for the calculation of effect sizes. DSTAT (Johnson, 1989), a statistical software developed exclusively for meta-analysis, was used for data analysis in the current study. Other diagnostic analyses and graphical displays were done using Statistical Package for the Social Sciences version 14.0.

Hunter and Schmidt (2004) identified two general frameworks for combining information from a group of studies: fixed-effects model and random-effects model. DSTAT uses a fixed-effects model.

*Fixed-effects and Random-effects Models*

Fixed- and random-effects models not only lead to different types of significance tests and confidence intervals for the mean effect size, but they also provide different tests for the effects of moderator variables. Hunter and Schmidt (2004) advised that researchers choose the model to be used based on how the results of the meta-analysis will be interpreted.

When a researcher uses a fixed-effects model, it is assumed that “the studies being analyzed are homogeneous at the level of the population effect sizes” (Hunter & Schmidt, 2004; p. 394). Stated differently, when using the fixed-effects model, the researcher is interested in generalizing the results of a meta-analysis only to future studies that are
identical to the ones included. With the fixed-effects model, it is assumed that the average effect size of the population of studies is the same for studies included in the meta-analysis. Fixed-effects models are assumed to have only one source of error variance: the variance due to the sampling of studies’ participants (Hunter and Schmidt, 2004; Rosenthal, Hoyt, Ferrin, Miller, and Cohen, 2006; Normand, 1995).

Rosenthal et al. (2006) cautioned: “if studies are treated as fixed effects, then the observed effect size tells us something about the results of these particular studies but gives no information about the generalizability of the effect sizes to future studies (or other existing studies) involving similar skills training intervention” (p. 242).

In the random-effects model, the researcher treats studies selected for inclusion in the meta-analysis as a random sample of all studies on the topic of interest and seeks to generalize findings from the meta-analysis to the entire population. Random-effects models are assumed to have two sources of error variance: variance due to sampling of studies and variance of the effect parameters (Hunter and Schmidt, 2004; Rosenthal et al., 2006; Normand, 1995; Radenbush, 1984).

**Effect Size Calculation in DSTAT**

Glass (1976) defined the effect size of a study as the mean difference between the experimental group (CAS group, in this case) and the control group (non-CAS group, in this case).

Wolf (1982) identified three different formulae for calculating effect sizes for between-group design studies—Hedges’ g, Cohen’s d, and Glass’ Δ:

\[
\text{Hedges’ } g = \frac{|M_1 - M_2|}{s_{\text{pooled}}}
\]  

(1)
Cohen’s d = \( \frac{|M_1 - M_2|}{\sigma} \)  \hspace{1cm} (2)

Glass’ Δ = \( \frac{|M_1 - M_2|}{sd_c} \)  \hspace{1cm} (3)

where \( M_1 \) is the mean of the experimental group,

\( M_2 \) is the mean of the control group,

\( sd_{pooled} \) is the pooled standard deviation of the experimental group and the control group,

\( \sigma \) is the standard deviation of the population, and

\( sd_c \) is the standard deviation of the control group.

The pooled standard deviation for equation (1) is calculated from the following formula:

\[
sd_{pooled} = \left\{ \frac{(n_e - 1)sd_e + (n_c - 1)sd_c}{n_e + n_c - 2} \right\}^{\frac{1}{2}}
\]  \hspace{1cm} (4)

where \( n_e \) is the sample size for the experimental group,

\( n_c \) is the sample size for the control group, and

\( sd_e \) is the standard deviation of the experimental group.

Hunter and Schmidt (2004) advocated for the use of equation (1) in determining the effect size because they discovered that the sampling error in the pooled standard deviation was only half of the sampling error in the control group standard deviation. DSTAT uses equation (1) to estimate effect sizes when means and standard deviations are available and this was the formula used in the current study.

In instances where a t-value was given, the program converted t to g using the formulae:

\[
g = t(2/n)^{\frac{1}{2}} \]  \hspace{1cm}  or  \hspace{1cm} (5)

\[
g = t\left[ \frac{N}{(n_e n_c)} \right]^{\frac{1}{2}}. \]  \hspace{1cm} (6)
Equation (5) was used when the sample size of the comparison groups were equal \((n_e = n_c = n)\) and (6) was used when the comparison groups had different sample sizes \((N = n_e + n_c)\) (Johnson, 1989).

When F-values were reported, DSTAT converted F to Hedges g via the formulae:

\[
g = (2F/n)^{1/2} \quad \text{or} \quad g = \{F[(n_e + n_c)/(n_en_c)]\}^{1/2}.
\]

Equation (7) was used when the sample sizes of the comparison groups were the same and equation (8) was used when the comparison groups sample sizes were unequal (Johnson, 1989).

When Hedges (1986) investigated the properties of g, he discovered that it is biased by sample size and provides an unbiased estimate of the effect size only when a study has a small sample size \((n<10)\). He proposed that, for larger samples, g must be corrected for biasness via the equation:

\[
d = g \{1 - [3/ (4n_e + 4n_c - 9)]\}
\]

where \(d\) is an unbiased estimate of the effect size for a given study. Per Hedges’ (1986) recommendation, DSTAT converted all values of g obtained using (1), (5), (6), (7) or (8) to \(d\) using (9) (Johnson, 1989).

After all effect sizes were calculated, a diagnostic analysis was conducted to identify and deal with outliers. Many researchers have cautioned of the impact of outliers on the outcome of a study. Hunter and Schmidt (2004) advised: “the presence of even a single outlier can produce a radical increase in the observed standard deviation and a somewhat smaller distortion of the mean” (p. 196). Hunter and Schmidt (2004) also
reported that deleting just the smallest and largest 2% of a data increased their ability to explain error variance by an additional 5%. Based on these findings, all extreme values were eliminated from the final analysis of the current meta-analysis.

Glass (1976) recommended that once the effect size for each comparison is obtained, meta-analysts should combine these values to determine the mean effect size for all studies. When Hedges (1986) investigated the properties of \( d \), he found that \( d \) has a normal distribution with mean zero and variance

\[
v = \left[ \frac{n_c + n_e}{n_c n_e} \right] + \left\{ \frac{d^2}{2(n_c + n_e)} \right\}.
\]

Hedges (1986) cautioned that the overall effect size should not be determined by simply finding the arithmetic mean of all effect sizes because the arithmetic mean does not take into account error variances. He advocated that if there are \( k \) comparisons, the unbiased mean effect size (\( d_{avg} \)) should be estimated using the formula

\[
d_{avg} = \frac{\sum w_i d_i}{\sum w_i} \text{ for } i = 1 \ldots k
\]

where \( w_i = 1/v_i \). The overall unbiased effect size for the current study was obtained in DSTAT through the use of (11).

**Test for Homogeneity of Effect Size**

Assuming a fixed-effects model, a test was conducted to determine whether the overall effect size could be used to represent all studies included in this meta-analysis. The test for homogeneity of effect size in DSTAT produces a \( Q \) statistic. “The \( Q \) test is computed by summing the squared deviations of each study’s effect estimate from the overall effect estimate, weighting the contribution of each study by its inverse variance” (Huedo-Medina et al., 2006; p. 194). Stated mathematically,
\[ Q = \sum_{i=1}^{k} w_i (d_i - d_{avg})^2 \]  

where all variables are as previously defined.

Q is assumed to have a chi-square distribution with k-1 degrees of freedom (k is the number of studies included in a meta-analysis) (Rosenthal et al., 2006; Normand, 1995; Johnson, 1989).

Researchers (e.g., Normand, 1995; Johnson, 1989; Wolf, 1982; Glass, 1976) found that the Q statistic is often significantly different from zero. A significant Q indicates that studies included in a meta-analysis are heterogeneous in their measures of effect sizes and further investigation is warranted to identify sources of heterogeneity.

Put another way, if the Q statistic is significant, then the overall effect size does not represent all of the studies in the population; therefore, the researcher tries to find reasons for the discrepancies. In instances where Q does not differ significantly from zero (this rarely ever happens), the researcher assumes that the mean effect size is the same across all studies in the meta-analysis (Normand, 1995; Johnson, 1989; Rosenthal et al., 2006).

Based on historical trends, it was hypothesized that the Q statistic would be significant and studies were coded a priori to determine if the overall effect size was moderated by the predefined independent variables.

**Test for Moderator Effects**

Studies were coded by moderator variables to determine which study characteristics were responsible for heterogeneity of the overall effect size. DSTAT
partitions the Q statistic in (12) into two sources of error variance: a between-group Q statistic ($Q_B$) and a within-group Q statistic ($Q_W$). Stated algebraically,

$$Q = Q_B + Q_W.$$ \hspace{1cm} (13)

The partitioning of Q in DSTAT is analogous to the partitioning of the sum of the square variances in ANOVA. A significant $Q_B$ suggests a significant difference between/among different levels of a moderator variable. If an independent variable has more than two levels, DSTAT includes a menu to conduct post-hoc analysis. The $Q_B$ obtained is assumed to have a chi-square distribution with p-1 degrees of freedom, where p is the number of levels of a moderator variable (Huedo-Medina et al., 2006; Rosenthal et al., 2006; Johnson, 1989; Hedges & Olkin, 1985).

A significant $Q_W$ suggests significant differences within a given level of the moderator variable. $Q_W$ has an approximately chi square distribution with m-1 degrees of freedom (m is the number of studies within a subgroup) (Johnson, 1989; Hedges & Olkin, 1985; Rosenthal et al., 2006). Using DSTAT, $Q_B$ and $Q_W$ were estimated for all predetermined moderator variables included in this meta-analysis.

**Fail-Safe N Calculation**

In addition to investigating whether published studies were more likely to report statistically significant differences between the comparison groups than unpublished studies, the fail-safe N (the number of nonsignificant studies needed to reverse the conclusion of a quantitative synthesis) was calculated (Glass, 1989).
This meta-analysis used Orwin’s (1983) formula to calculate the fail-safe, $N_{fs}$:

$$N_{fs} = \frac{N (d_1 - d_2)}{d_2}$$

(14)

where $N$ is the number of studies in the meta-analysis,

d$_1$ is the average effect size of the studies included in the meta-analysis,

and, d$_2$ is a value selected that d$_1$ would be equivalent to when a given number of studies is added to the meta-analysis.

Orwin also recommended that researchers use any of Cohen’s (1977) d values for d$_2$ (e.g., 0.2 (small effect), 0.5 (medium effect) or 0.8 (large effect)) but there is no standard value of d$_2$ used in the $N_{fs}$. Stated another way, if d$_2 = 0.2$ and $N_{fs}$ was found to be 20, then 20 studies are needed to reduce the average effect size (d$_1$) to 0.2. The value of d$_2$ was set to 0.001 to allow for the calculation of the number of studies in the file drawers that are needed to bring the average effect size close to zero.

**Summary**

This chapter provided a detailed outline of how studies included in this meta-analysis were located, retrieved, coded and analyzed. Also included was a description of the model selected and mathematical equations used to arrive at quantitative values of interest.
CHAPTER IV
RESULTS

Introduction

The purpose of this study was to investigate the overall effectiveness of CAS, in comparison to non-CAS instruction, on students’ achievement in mathematics at pre-college and post-secondary institutions. The study utilized meta-analysis on a group of primary studies that individually investigated the effectiveness of CAS instruction on students’ achievement in mathematics, and explored the extent to which this overall effectiveness of CAS was moderated by various study characteristics. This meta-analysis was guided by two research questions:

1. What is the overall effectiveness of CAS instruction on students’ achievement in mathematics, in comparison to non-CAS instruction, at pre-college and post-secondary institutions?

2. Does the effectiveness of CAS on students’ achievement in mathematics differ by type of publication (published or unpublished), type of usage (tutor or tool), location (US or international), time (1980-1989, 1990-1999, or 2000-2007), educational level (pre-college or postsecondary), study design (studies that controlled for the effect of teacher or studies that did not control for the effect of teacher), evaluation method (studies in which the experimental group used CAS during evaluation or studies in which the experimental group did not
use CAS during evaluation), or course-type (algebra, calculus, or other)?

The research questions were investigated via the following null hypotheses:

1. There is no significant difference between the effectiveness of CAS instruction and the effectiveness of non-CAS instruction in mathematics at pre-college and post-secondary institutions.

2. The effectiveness of CAS on students’ achievement does not differ by type of publication (published or unpublished).

3. The effectiveness of CAS on students’ achievement does not differ by type of usage (tutor or tool).

4. The effectiveness of CAS on students’ achievement does not differ by location (US or international).


6. The effectiveness of CAS on students’ achievement does not differ by educational level (pre-college or postsecondary).

7. The effectiveness of CAS on students’ achievement does not differ by study design (studies that controlled for the effect of teacher vs. studies that did not control for the effect of teacher).

8. The effectiveness of CAS on students’ achievement does not differ by evaluation method (studies in which the experimental group used CAS
during evaluation or studies in which the experimental group did not use CAS during evaluation).

9. The effectiveness of CAS on students’ achievement does not differ by course-type (algebra, calculus, or other mathematics courses).

All hypotheses were tested at the 5% significance level. The rest of this chapter contains descriptive analyses of effect sizes and results found from testing the null hypotheses.

Descriptive Analyses of Effect Sizes

Our search of the literature yielded 31 studies in which CAS were compared to traditional. There were 107 comparisons of CAS instruction and non-CAS instruction in these primary studies; 14 (13%) favored non-CAS instruction while the remaining 93 (87%) favored CAS instruction. The stem-and-leaf plot in Figure 1 provides a breakdown of all effect sizes obtained. A detailed list of each study, along with the number and range of effect sizes, is provided in Appendix 1. Effect sizes coded by levels of the independent variables can also be found in Appendix 2.

```
-2. 3
-0. 11111122234
  0. 00011111111111111111
  0. 2222222333333333
  0. 444444444444555555555
  0. 66666666666677777778889999
  1. 00011122688
  2. 5
```

*Figure 1.* Stem-and-leaf plot of all effect sizes.

Initial box and whiskers plot of all effect sizes revealed 4 extreme effect sizes (Figure 2) in Li and Edmonds (2005), Tiwari (1999), Runde (1997), and Campbell
(1994). These extreme values were temporarily removed from the dataset and a second box plot was done to determine whether all potential outliers had been identified. Further analysis revealed another extreme value in Portis (1991). This extreme value was removed from the analysis and the subsequent plot (Figure 3) suggested that all extreme values had been identified. Studies containing extreme values were reexamined.

Figure 2. Box and whiskers plot of all effect sizes.
The investigation of studies with potential outliers did not reveal any errors in data extraction and/or in calculation the effect sizes. All 5 extreme effect sizes were eliminated from the analysis (102 effect sizes were included in the final analysis). A breakdown of the number of effect sizes by levels of the independent variables is presented in Table 3.
Table 3

*Number of Effects Sizes per Category of Moderator Variable*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Categories</th>
<th>Number of Effect Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of Publication</td>
<td>published</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>unpublished</td>
<td>46</td>
</tr>
<tr>
<td>Type of Usage</td>
<td>tutor</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>tool</td>
<td>80</td>
</tr>
<tr>
<td>Location</td>
<td>US</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>International</td>
<td>17</td>
</tr>
<tr>
<td>Educational Level</td>
<td>pre-college</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>postsecondary</td>
<td>67</td>
</tr>
<tr>
<td>Time</td>
<td>1980-1989</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>1990-1999</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>2000-2007</td>
<td>47</td>
</tr>
<tr>
<td>Study Design</td>
<td>same Instructor</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>different Instructor</td>
<td>27</td>
</tr>
<tr>
<td>Evaluation Method</td>
<td>used CAS</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>did not use CAS</td>
<td>76</td>
</tr>
<tr>
<td>Course Type</td>
<td>algebra</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>calculus</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>13</td>
</tr>
</tbody>
</table>
Overall Effectiveness of CAS

Hypothesis 1

The overall effectiveness of CAS was examined using null hypothesis 1: There is no significant difference between the effectiveness of CAS instruction and the effectiveness of non-CAS instruction in mathematics at secondary and post-secondary institutions. A total of 102 comparisons (N=7,342) of CAS and non-CAS instruction were analyzed. The effect sizes (after removing extreme values) ranged from −0.47 to 1.23. The overall corrected unbiased effect size was 0.38. The lower bound (0.34) and the upper bound (0.43) of the average effect size hinted that this value was significantly different from zero and prompted the rejection of hypothesis 1.

The mean effect size, 0.38, also suggests that if a regular student exposed to CAS instruction was given the same exam as a regular student taught using non-CAS instruction, the former would score at the 65th percentile while the latter would score at the 50th percentile (Figure 4). The fail-safe N, $N_{fs}$, hinted that about 11,749 studies with non-significant results would be needed to nullify the current finding (i.e., reduce the effectiveness of CAS instruction vs. non-CAS instruction from 0.38 to 0.001).

There was evidence to indicate significant heterogeneity across the 102 effect size measures, $Q_{sh}(101) = 225.60, p < 0.05$. This triggered the need to further explore the extent to which the overall effectiveness of CAS instruction was moderated by the hypothesized study characteristics.
Test for Moderator Variables

Hypothesis 2

Research reports were coded by published and unpublished studies to investigate if publication bias was present (i.e., whether published studies were more likely to produce positive results than unpublished studies). The presence, or lack thereof, of publication bias was examined via null hypothesis 2: The effectiveness of CAS on students’ achievement does not differ by type of publication (published or unpublished).

While there was evidence to suggest heterogeneity within published studies ($Q_w(55) = 118.69, p < 0.05$) and unpublished ($Q_w(45) = 106.86, p < 0.05$) studies, the average effect size for published studies ($d = 0.38$) did not differ significantly from the average effect size for unpublished studies ($d = 0.39$), $Q_B(1) = 0.05, p > 0.05$. It appeared that any observed differences between the two groups of studies may have been due to sampling error; therefore, null hypothesis 2 was not rejected. A box plot of effect size by type of publication is presented in Figure 5.
Hypothesis 3

Independent evaluations were coded by type of usage to explore if studies in which the subjects/participants used CAS as a tutor were more likely to produce results significantly different from studies in which the subjects/participants used CAS as a tool. The effect of type of CAS usage on students’ performance in math was probed using null hypothesis 3: The effectiveness of CAS on students’ achievement does not differ by type of usage (tutor or tool).

Figure 5. Box and whiskers plot of effect sizes by type of publication.
The average effect size for CAS in a tutorial role (d = 0.40) did not differ significantly from the average effect size for CAS in a tool role (d=0.39), $Q_b(1) = 0.007$, $p > 0.05$. This finding prompted the failure to reject null hypothesis 3; however, studies in which participants used CAS as a tutor ($Q_w(21) = 35.46, p < 0.05$) and studies in which participants used CAS as a tool ($Q_w(79) = 175.82, p < 0.05$) reported significant heterogeneity in their measures of effect sizes. Figure 6 provides a box plot of effect sizes by type of usage.

*Figure 6.* Box and whiskers plot of effect size by type of usage.
**Hypothesis 4**

Because of reported differences in students’ performance between the US and other countries, primary studies were coded to determine whether the effectiveness of CAS was moderated by location. The effect of location was investigated using null hypothesis 4: The effectiveness of CAS on students’ achievement does not differ by location (US or international). A box plot of effect size by location is presented in Figure 7.

*Figure 7. Box and whiskers plot of effect size by location*
While the average effect size for studies conducted in the US \((d = 0.37)\) was less than the average effect size for studies conducted internationally \((d = 0.48)\), these values were not significantly different, \(Q_{B}(1) = 1.74, p > 0.05\). Null hypothesis 4 was not rejected.

It appeared that there was significant heterogeneity within the effect size measures for studies conducted in the US \(Q_{w}(84) = 202.73, p < 0.05\). On the contrary, international studies did not report significantly different effect sizes \(Q_{w}(16) = 21.14, p > 0.05\).

**Hypothesis 5**

To determine whether the effect of CAS was moderated by time, primary studies were coded and examined. This examination was conducted using null hypothesis 5: The effectiveness of CAS on students’ achievement does not differ by time (1980-1989, 1990-1999, or 2000-2007).

Studies conducted between 1980 and 1989 did not report significant heterogeneity in their measures of effect size \(Q_{w}(13) = 23.24, p > 0.05\); however, there was evidence to suggest significant heterogeneity within the effect size measures for studies conducted from 1990 to 1999 \(Q_{w}(40) = 83.15, p < 0.05\), and from 2000 to 2007 \(Q_{w}(46) = 91.33, p < 0.05\). Additionally, it appeared that the effect size measures differed significant across time, \(Q_{B}(2) = 27.88, p < 0.05\). This finding prompted the need to reject null hypothesis 5 and conduct a post-hoc analysis.

DSTAT post-hoc analysis revealed that the average effect size for studies conducted from 1990 to 1999 \((d = 0.51)\) was significantly larger than the average effect
size for studies conducted from 2000 to 2007 ($d = 0.24$), $\chi^2(1) = 27.78, p < 0.05$. While no other pair of comparison was significant, the difference between the average effect size for studies conducted in the 1980’s ($d = 0.34$) was less than that of studies conducted from 1990 to 1999. A box plot of effect size by time is presented in Figure 8.

*Figure 8.* Box and whiskers plot of effect size by time
Hypothesis 6

The extent to which the overall effect size was moderated by education level was investigated using null hypothesis 6: The effectiveness of CAS on students’ achievement does not differ by educational level (pre-college or postsecondary). A box plot of effect size by educational level is presented in Figure 9.

Figure 9. Box and whiskers plot of effect size by educational level

The average effect size for studies conducted on pre-college students \( (d = 0.35) \) did not differ significantly from the average effect size for postsecondary studies \( (d = \)
$Q_B(1) = 0.65, p > 0.05$; null hypothesis 6 was not rejected. Though there was significant within-group heterogeneity for studies conducted at the post-secondary level ($Q_w(66) = 181.42, p < 0.05$), studies conducted on pre-college students were not significantly heterogeneous ($Q_w(34) = 45.54, p > 0.05$).

**Hypothesis 7**

Independent evaluations were coded by study design to investigate if studies in which the researcher taught both the CAS group and the non-CAS group were likely to produce more favorable results than studies in which the comparison groups were taught by different instructors. The extent to which the overall effect size was moderated by study design was probed via null hypothesis 7: The effectiveness of CAS on students’ achievement does not differ by study design (studies that controlled for the effect of teacher vs. studies that did not control for the effect of teacher).

The average effect sizes for studies that controlled for the effect of teacher (different teachers) and studies that did not control for the effect of teacher (same teacher) were found to be 0.41 and 0.30, respectively. The difference between these values approached significance, $Q_B(1) = 3.95, p = 0.05$; null hypothesis 7 was rejected. Additionally, both studies that controlled for the effect of teacher ($Q_w(74) = 177.61, p < 0.05$), and studies that did not control for the effect of teacher ($Q_w(26) = 44.04, p < 0.05$) revealed significant heterogeneity in their measures of effect size. A box plot of effect size by study design is presented in Figure 10.
Research reports in this meta-analysis were also coded by evaluation method to determine whether studies in which the experimental group used CAS during evaluation were likely to produce results significantly different from studies in which the experimental group did not use CAS during evaluation.

The extent to which the overall effect size was moderated by evaluation method was investigated via null hypothesis 8: The effectiveness of CAS on students’
achievement does not differ by evaluation method (studies in which the experimental group used CAS during evaluation or studies in which the experimental group did not use CAS during evaluation). A box plot of effect size by evaluation method is presented in Figure 11.

![Box and Whiskers Plot](image)

**Figure 11.** Box and whiskers plot of effect size by evaluation method

The average effect size for studies in which CAS were used during evaluation ($d = 0.31$) was significantly lower than the average effect size for studies in which CAS were not used during evaluation ($d = 0.42$), $Q_{df}(1) = 4.35, p < 0.05$; null hypothesis 8 was
rejected. Additionally, both studies in which CAS were used by the experimental group during evaluation \((Q_w(25) = 70.34, p < 0.05)\) and studies in which CAS were not used by the experimental group during evaluation \((Q_w(75) = 136.59, p < 0.05)\) showed significant heterogeneity in their measures of effect sizes.

**Hypothesis 9**

The extent to which the overall effect size was moderated by course type was investigated using null hypothesis 9: The effectiveness of CAS on students’ achievement does not differ by course-type (algebra, calculus, or other mathematics courses). The average effect for studies that involved the teaching of algebra, calculus, and other mathematics courses were found to be 0.38, 0.43, and 0.26, respectively. These values did not differ significantly, \(Q_\beta(2) = 4.73, p > 0.05\); null hypothesis 9 was not rejected. There was evidence to suggest the presence of heterogeneity within the effect size measures for algebra \((Q_w(63) = 143.17, p < 0.05)\), calculus \((Q_w(24) = 52.27, p < 0.05)\), and other mathematics courses \((Q_w(12) = 25.43, p > 0.05)\). A box plot of effect size by course type is presented in Figure 12.
This chapter began with a descriptive overview of effect size measures and presented results found from investigating the 9 null hypotheses in the study. The overall effect size, 0.38, was significantly different from zero; a finding which prompted further examination of this value to determine the extent to which it was moderated by the various independent variables. Three of the independent variables (study design, time,
and evaluation method) were found to be significantly related to the average effect of CAS.
CHAPTER V

DISCUSSION

Introduction

This chapter summarizes findings from the current study, provides discussion/implications of the findings, and offers suggestions for future research.

Summary of Findings

Students’ performance in mathematics is an ongoing global concern for both educators and policy makers. Recent scoring trends of students on standardized exams continue to point to the nagging problem. This situation has compelled many individuals and organizations to call for modified methods to improve the learning of mathematics.

Many have suggested that the use of technology in the classroom could increase students’ learning of mathematics without compromising their knowledge acquisition. CAS have been introduced in classrooms to do just that, but findings about their effectiveness have been controversial. While one group of researchers reported that students taught using CAS instruction performed better than students taught using non-CAS instruction, others found no differences between the comparison groups. Some even hinted that students exposed to non-CAS instruction tend to acquire more mathematical skills than their peers taught using CAS. These contradicting findings mitigated the need to quantitatively combine studies that compared CAS and non-CAS instruction.

The purpose of this study was to investigate the overall effectiveness of CAS, in comparison to non-CAS instruction, on students’ achievement in mathematics at pre-
college and post-secondary institutions. The study utilized meta-analysis on a group of primary studies that individually investigated the effectiveness of CAS on students’ achievement in mathematics, and explored the extent to which this overall effectiveness of CAS was moderated by various study characteristics.

*Overall Effectiveness of CAS*

Findings from the current study revealed that students using CAS tend to perform better than students taught using non-CAS instruction. This finding is consistent with previous findings (e.g., Kulik et al., 1980; Roblyer, 1989; Liao, 1992; Khalili and Shashaani, 1994; Christmann et al., 2001; Vogel et al., 2006) on the effect of technology on education.

The overall unbiased effect size for CAS, when compared to non-CAS instruction, was found to be 0.38 and this value was significant different from zero. The magnitude and direction of this effect size implies that students taught using CAS were likely to perform better than 65% of the students receiving only non-CAS instruction.

The value for the overall effectiveness of CAS is within the range of results documented in previous meta-analyses that focused on the effect of technology on education. Effect sizes for technology found in a review of existing literature ranged from 0.12 (Timmerman and Kruepke, 2006) to 0.52 (Schmidt et al., 2001). This suggests that the difference between the effectiveness of technology, when compared to non-CAS instruction, ranged between 5 and 20 percentile points (in favor of technology instruction). The present study found that students in the experimental group gained 15 percentile points over their peers in the control group.
Moderator Variables

The failure of the overall effect size to meet the assumption of homogeneity prompted an investigation into the presence of moderator variables. Atkinson et al. (1982) inferred that published studies were about three times more likely to report statistically significant results than unpublished studies.

In their 1991 meta-analytic review of studies from online sources on the effectiveness of CBI, Kulik and Kulik detected a significant presence of publication bias in favor of published studies. Their study covered several content areas including math, science, reading, language arts, and social science. In a meta-analysis on the effects of CAI, Liao (1992) hinted that ERIC documents tend to report significantly higher effect sizes than published studies or dissertations.

Findings from this study did not reveal the presence of publication bias for studies on the effectiveness of CAS. It appeared that unpublished studies were more likely to produce positive results in favor of CAS than published studies, though this difference was not statistically significant.

Some researchers sought to determine whether the effect of CBI differed based on how the technology is used during instruction. Liao (1992) reported that tutorial software were significantly more effective than drill-and practice software, problem-solving software, and simulation software. Kulik et al. (1985) also found CAI to be significantly more effective than CMI.

The current study did not find significant effect for type of usage. This finding is consistent with findings from Roblyer (1989) who compared drill-and-practice programs
and tutorial programs, and found no significant differences between them. Additionally, Kulik et al. (1980) earlier reported no significant differences among tutorial programs, computer-managed teaching programs, and simulation programs.

I also discovered that students in the US who were exposed to CAS were not significantly more or less likely to perform better than their peers in other countries who were taught using CAS. Moreover, the effectiveness of CAS was not significantly related to type of mathematics course or educational level.

Barayktar (2001) reported that the effect of CAI was not moderated by educational level, and this result is consistent with the current findings. Elsewhere, Kulik et al. (1985) and Roblyer (1989) found a significant relationship between the effect of technology and educational level.

Some researchers suggested a novelty effect for technology on education; i.e., the tendency for students’ performance to initially improve when a new technology is introduced, not because of any actual improvement in learning but in response to increased interest in the new technology. Christmann et al. (2001) found that the average effect of technology was significantly lower for studies done between 1990 and 1995 than it was for studies done between 1984 and 1990. Bayraktar (2001) indicated that the average effect of CAI reduced significantly from 1970 – 1979 to 1980 – 1989 and also from 1980 – 1989 to 1990 – 1999. Timmerman and Kruepke (2006) reported that the effect of technology diminished significantly when they compared studies conducted between 1985 and 1994, and studies conducted between 1995 and 2004. In contrast, Kulik (1983) reasoned that the effect of CBI should increase with time since
technological advancements were expected to bring in newer innovations to increase flexibility, capabilities, and applications.

Results from the present study were in agreement with reports from Christmann et al. (2001), Bayraktar (2001), and Timmerman and Kruepke (2006). The mean effect size for studies conducted between 1980 and 1989 was smaller than the mean effect size for studies conducted between 1990 and 1999; this difference was not significant. On the other hand, the mean effect size for studies conducted between 1990 and 1999 was significantly larger than the mean effect size than studies conducted between 2000 and 2007.

The current result also offers some speculations about the trend of CAS use. It appeared that the effect of CAS was lower in the 1980’s when the technology was just introduced to the classroom. This could be attributed to the time taken by both instructors and students to become familiar with the new technology. Additionally, researchers needed time to be familiar with the new system before they could investigate its effectiveness. This hunch is buttressed by the fact that only three of the studies in this meta-analysis were conducted in the 1980’s.

By 1990, the knowledge about the presence of CAS may have spread to other practitioners thus leading to more studies (19 of the studies in this meta-analysis) been conducted. Between 2000 and 2007, studies began reporting smaller effect sizes suggesting the possibility of a novelty effect. The extent to which this trend could continue is a topic for future research.
One surprising discovery was the relationship between evaluation method and the overall effect size. A reasonable individual might assume that students in the experimental group would perform better if they were allowed to use CAS during evaluation than if they were not allowed to use CAS during evaluation. The current study suggested a different result. Experimental group students who did not use CAS during evaluation were more likely to perform better than their other peers in the experimental group who used CAS during evaluation.

Another phenomenon that has been a topic of prominent discussion among researchers who have investigated the effectiveness of technology is the moderating tendency of teacher effect. Kulik et al. (1980), and Kulik and Kulik (1991) stated that studies in which the experimental group and the control were taught by the same instructor tended to report significantly higher effect sizes than studies in which the comparison groups were taught by different instructors. In contrast, Bayraktar (2001), and Khalili and Shashaani (1994) found no significant teacher effect.

The current investigation discovered significant teacher effect: studies in which the comparison groups were taught by the same teacher reported significantly higher effect sizes than studies that in which the comparison groups were taught by different instructors. This finding was consistent with results from Kulik et al. (1980), and Kulik and Kulik (1991).

Limitations

It is important to mention that the primary studies included in this meta-analysis were retrieved only from online research databases and search engines. Further, a fixed-
effects model was used for data analysis. Any generalization beyond studies meeting the predefined search criteria must be done with caution. Additionally, meta-analysis is a technique that quantitatively summarizes findings from previous studies. These results do not predict future trends on the effectiveness of CAS, in particular, or technology, in general.

An area of caution is the type of evaluation that was used by the various researchers to measure achievement. For example, Runde (1997) used a conceptual (non-multiple choice) word problem exam to measure achievement while Stephens and Konvalina (1998) administered a multiple choice exam to measure the same variable. Elsewhere, Hollard and Norwood measured achievement by means of a departmental final exam while Ford and Klicka (1994) used Accuplacer (a standardized exam designed by the College Board which is generally used for placement into college courses) to measure the variable. It can be argued that all of these exams may not have measured the same construct. Researchers with larger samples could investigate the extent to which there is homogeneity, or lack thereof, among studies using different tests to measure achievement.

The lack of within-group homogeneity is also an issue of concern. A case in point: I did not find any significant effect for location or education level but an investigation into location by educational level revealed that the mean effect of CAS on achievement for international postsecondary students ($d = -0.46$) could be significantly higher than the mean effect of CAS on achievement for US postsecondary students ($d = 0.39$). However, this phenomenon could not be thoroughly investigated because I found
only one international study (with one effect size) that examined the effectiveness of CAS at the postsecondary level. Elsewhere, Roblyer (1989) discovered that technology had a lesser effect on learning outcomes of students in secondary school than it had for students in elementary school. This discovery could probably be a source of heterogeneity in the effect sizes within the pre-college level.

Moreover, there is a multiplicity of CAS software available in the marketplace. MATHEMATICA, DERIVE, ALEKS, muMath/muSimp-83, MACSYMA, Academic System software, FUNdamentallyMATH, and Texas Instruments are but a few from this continuously expanding list of software options. A finding that CAS is more effective than non-CAS instruction does not indicate if there were differences among the various software packages, and this could be another source within group heterogeneity.

While this study did not find significant moderating effect for the independent variables (except for time, design, and evaluation method) the extent to which there was heterogeneity within groups or interaction between/among independent variables should be an area of investigation for future researchers with large enough samples.

Implications

In the debate on mathematical instruction, both traditionalists and constructivists agree that there is a need for mathematical reform to improve students learning. The quest for a more viable way to deliver mathematics education is ongoing. It is important that students who acquire academic skills, both at the pre-college and postsecondary educational levels, be adequately prepared to meet the demands of a technologically savvy workplace. Ignoring the presence of technological tools that have the propensity to
increase students’ acquisition of knowledge is a disservice to both the students and the larger society. I believe the discussion is beyond whether technology should be used in the classroom but how it can be imbedded in our pedagogy so that the learning of academic skills is enhanced.

These findings could serve as resource for policy makers who are considering adopting CAS as options for content delivery especially when considering that CAS appear to allow students to do more complicated problems than they would have otherwise been able to do by hand, and enable them to generate and manipulate symbolic expressions that were too time-consuming and complicated (Heid, 1997; Heid & Edwards, 2001; Heid, 2002).

In their 2001 study on how Logo (a computer program) affects students’ learning, Clements, Battista, and Sarama suggested: (a) students generally tend to master concepts emphasized by their teachers, (b) little learning is achieved in the Logo environment in the absence of teacher’s guidance, (c) “Dialogue between teacher and students is essential for encouraging predicting, reflecting, and higher level reasoning” (p.9), and (d) studies that showed positive effects for Logo depended on carefully planned activities.

This line of reasoning concerning Logo could be extended where CAS are concerned. Teachers and other professionals must recognize that CAS are merely resources to aid in the learning process. The optimum benefits of these systems may only be realized in a carefully planned learning environment that fosters dialogue between teachers and learners, and emphasizes the core curricula concepts that students need to acquire. Students and instructors have to use thinking and logical skills to interpret the
results from their computation. At the very least, both students and instructors in CAS environment should go beyond just doing a computation to inquiring why the computation was necessary.

Those who decide to adopt CAS must also take into account that, except for one of the studies in this analysis, all other researchers used CAS to supplement traditional instruction. Kulik and Kulik (1991) found that the effect of CBI was significantly higher when the duration of treatment was 4 weeks or less than when the length of time was longer.

The criticisms of CAS should not be ignored. Practitioners and policy makers are cautioned that, as it is the case with technological tools, it would require a learning curve and a change in classroom practices to incorporate CAS. The first issue is the amount of time to learn how and when to use CAS in the classroom. The second issue is getting students to develop essential mathematical skills. Finally, there is a need to form a marriage between the use of CAS and the acquisition of mathematical skills.

Limitations aside, the findings from the current study tend to suggest that CAS have the potential to improve learning in the classroom. Regardless of how CAS were used (tutor or tool), the current study found that CAS contributed to a significant increase in students’ performance. Additionally, even when CAS were taken from students during evaluation, the experimental group still performed better than their peers in the control group.

It is critically important that students, at all educational levels, learn technological skills (in doses commensurate with their academic needs), but this must not be done in
disregard of their need to learn computational skills. At present, almost all standardized exams demand computational skills. Even if this trend were to vanish, both computational skills and logical thinking skills necessary to assist students learn mathematics.

Conclusion

The current study quantitatively combined results from primary studies that compared the effectiveness of CAS instruction against the effectiveness of non-CAS instruction. Results indicated that students exposed to CAS were likely to perform better than their peers taught using non-CAS instruction if both groups were given a common mathematics exam.

Constructivists and traditionalists, and those along the philosophical continuum, must recognize that technology is here to stay. Our continuous growth depends on the degree to which we embrace it. We must strive to make technology our friend and not our foe. In this era of a continuous search for viable ways to assist learners, CAS could serve as one of the alternatives to supplement learning in mathematics classrooms.

However, as it is with many things in nature, over-reliance could be harmful. While we strive to educate technologically literate citizens who will meet the demands of the modern workplace, we should also recognize that there is need to strike a balance between equipping future employees with both computational skills and technical skills. We must not slay one mathematical skill on the academic alter at the expense of the other.
APPENDICES
APPENDIX A

Reference List of Studies Included in the Meta-analysis


APPENDIX B

Number and Range of Effect Sizes for Independent Studies
Table 4

*Number and Range of Effect Sizes for Independent Studies*

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of Effect sizes</th>
<th>Effect size range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batchelder (2000)</td>
<td>1</td>
<td>0.334</td>
</tr>
<tr>
<td>Brown (2000)</td>
<td>6</td>
<td>–0.024 to 0.748</td>
</tr>
<tr>
<td>Campbell (1996)</td>
<td>2</td>
<td>1.107 to 1.615</td>
</tr>
<tr>
<td>Chen &amp; Lui (2007)</td>
<td>1</td>
<td>0.706</td>
</tr>
<tr>
<td>Cooley (1997)</td>
<td>8</td>
<td>0.101 to 0.742</td>
</tr>
<tr>
<td>Cooley (1995)</td>
<td>2</td>
<td>0.416 to 0.488</td>
</tr>
<tr>
<td>Fletcher et al. (1990)</td>
<td>2</td>
<td>0.378 to 0.461</td>
</tr>
<tr>
<td>Ford &amp; Klicka (1994)</td>
<td>1</td>
<td>0.593</td>
</tr>
<tr>
<td>Gesshel-green (1986)</td>
<td>9</td>
<td>0.051 to 0.616</td>
</tr>
<tr>
<td>Graham &amp; Thomas (2000)</td>
<td>2</td>
<td>0.724 to 0.730</td>
</tr>
<tr>
<td>Hagerty &amp; Smith (2005)</td>
<td>5</td>
<td>–0.146 to 0.619</td>
</tr>
<tr>
<td>Hamide (2001)</td>
<td>10</td>
<td>–0.340 to 0.673</td>
</tr>
<tr>
<td>Hawker (1986)</td>
<td>1</td>
<td>0.164</td>
</tr>
<tr>
<td>Hollard &amp; Norwood (1999)</td>
<td>4</td>
<td>0.597 to 1.06</td>
</tr>
<tr>
<td>Hopkins &amp; Kinard (1998)</td>
<td>1</td>
<td>0.755</td>
</tr>
<tr>
<td>Keepers (1995)</td>
<td>3</td>
<td>0.015 to 0.14</td>
</tr>
<tr>
<td>Kramarski et al. (2006)</td>
<td>8</td>
<td>0.080 to 0.754</td>
</tr>
<tr>
<td>Leng et al. (2005)</td>
<td>1</td>
<td>–0.467</td>
</tr>
<tr>
<td>Study</td>
<td>Number of Effect sizes</td>
<td>Effect size range</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Li &amp; Edmonds (2005)</td>
<td>6</td>
<td>0.360 to 2.479</td>
</tr>
<tr>
<td>Melin-Conejeros (1992)</td>
<td>6</td>
<td>0.088 to 0.61</td>
</tr>
<tr>
<td>Palmiter (1986)</td>
<td>4</td>
<td>–0.124 to 1.002</td>
</tr>
<tr>
<td>Palmiter (1991)</td>
<td>2</td>
<td>0.935 to 1.031</td>
</tr>
<tr>
<td>Portis (1991)</td>
<td>2</td>
<td>0.716 to 1.830</td>
</tr>
<tr>
<td>Power et al. (2005)</td>
<td>4</td>
<td>0.117 to 0.392</td>
</tr>
<tr>
<td>Runde (1998)</td>
<td>3</td>
<td>0.179 to 1.131</td>
</tr>
<tr>
<td>Runde (1997)</td>
<td>1</td>
<td>–2.277</td>
</tr>
<tr>
<td>Shaw et al (1997)</td>
<td>1</td>
<td>0.571</td>
</tr>
<tr>
<td>Stephens &amp; Kovalina (1999)</td>
<td>2</td>
<td>0.403 to 0.498</td>
</tr>
<tr>
<td>Tab (2001)</td>
<td>4</td>
<td>–0.248 to 0.428</td>
</tr>
<tr>
<td>Tiwari (1999)</td>
<td>3</td>
<td>0.134 to 1.217</td>
</tr>
<tr>
<td>Trout (1993)</td>
<td>2</td>
<td>0.299 to 1.229</td>
</tr>
</tbody>
</table>
APPENDIX C

Studies Coded by Levels of the Independent Variables
Table 5

*Studies Coded by Publication Type, Type of Usage, and Location*

<table>
<thead>
<tr>
<th>Study</th>
<th>Publication Type</th>
<th>Type of Usage</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batchelder (2000)</td>
<td>published</td>
<td>tutor</td>
<td>US</td>
</tr>
<tr>
<td>Campbell (1996)</td>
<td>unpublished</td>
<td>tool</td>
<td>US</td>
</tr>
<tr>
<td>Chen &amp; Lui (2007)</td>
<td>published</td>
<td>tutor</td>
<td>International</td>
</tr>
<tr>
<td>Cooley (1997)</td>
<td>published</td>
<td>tool</td>
<td>US</td>
</tr>
<tr>
<td>Fletcher et al. (1990)</td>
<td>published</td>
<td>tutor</td>
<td>US</td>
</tr>
<tr>
<td>Ford &amp; Klicka (1994)</td>
<td>unpublished</td>
<td>tutor</td>
<td>US</td>
</tr>
<tr>
<td>Gesshel-green (1986)</td>
<td>unpublished</td>
<td>tool</td>
<td>US</td>
</tr>
<tr>
<td>Graham &amp; Thomas (2000)</td>
<td>unpublished</td>
<td>tool</td>
<td>International</td>
</tr>
<tr>
<td>Hagerty &amp; Smith (2005)</td>
<td>published</td>
<td>tutor</td>
<td>US</td>
</tr>
<tr>
<td>Hawker (1986)</td>
<td>unpublished</td>
<td>tool</td>
<td>US</td>
</tr>
<tr>
<td>Kramarski et al. (2006)</td>
<td>published</td>
<td>tool</td>
<td>International</td>
</tr>
<tr>
<td>Leng et al. (2005)</td>
<td>published</td>
<td>tool</td>
<td>International</td>
</tr>
<tr>
<td>Study</td>
<td>Publication Type</td>
<td>Type of Usage</td>
<td>Location</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------</td>
<td>---------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Li &amp; Edmonds (2005)</td>
<td>published</td>
<td>tutor</td>
<td>International</td>
</tr>
<tr>
<td>Palmiter (1986)</td>
<td>unpublished</td>
<td>tool</td>
<td>US</td>
</tr>
<tr>
<td>Palmiter (1991)</td>
<td>published</td>
<td>tool</td>
<td>US</td>
</tr>
<tr>
<td>Power et al. (2005)</td>
<td>published</td>
<td>tool</td>
<td>US</td>
</tr>
<tr>
<td>Runde (1997)</td>
<td>unpublished</td>
<td>tool</td>
<td>US</td>
</tr>
<tr>
<td>Stephens &amp; Kovalina (1999)</td>
<td>published</td>
<td>tool</td>
<td>US</td>
</tr>
<tr>
<td>Tab (2001)</td>
<td>published</td>
<td>tool</td>
<td>US</td>
</tr>
<tr>
<td>Tiwari (1999)</td>
<td>unpublished</td>
<td>tool</td>
<td>US</td>
</tr>
</tbody>
</table>
Table 6

*Studies Coded by Time, Educational level, and Course Type*

<table>
<thead>
<tr>
<th>Study</th>
<th>Time</th>
<th>Educational level</th>
<th>Course type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chen &amp; Lui (2007)</td>
<td>2000-2007</td>
<td>pre-college</td>
<td>other</td>
</tr>
<tr>
<td>Cooley (1997)</td>
<td>1990-1999</td>
<td>postsecondary</td>
<td>calculus</td>
</tr>
<tr>
<td>Fletcher et al. (1990)</td>
<td>1990-1999</td>
<td>pre-college</td>
<td>other</td>
</tr>
<tr>
<td>Kramarski et al. (2006)</td>
<td>2000-2007</td>
<td>pre-college</td>
<td>algebra</td>
</tr>
<tr>
<td>Leng et al. (2005)</td>
<td>2000-2007</td>
<td>postsecondary</td>
<td>algebra</td>
</tr>
<tr>
<td>Study</td>
<td>Time</td>
<td>Educational level</td>
<td>Course type</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------</td>
<td>-------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Li &amp; Edmonds (2005)</td>
<td>2000-2007</td>
<td>pre-college</td>
<td>algebra</td>
</tr>
<tr>
<td>Portis (1991)</td>
<td>1990-1999</td>
<td>pre-college</td>
<td>algebra</td>
</tr>
<tr>
<td>Power et al. (2005)</td>
<td>2000-2007</td>
<td>postsecondary</td>
<td>other</td>
</tr>
<tr>
<td>Tiwari (1999)</td>
<td>1990-1999</td>
<td>postsecondary</td>
<td>calculus</td>
</tr>
</tbody>
</table>
Table 7

*Studies coded by Design and Evaluation Method*

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Design</th>
<th>Evaluation Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batchelder (2000)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Brown (2000)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Campbell (1996)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Chen &amp; Lui (2007)</td>
<td>different instructors</td>
<td>CAS used</td>
</tr>
<tr>
<td>Cooley (1997)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Cooley (1995)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Fletcher et al. (1990)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Ford &amp; Klicka (1994)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Gesshel-green (1986)</td>
<td>same instructor</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Graham &amp; Thomas (2000)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Hagerty &amp; Smith (2005)</td>
<td>same/different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Hamide (2001)</td>
<td>different instructors</td>
<td>CAS used</td>
</tr>
<tr>
<td>Hawker (1986)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Hollard &amp; Norwood (1999)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Hopkins &amp; Kinard (1998)</td>
<td>different instructors</td>
<td>CAS used</td>
</tr>
<tr>
<td>Keepers (1995)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Kramarski et al. (2006)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Leng et al. (2005)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
</tbody>
</table>
Table 7 continued

<table>
<thead>
<tr>
<th>Study</th>
<th>Study Design</th>
<th>Evaluation Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li &amp; Edmonds (2005)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Melin-Conejeros (1992)</td>
<td>same instructor</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Palmiter (1986)</td>
<td>different instructors</td>
<td>CAS used/not used</td>
</tr>
<tr>
<td>Palmiter (1991)</td>
<td>different instructors</td>
<td>CAS used</td>
</tr>
<tr>
<td>Portis (1991)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Power et al. (2005)</td>
<td>different instructors</td>
<td>CAS used</td>
</tr>
<tr>
<td>Runde (1998)</td>
<td>different instructors</td>
<td>CAS used/not used</td>
</tr>
<tr>
<td>Runde (1997)</td>
<td>same instructor</td>
<td>CAS used</td>
</tr>
<tr>
<td>Shaw et al (1997)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Stephens &amp; Kovalina (1999)</td>
<td>same instructor</td>
<td>CAS not used</td>
</tr>
<tr>
<td>Tab (2001)</td>
<td>same instructor</td>
<td>CAS used</td>
</tr>
<tr>
<td>Tiwari (1999)</td>
<td>same instructor</td>
<td>CAS used/not used</td>
</tr>
<tr>
<td>Trout (1993)</td>
<td>different instructors</td>
<td>CAS not used</td>
</tr>
</tbody>
</table>


Khalili, A., & Shashaani, L. (1994). The effectiveness of computer applications: A meta-

research. *Communications of the Association for Information Systems, 16*, 665 – 
686.


www.edexcellence.net

century*. Retrieved August 9, 2007 from www.csun.edu/~vcmth00m/

Ahistory.html


effects of instructional methods on mathematical literacy. *The Journal of 

Reproduction Service No. EJ286565).


