Effects of Technology on Students' Achievement: A Second-Order Meta-Analysis

Rana M. Tamim

A Thesis

In the Department

of

Education

Presented in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy (Educational Technology)

at

Concordia University

Montreal, Quebec, Canada

February, 2009

Your file Votre référence
ISBN: 978-0-494-63392-2
Our file Notre référence
ISBN: 978-0-494-63392-2

ABSTRACT

Effects of Technology on Students' Achievement: A Second-Order Meta-Analysis

Rana M. Tamim, Ph.D.
Concordia University, 2009

Numerous meta-analyses addressing the effect of technology on student achievement differ by focus, scope, content, sample, and methodological quality, making the interpretation of the overall effect challenging. To overcome this problem, this dissertation implemented a systematic quantitative synthesis procedure (second-order meta-analysis) to answer the question: does technology use enhance student achievement in formal face-to-face classroom settings as compared to traditional (no/low technology) settings, while taking methodological quality into consideration.

Literature searches and review processes resulted in 37 relevant meta-analyses involving 1253 different primary-studies (approximately 130,300 participants). After examining the lists of primary studies, 25 meta-analyses incorporating 1055 primary studies (approximately 109,700 participants) were found to have greatest coverage of the overall set of primary-studies while minimizing the problem of overlap in primary literature.

Analyses revealed a variety of weaknesses in the implementation of the meta-analytic procedures. To synthesize the 25 effect-sizes from the unique meta-analyses, two standard error approaches were used, one based on sample sizes in the primary studies, and one based on number of studies included in individual meta-analyses. The weighted mean effect-sizes from the two approaches, 0.315 and 0.333 respectively, were significantly different from zero. Results from the first approach revealed a high level of heterogeneity while those from the second one were homogeneous. Moderator analysis for results from the first approach revealed that higher methodological quality meta-analyses and higher inclusivity regarding the covered literature and incorporated research designs in a meta-analysis were associated with lower average effect-sizes.

To validate these findings, 574 individual effect-sizes (60,853 participants) were extracted from 13 meta-analyses that provided sufficient information. The weighted mean effect-size of 0.304 was significantly different from zero and highly heterogeneous thus supporting the findings of the second-order meta-analysis with both approaches. The results consistently represent a medium strength effect-size, favouring the utilization of technology.

Guidelines for conducting a second-order meta-analysis with advantages and disadvantages of the used approaches are presented and discussed with suggestions for applicability in different settings. Implications for technology use are offered and recommendations for future meta-analyses are suggested, including the need for greater systematicity, rigour and transparency in implementation and reporting.

ACKNOWLEDGEMENTS

Pursuing a post graduate degree has been a personal dream for a long time, and like many things in life its completion could not have been achieved without the help and support of amazing individuals whom I have been blessed with knowing, and to whom I owe a few words of appreciation.

To my supervisor, Dr. Richard Schmid, I would like to express my most sincere gratitude. Your continuous support, guidance, and willingness to help are highly appreciated. Throughout the five years I have known you, my ability to address issues from different perspectives has matured and my appreciation of the term *depends* has developed. To Dr. Robert Bernard, this dissertation was made possible because of your skills, expertise, and willingness to help. Your support and friendship are highly valued and words are not capable of expressing my appreciation for your encouragement. Working with both of you has helped me see how a professor may be a colleague first and foremost, and allowed me to appreciate the fact that one gains most through giving, particularly in academics and research.

Great thanks are in order to Dr. Phil Abrami who in a moment of reflection on a Tuesday morning suggested the notion of conducting a review of meta-analyses that morphed into this second-order meta-analysis. Your compelling and challenging questions have raised the bar for me, and helped me refine my work and improve the quality of my dissertation.

To Dr. Evgueni Borokhovski, I would like to express my gratitude for your assistance, and your continuous readiness to offer your feedback and opinion throughout the various stages of this dissertation, and on the final draft. Your enthusiasm for my project helped keep my interest alive at moments when I started questioning it myself. Thanks for the help and for being a friend.

To Ms. Anne Wade, this dissertation would not have been as comprehensive without your help and expertise. Working with you made me see information retrieval with an all new light that goes beyond ERIC and Google.

To my colleague Mike Surkes, I would like to extend my deepest gratitude for participating in the coding process. Your commitment and efficiency were a continuous inspiration for me to work harder on more than one occasion. I would also like to thank Ms. Katehrine Hanz for her help with *Endnote* and the reference list.

As for my family, I owe my existence and all my achievements, to my wonderful parents, Ilham and Mohamad, to whom I am grateful for their unconditional love and for instilling in me the commitment for learning and the appreciation for education. I also want to thank my sisters and brothers for their continuous love and support, and their setting a great standard for me to live by, at personal, academic, and professional levels. Particularly, I will forever be thankful for my sister Nada who witnessed my ups and downs in the last five years, and even the subterranean moments and helped me through it all. Thanks for being there for me. I guess I also need to thank Vonage and Fido for their plans which made it possible.

For the greatest kids in the world, Mohamad, Jad, and Rand, I cannot claim that this dissertation is as much yours as it is mine, because in reality it is yours. You made it happen with your endurance and understanding. You are the ones who felt on a first hand bases the real challenges and frustrations of the past few years. Thank you for your patience and love, for being there for me and giving me a reason to live, and for being

more than anything a parent may dream of or ask for. I hope you know how much I love you.

Most of all I would like to thank my greatest friend, my partner, my lover, and my husband Abdul Rahman Habbal for a love beyond the limitations of words. I know that I am truly blessed to have found a person like you, and I thank you for accepting my weaknesses and reinforcing my strengths. No words can express how grateful I am for the happiness and meaning you bring to my life, so I will go with a simple *I love you.*

I dedicate this with all my love to

Mohamad & Jad, *my best friends*

Rand, *my beautiful princess*

&

Abdul Rahman, *my everything*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

General Introduction

The level to which computer technology has permeated our lives is undeniable. Whether a believer in its advantages or not, one has to admit that it is a central part of the daily life in the 21st century. Currently, the impact of computer technology on all aspects of our lives is being sensed more than ever before, due to a variety of factors including affordable prices of desktop computers (Winn, 2002). At the most superficial level, the affordability of computers is leading to their use in almost every area in our societies. Whether around the household, in public service operations, within the corporate sector, or in the academic field, computers are becoming an integral part of day-to-day lives. The pervasiveness of computer technologies, including information and communication technologies, has reached a level where it is almost impossible to find an institution which is computer and internet free.

This profusion of computer technology has not always been the case nor has it been a predictable progression of events during the earlier years of development in the computer technology arena. Although not fully supported by documented evidence, it is alleged that in 1943 Thomas Watson, the chairman of IBM said: "I think there is a world market for maybe five computers." ("Thomas J. Watson", 2008). While it is not fully clear whether Watson is the real author of this quotation, the statement reflects the overall perception about the future of computers at that time. Nevertheless, the advent of technology and the forward march has been so steady and quick it led Bill Gates to claim that: "If GM had kept up with technology like the computer industry has, we would all be driving $25 cars that got 1000 MPG" (*Bill Gates quotes*, 2008).

The high level of pervasiveness of computer technology in the different aspects of society has influenced many facets of our lives, including language to the point that the noun "mouse potato" and the verb "Google" have found their way to the Merriam-Webster's Collegiate Dictionary. It is not surprising that the education sector experienced the ripples of the computer technology wave since its early days. Ever since the late 1970's when microcomputers became available and Apple II microcomputers succeeded in accessing schools (Alessi & Trollip, 2000) the computer technology march into the classroom has grown stronger by the day.

The list of technologies that are thought by many to enhance learning and offer the solution that will help change the role of the teacher from the *sage on the stage* to the *guide on the side*, allowing for more active and meaningful learning (Jacobson, 1998) is quite long. Some of these learning technologies include computer-assisted instruction, computer-based instruction, intelligent tutoring systems, videoconferencing, interactive multi-media, web-based instruction, and e-learning. However, the impact of such technologies on the learning process and students' achievement is still elusive and debatable. Research addressing the relationship between technology use and students' cognitive outcomes has been increasing exponentially with the hope of offering conclusive results and the intention of giving guiding principles for adequate technology integration procedures for maximal student advantages. Similar to the situation in other areas of interest, the fact that no single study can provide conclusive evidence has caused attention to turn to literature reviews, especially meta-analyses, in order to make sense of what the overall body of research has to say.

Meta-analysis is a systematic review technique that was developed by Glass in the 1970's. Olkin described the relationship between primary research and meta-analysis by comparing it to being in a helicopter and moving further from the ground where the focus and visibility of the trees diminishes allowing patterns that are not detectable from the ground to emerge (Hunt, 1997). Glass, McGaw, and Smith (1981) stress that it is a technique for integrating empirical research which was initiated because of necessity. The overwhelming exponential growth in empirical research in the 20th century made it very hard to depend on regular narrative literature reviews to capture the essence of what the body of literature has to say. Similarly, vote counts were not adequate enough to extract information from the vast body of literature, organize it, analyze it, and present it (Glass, 1977). The popularity of the meta-analytic approach is reflected by a quick *Google* search for the term which returns 3,160,000 hits. The increased interest and attention given to meta-analysis is also highly evident in the educational field. A recent search of the ERIC database revealed more than 1726 documents that implement or discuss meta-analytic procedures. Particularly concerning computer technology and its impact on students' achievement and attitudes, a preliminary search of the ERIC data-base at the onset of this study revealed 62 meta-analyses, published between 1980 and 2006.

Statement of the Problem

Upon checking the different meta-analyses in a particular area, including technology integration and student learning, we find that they differ in the adequacy of the implemented procedures and thus their methodological quality. Such an issue makes it hard to decide on which meta-analysis to trust especially in the absence of approaches to assess the methodological quality of a given meta-analysis. Other aspects that the meta-analyses differ on include the scope of the questions answered, the time frames covered, the grade levels, and subject matter targeted. For example, the meta-analysis conducted by Timmerman and Kruepke (2006) addressed CAI and its influence on students' achievement at the college level, Christmann and Badgett (2000a) investigated the impact of CAI on high school students' achievement, and Bayraktar (2000) focused on the impact of CAI on K-12 students' achievement in science. Alternately, Bangert-Drowns (1993) studied the influence of word processors on student achievement at various grade levels while Cohen and Dacanay (1992) focused on the impact of CBI on students' achievement at the post-secondary levels. Although there might be some redundancy in the issues addressed by the different meta-analyses or some overlap in the empirical research included in some of them, they offer a rich and invaluable source of information that might prove to be complementary if synthesized and analyzed appropriately.

As producers or consumers of meta-analyses, we find ourselves in a situation similar to that with primary research where the need for judging methodological quality is essential, along with the need to synthesize the growing body of literature to answer big and broad questions in a given area, including that of computer technology use in education. Capturing the essence of what a collection of meta-analyses in a given area has to offer may be done through regular narrative reviews or through a systematic quantitative approach that emulates the meta-analytic process at a secondary level. The former approach suffers from the various flaws pertinent to narrative reviews the most important of which is biasness and inability to account for sample sizes or strength of the effects of a given treatment. As for the latter approach, it has been experimented with by several researchers such as Lipsey and Wilson (1993), Wilson and Lipsey (2001), Sipe and Curlette (1997), Møller and Jennions (2002), Barrick, Mount and Judge (2001), Peterson (2001), Sheeran (2002), and Luborsky et al. (2002). These syntheses did not follow a common or standard set of procedures, nor did they specifically address the methodological quality of the included meta-analyses, but the approach is thought to offer potential advantages in making sense of the growing body of literature and reaching more reliable and generalizable inferences than individual studies (Peterson, 2001). Moreover, a systematic quantitative synthesis of meta-analyses may prove to be an easier task to complete than conducting a full scale comprehensive meta-analysis to answer one big question, which in certain situations may include reviewing thousands of primary studies. Particularly in the case of technology integration and its impact on student achievement, a search in 2006 of the ERIC database for primary research using a combination of different terms related to computer technology use in post secondary educational settings only, with no restriction on the publication date, yielded 9372 records. The number would be quite more substantive if a search is conducted at the present time while including the variety of grade levels.

Overall Objectives

With the ongoing interest in the impact of technology on learning, and the growing attention to and increasing number of published meta-analyses this dissertation had two main components. The first component is methodological aiming at: a) designing an approach to assess the methodological quality of meta-analyses in the social science field; b) piloting a second-order meta-analysis procedure that takes methodological quality into consideration; and c) validating the results of the second order-meta-analysis. The second component aims at answering substantive questions related to meta-analyses addressing the impact of computer technology on student achievement in formal educational contexts through implementing the second-order meta-analysis methodology. The objectives of the second component are to: a) critically examine the meta-analyses addressing the impact of computer technology on learning; b) synthesize the findings of meta-analyses addressing technology integration and student achievement through a second-order-meta-analysis; and c) explain the variance in the effect sizes if possible

CHAPTER 2: REVIEW OF RELATED LITERATURE

"All research begins and ends in the library"

P.C. Abrami (personal communication, November 27, 2008)

This section reviews the literature addressing issues pertinent to technology integration within educational contexts as well as theoretical and procedural aspects related to meta-analysis. First, the literature on computer technology and learning is discussed. Next, the theoretical framework for meta-analysis and procedural aspects are presented. This is followed by an overview of meta-analyses addressing learning and computer technology. Finally, current examples of reviews of meta-analyses are discussed. The section concludes with the rationale and objectives for the current study.

Learning and Computer Technology

"They say one of a baby's first non-verbal forms of communication is pointing. Clicking must be somewhere just after that."

(Anonymous, *Computer Quotes*, 2008)

Pervasiveness of Computer Technology

No one can deny the current importance of computer technology, and the level to which it has pervaded our daily lives. Its impact on different aspects of our communities is escalating on a daily bases and is being sensed more than ever before. The areas in which technology is getting to be significant and fundamental is highly varied and includes entertainment, knowledge retrieval, business transactions, health services, formal governmental correspondence with citizens, communication across various areas around the globe, and transmission of information between individuals on earth and those orbiting around it.

Numbers and statistics pertaining to the ownership and use of computer and computer related communication tools reveals the level of dependency, and the amount to which computers are becoming a central part of our lives and environments. According to a report published by the *Pew Research Center*, 82% of Swedes, 81% of South Koreans, 80% of Americans, and 76% of Canadians were computer users in the year 2007 (Kohut, Wike, & Horowitz, 2007). According to the same report, although there still is a digital divide between developed and developing countries, the overall use of computers in many poor or middle income countries has witnessed an increase over the five year period from 2002 to 2007. For example, computer usage in India has increased from 22% to 28% while in Peru it increased from 26% to 39%.

According to the *Internet World Stats*, the world total of internet users has increased from *360,985,492* individuals in the year 2000 to *1,463,632,361* in the year 2008 indicating a 305.5% growth over a six year period (Internet World Stats, 2008b). In 2008, 84.3% of the Canadian population and 72.5% of the United States of America's population are internet users (Internet World Stats, 2008a).

Computer Technology in the Educational Context

Particularly within the educational contexts the introduction of computer technology into the classroom dates to the 1978 when Apple II microcomputers were introduced to the school setting (Alessi & Trollip, 2000). The pace at which the integration of computer technology into different classrooms may not have been very fast, nevertheless, at the present time, the pervasiveness of computer technologies has reached a degree where it is almost hard to find an educational institution in the developed countries which is computer and internet free. In the United States of America, more than 91% of students in formal education (preschool to grade 12) were computer users in 2003 with 59% being internet users (DeBell & Chapman, 2006). As for Canada, it was reported that in 2003-2004, over 90% of elementary and secondary schools in Canada were connected to the internet, while 99% of the schools had computers, with a total of more than a million computers being accessible to students and teachers (Plante & Beattie, 2004).

Computer technology has been used to enhance instruction through a variety of approaches or strategies including computer-assisted instruction, computer-based instruction, drill and practice, simulations, tutorials, computer gaming, online learning, and computer-mediated communication. Some of the technological approaches are clearly understood and defined such as drill and practice which refers to software programs that offer the students the chance to work on structured problems or exercises while providing immediate feedback. Another example is computer-mediated communication which refers to "communication between two or more individuals with text-based tools such as e-mail, instant messaging, or computer-based conferencing systems" (Spector, Merrill, Van Merrienboer, & Driscoll, 2008, p. 819).

Other technological applications have more than one definition such as a simulation which is: "A working representation of reality; used in training to represent devices and process and may be low or high in terms of physical or functional fidelity. Also, an executable (runnable) model; computer software that allows a learner to manipulate variables and processes and observe results. Also, a computer-based model of a natural process or phenomenon that reacts to changes in the values of input variables by displaying the resulting values of output variables." (Spector, Merrill, Van Merrienboer, & Driscoll, 2008, p. 826).

Moreover, some educational technology terms are not clearly defined in the literature such as computer-assisted instruction. It may be used as a general term to represent a variety of technology uses for the enhancement of instruction such as drill and practice and tutorials, or as a specific approach to technology such as computer-based programmed instruction (Schenker, 2007). Finally, some terms are used flexibly and interchangeably such as computer-based instruction which is considered to be the newer version of computer-assisted instruction ("Computer-assisted instruction", 2008). Despite the multiplicity of situations regarding the clarity of terms in the field, one thing is absolutely clear and highly straightforward; computer technology is unquestionably a central element in the 21st century classroom.

Computer Technology and Learning

With the current wide spread of computers, and the availability of information communication technology, computer skills are becoming a central and important goal for all school systems and at all different levels (Plante & Beattie, 2004). After all, one might consider it as important a tool for today's student as a paintbrush to a painter (Jonassen & Reeves, 1996). However, the debate around the influence of technology on learning has been going on for a long time. Clark (1983) started the argument with the stand that computer technology has no impact on learning, and that media is a mere vehicle that delivers goods (knowledge) to the learner. This led Kozma (1991) to retaliate by arguing that computer technology is much more than a mere truck, and that it has an actual impact on the learning process. No main findings were reported by researchers to back up Kozma's argument (Clark, 1994).

A more recent call by Kozma was to restructure the debate to *will media influence learning* (Kozma, 1994). Clark (1994) responded by noting that media will never influence learning, and the active ingredient in the learning process is the learning strategy confounded with the use of a certain medium. This brought a third party to the debate, where Jonassen, Campbell, and Davidson (1994) argued that there is no use in going on with an instruction/media centered debate. In their opinion, the focus should be on a learner-centered debate where the main attention should be on how to use computer technology most effectively to support a learner-centered environment. Almost ten years later, and the debate is still ongoing, with researchers still trying to find support for either one of the two standpoints (Akyol & Cagiltay, 2007; Mayer, 2003).

Beyond the Clark/Kozma debate, and within the educational field, computer technology has been advocated by many to be the "magic bullet" that will make education more accessible, affordable, and effective (Van Dusen, 1998). Van Dusen stressed that technology has proved to be cost effective from an administrative point of view, and has instigated new ways of looking at teaching and learning. Major uses of computer technology throughout the world include: a) gathering information; b) keeping records; c) creating proposals; d) constructing knowledge; e) performing simulations to develop skills; f) distance learning; and g) global collaboration for lifelong learning and work (Jacobson, 1998; Kimble, 1999).

Higher achievement, increased motivation, enhanced self confidence, greater student satisfaction, and more effective support for special needs students are only some of the desired and promised benefits. Many researchers believe that computer technology has changed teaching and learning in post secondary classess (Lowerison, Sclater, Schmid, & Abrami, 2005) and has improved learning outcomes (Bransford, Vye, & Bateman, 2002; Kuh & Vesper, 2001; McCombs, 2000). Others believe that it has the potential for enhancing students' problem solving skills (Jonassen, 2003), helping the students by increasing access to information (Bransford, Vye, & Bateman, 2002; Hill & Hannafin, 2001), offering more convenient access to the instructor, easier presentation of course content, and more effective studying strategies (Grabe & Sigler, 2002), and furnishing richer learning environments (Bransford, Vye, & Bateman, 2002). Still, others believe that computer technology presents learners with the chance to develop critical thinking skills as authors, designers, and constructors of knowledge (Jonassen & Reeves, 1996).

Nevertheless, research findings offer a variety of contradictory results regarding the impact of computer technology on student achievement. This only adds to the controversial issue and debate of the impact of computer technology on the learning process and its outcomes. Different research studies have reported positive results regarding the impact of technology on student learning and achievement. Kulik (1994) reported that students tend to learn more in less time in classess that apply computer-based instruction. Furthermore, he noted that students reported enjoying classess when computer help is provided, and learned the same or more than from peers or cross-age tutoring. Later, Kimble (1999) noted that research has demonstrated that student learning and self confidence tend to increase when computer software are used to solve authentic real world problems.

In 1999, Pisapia, Knutson, and Coukos, found that student achievement can be influenced by appropriate integration of computer technologies into instruction (Pisapia, Knutson, & Coukos, 1999). Moreover, Mitra conducted a study that checked student attitudes and use of computers in a "computer-enriched" environment (Mitra & Steffensmeier, 2000). Findings reflected that the computer enriched environment is positively correlated with students' attitudes towards computers in general, their role in teaching and learning, and their ability to facilitate communication. From another perspective, Laurillard emphasized that computer technology can offer benefits to learning, but the effectiveness is influenced by many factors such as instructional design, learner characteristics, and nature of the learning task (Laurillard, 2002).

With all these positive attitudes and findings, many researchers still address computer technology with a critical outlook (Van Dusen, 2000). The majority of critics do not refute the positive research findings, but mainly criticize the way it is being used in classrooms, the teacher preparedness, and the relative cost to acquiring technology in the academic context (Kimble, 1999). For example, Salomon (2000) argues that no major changes have occurred in education as a result of computer technology integration. He believes that computer technology has impacted the medical field, advertising, and travel, way more than education. Similarly, Cuban, Kirkpatrick, and Peck (2001) conducted an investigation with two highly technological schools and concluded that access to computer technology, including equipment and software, rarely leads to extensive teacher or student use. Moreover, their investigation led them to the conclusion that computer technology was used in ways to support teachers existing teaching practices and strategies rather than alter them.

On a different note, Becker (2000) says that not all computer activities attract the same degree of student interest and effort, neither is it that all computer activities are equally valuable in improving learning outcomes. A summary of the debate and a critical analysis of a sample of studies conducted by Joy & Garcia (2000) led to the conclusion that "learning effectiveness is a function of effective pedagogical practices" (p.33). On a much stronger note, it is considered by some to pause a threat to the new generation's intellectual skills to the point where they wonder if it has the power to make "kids stupid" (Ferguson, 2005).

With the ongoing debate, the contradictory research findings, and opposing attitudes towards computer technology, many researchers and practitioners are trying to find best methods, practices, and approaches, to make the most out of what computer technologies have to offer. Adding the fact that no single study can provide conclusive evidence, attention has turned to meta-analysis as a technique, considered by many, to help in capturing the essence of the expanding body of literature (Bernard & Naidu, 1990).

Meta-Analysis

"Scientists have known for centuries that a single study will not resolve a major issue. Indeed, a small sample study will not even resolve a minor issue. Thus, the foundation of a science is the culmination of knowledge from the results of many studies."

(Hunter, Schmidt, & Jackson, 1982, p. 10)

What is a Meta-Analysis?

Meta-analysis is a systematic review technique that was developed by Gene Glass in the 1970's. He first defined it as "analysis of analyses" referring to the examination of a large collection of analyses presented in different studies with the goal of integrating them to help in higher generalizability of the findings (Glass, 1976). It is a form of survey in which research reports are investigated through a statistical standardization procedure of the study findings so that the resulting numerical values can be interpreted in a consistent way across all measures and variables involved (Lipsey & Wilson, 2000). It depends on the use of effect size as a metric to measure the difference between the control and treatment conditions (Bernard et al., 2004).

Meta-analysis helps in resolving contradiction in research findings (Bangert Drowns & Rudner, 1991) while addressing the need to: a) capture the essence of the expanding body of literature; and b) overcome biasness in other forms of reviews (Bernard & Naidu, 1990; Glass, McGaw, & Smith, 1981). After all, the overwhelming exponential growth in empirical research in the 20^{th} century rendered it very hard to depend on regular narrative literature reviews or vote counts to capture the essence of what the body of literature has to say.

Advantages of Meta-analysis

Although narrative reviews and vote counts have been used for a substantial amount of time, they are neither scientifically sound (Kline, 2004) nor adequate in extracting information from the vast body of literature, organizing it, analyzing it, and presenting it (Glass, 1977). They do not account for different sample sizes and the varying strength of results in different studies, they are not statistically powerful, and they do not address the size of the effect in a given study (Abrami, Cohen, & D'Appollonia, 1988; Hunt, 1997). They rely heavily on statistical test outcomes, namely the *p* value, which is subject to all the null hypothesis testing limitations emphasized by many researchers including Meehl (1967), Cohen (1990), and Glass (1976). In addition, they are highly subjective (Bernard & Naidu, 1990; Slavin, 1984) and most of them are restricted to published research studies which includes the overestimation bias entailed with that (Kline, 2004).

Since no single study can ever give conclusive evidence the need for systematic quantitative research syntheses gets to be stronger since they offer greater coverage of the population to be studied and help in overcoming chance fluctuations within samples thus allowing for more generalizable findings (Bernard & Naidu, 1990). Wolf (1986) asserts that meta-analysis is a process which addresses major problems in traditional literature reviews, including: a) selective inclusion and exclusion of studies; b) subjective differential weighting of studies; c) misleading interpretations; d) overlooking study characteristics and features that explain different findings; and e) failing to address moderating variables.

According to Lipsey and Wilson (2000) advantages of the meta-analytic approach include it being:

a) Organized and systematic in handling information from a variety of sources.
b) Sophisticated and takes into consideration the magnitude and direction of each relevant statistical relationship.
c) Capable of addressing different study features that are not dealt with in other forms of reviews.
d) Able to offer more statistical power due to the pooling of the effect sizes from different research studies.
e) Systematic and explicit allowing the reader to judge the value of the findings.

Main Meta-analytic Approaches

There are four main approaches to meta-analysis in the literature: a) the Glassian; b) the study effect; c) the homogeneity; and d) the psychometric (Bangert Drowns & Rudner, 1991).

The classic or Glassian meta-analysis (Glass, McGaw, & Smith, 1981) is an approach that has proved to be robust upon re-analysis. It applies liberal inclusion criteria, with the unit of analysis being the "study finding" where more than one comparison per study may be calculated leading to dependency issues. Moreover, it allows for aggregating effect sizes from different dependent variables even if they are measuring different constructs jeopardizing the reliability of the findings (Bangert Drowns & Rudner, 1991).

The study effect meta-analysis approach mainly includes studies with specific methodological quality resulting in higher selectivity. It considers the study to be the unit of analysis with one effect size per study and thus overcomes the dependency problem. Although it gives same weighting to each of the included studies, it reduces the number of effect sizes and may be influenced by researcher's bias in the inclusion/exclusion criteria (Bangert Drowns & Rudner, 1991).

As for the tests of homogeneity approach (Hedges & Olkin, 1985) it is used to determine if the observed variance is due to sampling error. If the homogeneity test is significant, studies are repeatedly subdivided to find groups that offer non-significant within group variation. With the implementation of this approach heterogeneity is often found due to a variety of factors that may influence the variance. Having multiple divisions of the studies may lead to chance findings that may lead to the identification of incorrect moderators (Bangert Drowns & Rudner, 1991).

Finally, the psychometric meta-analysis approach (Hunter, Schmidt, & Jackson, 1982) combines the advantages of the other approaches where studies are included regardless of quality, the effect size distribution is corrected for a variety of errors including sampling and measurement errors, and the identification of subgroups based on study features. The subgroups are further meta-analyzed separately if the variance remains large. The major issue with this approach is the need for substantial information from the included primary studies for accurate effect size corrections, and which is unfortunately not always the case (Bangert Drowns & Rudner, 1991).

Criticisms and Defence of Meta-analysis

Similar to all other techniques, meta-analysis has had its share of criticism, the most important of which is its invalidity because of the mixing "apples and oranges" argument (Eysenck, 1978). Glass and his colleagues argue against this by noting that both apples and oranges are included in the overall category of "fruits" (Smith, Glass, & Miller, 1980). One may go further by counter arguing that the analogy needs to be modified. The apples and oranges analogy is adequate if a meta-analysis attempts to synthesize research answering unrelated questions; however, a meta-analysis usually targets the synthesis of research addressing an explicit topic within a specific field (Light & Pillemer, 1984). This is usually addressed by the inclusion and exclusion criteria for each meta-analysis (Sharpe, 1997; Sipe & Curlette, 1997). Some examples of questions addressed by meta-analyses include: what is the relationship between class-size and achievement (Glass & Smith, 1979) and how does distance education compare to face-to-face instruction (Bernard et al., 2004). Looking at it more closely, an adequate meta-analysis seems to be a mix of different varieties of apples rather than a mix of apples and oranges. With this analogy, one would argue that a meta-analysis is calculating the average nutritional value of all varieties of apples. Moreover, it is comparing the nutritional value, the quality, the color, the taste, the lustre, and the best usage of different varieties of apples.

Advocates of meta-analyses argue that comparison of different studies is the only comparison that makes sense, since they are the only studies that need to be compared (Glass, McGaw, & Smith, 1981; Lipsey & Wilson, 2000). To them, synthesizing studies that are fully similar is self-contradictory since they should all lead to the same finding, which does not need any synthesis. Kline (2004) stresses that if a situation is present where the synthesized studies are exact replications, all that a quantitative synthesis will have to offer would be: a) an estimation of the central tendency which is a better estimation of the population parameter than the results of any single study; and b) an estimate of the variability of the results which could be used to better identify individual outliers. On the other hand, he argues that a meta-analysis in behavioral sciences depends on synthesizing studies that tend to be generally construct replications, where the meta-analyst attempts to identify and measure the characteristics that give rise to variability in the results of those construct replications. This issue is addressed by coding study features and study characteristics, and empirically testing the impact of each on the results of the meta-analysis (1986). Sharpe (1997) also stresses that researchers usually assess if a set of effect sizes is to be pooled by statistically testing for homogeneity, with this being reported in the majority of published meta-analyses (Matt & Cook, 1994).

Another criticism is the "garbage in, garbage out" (Eysenck, 1978) argument which questions the quality of the findings from a meta-analysis when studies are included regardless of methodological quality. This is particularly true of all types of endeavours including different forms of literature reviews. To resolve this problem, two approaches have been followed. The stricter approach would be the best evidence synthesis method that limits inclusion of studies in a meta-analysis to randomized control studies (Lipsey & Wilson, 2000) such as the Cochrane Collaboration. Other proponents argue that studies with different methodological qualities should be included in a meta-analysis as long as the quality is coded and accounted for, and interaction between effect size and methodological quality is investigated (Cooper and Arkin, 1981; Glass et al., 1981).

Another criticism is the file drawer weakness (Rosenthal, 1979) which highlights the issue of publication bias resulting from including published studies which usually report significant findings. Consequently, meta-analyses that include only published research are liable to reach conclusions that are biased against a given null hypothesis. To overcome this problem, meta-analysts are urged to include both published and unpublished primary research which should be targeted at the inclusion/exclusion criteria level as well as the literature search strategies utilized.

With all this being said, and regardless of the arguments for or against meta-analysis, it seems that it is a technique that is here to stay. Meta-analyses are widely spread and used in different scholarly fields. A number of associations has been formally established for the purpose of supporting systematic reviews in general and meta-analyses in particular (Bernard, Abrami, Lou, & Borokhovski, 2004). Examples include the What Works Clearing House, the Cochrane Collaboration, and the Campbell Collaboration. What is more interesting is the fact that in many graduate and post-graduate programs, many research design courses are including a section on meta-analyses, with some specific courses being designed to acquaint future researchers with skills for understanding or conducting future meta-analyses.

A recent search of the ERIC database revealed more than 1100 documents that implement or discuss meta-analytic procedures. Major journals such as the American Educational Research Association's journal, namely the Review of Educational Research, consider meta-analyses to be of "particular interest when they are accompanied by an interpretive framework" (2008). Overall, the number of published meta-analyses in respectable journals has been growing rapidly in the fields of education, psychology, medicine, and business; in addition to un-published dissertations (Sharpe, 1997). Technologies' impact on or relationship with learning is not any different as revealed by the growing number of meta-analyses addressing this topic.

Meta-Analyses Addressing Computer Technology and Learning

The increased interest and attention given to meta-analyses is generally evident in the educational field, and particularly in addressing the impact of computer technology on learning. As noted earlier, and since the debate is still ongoing, researchers turned to meta-analyses to try and make sense of what the body of literature has to say. This is evident in the number of published meta-analyses addressing the issue.

At the onset of this project, a preliminary search of the ERIC data-base resulted in the location of 62 meta-analyses, published between 1980 and 2006, that addressed the impact of computer technology on students' learning and motivation. A sample is presented in Table 1. It goes without saying that if other resources, such as Psychinfo, ProQuest, Digital Dissertations and Theses Full-text, and EBSCO Academic Search Premier, are to be searched; this number is bound to increase.

With such an expansion in the number of meta-analyses, two challenges arise: a) how to assess the methodological quality of a given meta-analysis and subsequently which meta-analyses are to be trusted more; and b) how to capture the essence of what all this body of literature is offering. So, similar to the need for approaches for the appraisal of the methodological quality of primary research and its synthesis, there seems to be a need for a systematic review procedure to synthesize findings of different meta-analyses in different fields, including that of computer technology use in education.

Table 1. Partial list of meta-analyses conducted since 1985

Author	Publication Year	Technology addressed	Level	Number of *ES*s	Overall Mean *ES*
Timmerman & Kruepke	(2006)	Computer-Assisted Instruction (CAI)	College	118	r = 0.12
Pearson et al.	(2005)	Digital Tools	Middle School	20	0.49
Ungerleider & Burns	(2003)	Networked information communication technology	NS	12	0.00
Bayraktar	(2001)	CAI	K-12	108	0.27
Lou et al.	(2001)	Technology	NS	486	NS
Christmann & Badgett	(2000a)	CAI	High school	24	0.27
Liao	(1999c)	Hypermedia	NS	46	0.41
Whitley	(1997)	Attitudes toward computers	NS	104	0.23
Azevedo & Bernard	(1995)	Feedback in CBT	Higher Education	34	0.73
Walther, Anderson & Park	(1994)	Computer-mediated interaction	NS	35	NS
Bangert-Drowns	(1993)	Word processors	NS	33	0.21
Cohen & Dacanay	(1992)	Computer-based instruction	Adult and university	37	0.41
Kulik & Kulik	(1991)	Computer-based instruction	NS	248	0.30
Fletcher	(1989)	Interactive videodisc instruction	Adults	47	0.50
Roblyer	(1988)	Computer-based instruction	NS	NS	NS
Kulik & Kulik	(1987)	Computer-based education	All	NA	NA
Kulik & Kulik	(1986)	Computer-based education	College	99	0.26

Note: NS refers to "not specified" Unless otherwise noted, *ESs* are standardized mean differences.

Methodological Quality of Meta-Analyses

Considering that meta-analysis has been established as a useful methodology for research synthesis over the past two decades, questions about methodological quality become more pressing. For example, some researchers have long argued that the effect size offered by meta-analysis may have "mischievous" outcomes especially with naïve readers who may take the apparent "objectivity", "precision" and "scientism" as a seal of credibility (Cook & Leviton, 1980). In their counter argument Cooper and Arkin (1981) stress that there is nothing "mischievous" in the method itself, rather it may become so in the hands of specific researchers due to "intention" or "ignorance", which is applicable with any innovative methodology. This is heightened by the fact that it has been stressed by many researchers that the main concern about the quality of meta-analytic findings is not related to its theoretical construct or procedural aspects rather in the quality of implementation by different researchers (Abrami, Cohen, & D'Appollonia, 1988; Slavin, 1984).

To improve the quality of meta-analyses, different associations have worked on designing and implementing a set of specific standards to be followed by researchers interested in registering systematic reviews with them. Examples include the *What Works Clearing House*, the *Cochrane Collaboration*, and the *Campbell Collaboration.*

A few assessment tools have been designed to assess methodological quality of meta-analyses. One prominent example from the medical field is *The Quality of Reporting of Meta-analyses* (QUOROM) which was the outcome of a group conference consisting of 30 clinical epidemiologists, clinicians, researchers, statisticians, and editors. The group's objective was to identify items for a checklist of standards that would be used to assess and evaluate the quality of a meta-analytic report to help in improving the quality of reviews of randomized control trials in the medical field.

Another notable tool in the health science area is the one developed by the health.evidence.ca investigative team. As part of their mandate to help research consumers become more capable decision makers, they offer free access to their evaluation tool, *The Quality Assessment Tool*, and its dictionary.

Nevertheless, currently there is no specific approach that has been designed and utilized for the evaluation of the methodological quality of meta-analyses in the social science area. Although one can make use of some of the described tools, it would be more appropriate if a methodology is particularly designed for implementation within the educational field.

Reviews of Meta-Analyses

Synthesizing findings from different meta-analyses addressing a specific issue may be conducted through a general narrative literature review approach. This would be similar to narrative reviews of empirical research, and are subject to the same flaws and criticisms. An alternative approach would be a quantitative synthesis of findings reported in different meta-analyses in order to offer a conclusion of what they have found based on relevant and related empirical research. The second approach entails addressing the synthesis of meta-analyses quantitatively while considering each meta-analysis as the unit of analysis. Researchers who have experimented with such a quantitative approach to summarizing meta-analytic results include Lipsey and Wilson (1993), Wilson and Lipsey (2001), Sipe and Curlette (1997), Møller and Jennions (2002), Barrick, Mount and Judge

(2001), Peterson (2001), Sheeran (2002), Luborsky et al. (2002), and Butler, Chapman, Forman, and Beck (2006) . Although these syntheses did not follow specific or similar procedures, the number may be considered as an indication that the need for a quantitative means of synthesizing meta-analyses is becoming more pertinent. Such an approach is thought to offer potential advantages in making sense of the growing body of literature, reaching reliable and generalizable inferences than individual studies (Peterson, 2001).

Lipsey and Wilson (1993) conducted their first attempt with a synthesis of meta-analyses of research addressing the impact of treatments based on psychological variables manipulation on psychological change. They included 290 meta-analyses and examined 302 effect sizes in their analysis. This was criticized by Eysenck (1995) who argued that meta-analysis squared does not make sense because it "averages apples, lice, and killer whales" (p. 110). Lipsey and Wilson (1995) answered by saying that if they had combined: "eye blink conditioning with rhesus monkeys, the influence of instructional sets on the Stroop effect, and the impact of deinstitutionalization on the prevalence of homelessness, the results might indeed be a 'gigantic absurdity" (p. 113). They stressed that the meta-analyses they synthesized had a broad but common aspect, namely the implementation of psychologically based treatment with individuals in comparison with less or no treatment conditions. In a more recent attempt, and with a follow up approach to the 1993 analysis, Wilson and Lipsey (2001) conducted another synthesis that included 319 meta-analyses, with 250 of them providing relevant data for final analysis.

In a different field, specifically in biology, Møller and Jenions (2002) conducted a quantitative review of ecological and evolutionary studies to investigate the variance they explain. Their synthesis included 43 published meta-analyses, which yielded 93 estimates of mean effect sizes using Pearson's r and 136 using Cohen's *d* or Hedges'g.

In the area of organizational behaviour and human resource management, Steiner, Lane, Dobbins, Schnur, and McConnell (1991) conducted a review of published meta-analyses. While they did not attempt to quantitatively synthesize the findings of the individual meta-analyses their focus on the methodological aspects of the included meta-analyses was very systematic and comprehensive. Another review was conducted by Torgerson (2007) in which she reviewed and assessed the methodological quality of meta-analyses addressing literacy learning in English, however, the codebook and the coding progress were not as extensive as the one presented in Steiner et al.'s work.

In the area of computer technology and learning, although some qualitative reviews of meta-analyses have been published (Schacter & Fagnano, 1999); no quantitative synthesis of meta-analyses has been reported or published. Moreover, none of the previously conducted quantitative syntheses of meta-analyses addressed the methodological quality of the studies they included in an explicit and clear fashion.

Therefore, the main objective of this dissertation is to conduct a systematic quantitative review of meta-analyses addressing computer technology and its impact on learning while developing an approach to assess the methodological quality of the included meta-analyses. Hunter and Schmidt's term "second-order meta-analysis" will be used to refer to this approach (Hunter & Schmidt, 1990).

Dissertation Objectives

Upon checking the different meta-analyses addressing computer technology and education located by the ERIC search, we find that they differ in one or more aspects including the type of computer technology addressed, the scope of the questions answered, the time frames covered, the grade levels or subject matter targeted, the outcome measures under investigation, and the methodological quality. There might be some redundancy or repetition in the issues addressed by the different meta-analyses or some overlap in the empirical research included in some of them. However, these meta-analyses offer a rich and invaluable source of information that might prove to be complementary if synthesized and analyzed appropriately.

By systematically and critically reviewing these meta-analyses, we can have a more informed idea about what the body of literature has to say, what is known, and what gaps still need to be addressed. Moreover, by quantitatively synthesizing these and other relevant meta-analyses, one has the chance to reach more encompassing conclusions about what the literature has to say without having to re-invent the wheel and conduct a large meta-analysis to encompass the ever so growing body of literature.

Such a synthesis would lead to a more comprehensive understanding of the empirical research addressing the effectiveness of computer technology use in educational contexts. This could be alternatively achieved by conducting a large scale comprehensive meta-analysis that addresses primary research in this area. However, such a task will prove to be a challenging process that is both time consuming and resource depleting simply because of the large number of primary research in the field. A search in 2006 of the ERIC database for primary research using a combination of different terms related to computer technology use in post secondary educational settings, with no restriction on the publication date, yielded 9372 records. The mere review of such a number of records at the abstract level, to decide on full text retrieval for further review, would be extremely time-consuming. This is magnified if one thinks of all the other data bases and resources to be searched in order to ensure that the meta-analysis is adequately inclusive and comprehensive. This challenge, added to the fact that meta-analyses in the field are quite varied and substantial in number, renders it more reasonable and feasible to synthesize their findings. Therefore, by applying the steps and procedures utilized by systematic reviewers to the synthesis of meta-analyses in the field, this dissertation will help in capturing the essence of what the existing body of literature says about computer technology use and learning.

In addition, by designing an approach that helps in assessing the methodological quality of a meta-analysis, the dissertation will also enable meta-analysis users to judge the quality and thus reliability of a given meta-analysis. This may be a very useful when deciding on conducting a systematic review targeting a specific topic to avoid redundancy with previous meta-analyses particularly if they are methodologically strong and trust worthy.

This dissertation has two main components, the first being methodological aiming at: a) designing an approach to assess the methodological quality of meta-analyses in the social science field; b) piloting a second-order meta-analysis procedure that takes methodological quality into consideration; and c) validating the results of the second-order meta-analysis. The second component aims at answering substantive questions related to meta-analyses addressing the impact of computer technology on student achievement through implementing the second-order meta-analysis methodology. The objectives of the second component are to: a) critically examine the meta-analyses addressing the impact of computer technology on learning; b) synthesize the findings of meta-analyses addressing technology integration and student achievement through a second-order-meta-analysis; and c) explain the variance in the effect sizes if possible. Beyond the methodological aspects, particular research questions to be addressed in this dissertation are:

1. Does technology use enhance student achievement in formal face-to-face classroom settings as compared to traditional settings? If so, to what extent?
2. What features, if any, moderate the overall effects of technology use on students' achievement?

CHAPTER 3: METHODOLOGY AND PROCEDURES

"We [meta-analysts] are not advocates, we are reporters. We document and report what is there and what is not there"

P.C. Abrami (personal communication, April 22, 2008)

This second order meta-analysis was designed in order to answer the question of whether technology use enhances student achievement in formal face-to-face classroom settings as compared to traditional settings, and if so to what extent. Furthermore, it was designed to help in investigating different features that moderate the overall effects of technology use on students' achievement. The general systematic approach used in conducting a regular meta-analysis was followed in this second-order meta-analysis with some modifications to meet the specified objectives. After specifying the research question the following steps were followed:

1. Creating inclusion/exclusion criteria.
2. Developing and implementing search strategies.
3. Reviewing and selecting meta-analyses.
4. Extracting effect sizes and standard errors.
5. Developing a codebook.
6. Coding study features.
7. Designing and calculating the methodological quality index.
8. Identifying unique set of meta-analyses.
9. Conducting statistical analyses
10. Interpretation.

Although the steps are presented in a sequential format, they are rather implemented with a level of flexibility that allows for revisiting earlier stages throughout the whole review process. For example, the searches were updated in the latest stages of the project to make sure that no newer relevant publications are missed. This definitely led to a new cycle of coding and data extraction before the final analyses were run. The following sections present the methodology with full details regarding the implementation of each step.

Inclusion and Exclusion Criteria

Similar to all forms of systematic reviews, a set of inclusion and exclusion criteria had to be specified. Lipsey and Wilson (2000) stress that assigning explicit inclusion and exclusion criteria is one characteristic of a good review. This tends to facilitate the communication of the research area of interest and guide the process of inclusion and exclusion of primary research studies.

Keeping in mind that the overall research question was whether technology use enhances student achievement in formal face-to-face classroom settings as compared to traditional settings and if so to what extent, a meta-analysis was considered for inclusion if it:

- Addressed the impact of any form of computer technology as a supplement for in-class instruction as compared to traditional in-class instruction.
- Focused on the impact of the computer technology on students' achievement or performance to the exclusion of cost effectiveness of technology use, or gender differences and student attitudes.

- Dealt with students at different levels of formal education including kindergartens, elementary, high schools, college, and university to the exclusion of workplace or on-the-job training.
- Was published during or after the year 1985.
- Is publicly available or archived.
- Addressed the use of computer technology with learners in regular classroom settings to the exclusion of challenged and gifted students.
- Provided an average effect size that could be extracted.

Meta-analyses that satisfied the above listed criteria were included in the second-order meta-analysis. If any of the above mentioned criteria was not met, the study was disqualified and the reason for exclusion was reported. Reasons for exclusion are summarized by the following list:

- Primary study **(PS)**: a primary study and not a meta-analysis.
- Review **(REV)**: a narrative or qualitative literature review and not a meta-analysis.
- Distance education **(DE)**: addresses technology use in distance education and not in a face-to-face or blended condition.
- Technology in control group **(TCG)**: A meta-analysis which includes studies that have computer technology use in the experimental and control groups.
- Opinion article **(OA)**: Articles that reflect personal opinion regarding technology in education.
- Not technology in education **(NTE)**: A meta-analysis that is addressing educational issues different from technology in education.

- Duplicate **(DUP)**: Article which presents the same data as another meta-analysis or a preliminary report of a subsequently completed meta-analysis (the most comprehensive paper was included, or the more recent if both are as comprehensive).
- Irrelevant **(IRR)**: Meta-analysis addressing issues that are irrelevant for the purposes of the current project, including cost-effectiveness, gender differences, and attitudes to computer technology.
- Not institutionally based **(NIB)**: Meta analysis that has the main focus on technology use for on-the-job training, adult continuous learning or the military or corporate sector.

Specific examples of studies that were excluded based on the above listed criteria will be presented in the section discussing the review process.

Developing and Implementing Search Strategies

Any form of systematic literature review, particularly a meta-analysis, should utilize an adequate search strategy that would help identify relevant studies. This step is extremely critical (Wade, Abrami, Bernard, Turner, & Peretiatkowicz, 2005) and practically determines whether the review will be a comprehensive one or not. The search phase also has an impact on whether the included sample of relevant studies is biased or not (Egger, Juni, Bartlett, Holenstein, & Sterne, 2003). The advantage of having an extensive literature search while addressing different sources was highly evident in the meta-analysis conducted by Bernard et al. (2004) addressing the comparison between distance education and face-to-face instruction (Bernard et al., 2004). In their meta-analysis, Bernard et al. included 232 studies, which would not have been possible has it not been for the adequate and wide-ranging search strategies used.

Similar to a regular meta-analysis an appropriate and adequate search strategy for a second-order meta-analysis would lead to the location of the most relevant body of literature. The use of different sources is also important to ensure a more comprehensive view of the literature since no single source would be able to identify all relevant studies that are potentially eligible for a given research question (Lipsey & Wilson, 2000).

For the purpose of this second-order meta-analysis, a comprehensive search strategy was designed with the help of an information retrieval specialist in order to capture the largest number of meta-analyses addressing the impact of computer technology on learning and educational outcomes. Moreover, the search and retrieval process was iterative and ongoing throughout more than one phase of the project, to ensure the inclusion of as many relevant meta-analyses as possible.

Both electronic and manual searches were conducted using major databases including: ERIC, Education Index, Education index, PsycINFO, PubMed (Medline), EBSCO Academic Search Premier, AACE Digital Library, British Education Index, Australian Education Index, and ProQuest Dissertations and Theses Full-text. Additional searches included the EDITLib, Education Abstracts, and EBSCO Academic Search Premier.

The search strategy included the term "meta-analysis" and its synonyms, including "quantitative reviews" and "systematic reviews". In addition an array of search terms related to computer technology use in educational contexts were used and they varied according to the database searched but generally included terms such as: computer based instruction, computer based teaching, electronic-mail, information communication technology, technology-uses-in-education, electronic learning, hybrid courses, blended learning, teleconferencing, web-based-instruction, technology-integration, and integrated-learning-systems.

For example in ERIC the following terms were used: "Electronic-Mail" or "Electronic-Text" or "Internet-" or "Online-Systems" or "Educational-Technology" or "Technology-Uses-in-Education" or (computer* in de) or "CD-ROMs" or "Calculators-" or "Cybernetics-" or "Data-Processing" or "Electronic-Publishing" or "Electronic-Text" or "Expert-Systems" or "Hypermedia-" or "Multimedia-Materials" or "Online-Systems" or "Telecommunications-" or "Virtual-Reality" or electronic learning or "hybrid courses" or "blended learn*" or "Online-Courses" or "Online-Systems" or "Teleconferencing-" or "Virtual-Classrooms" or" Virtual-Universities" or "Web-Based-Instruction" or "Technology-Integration" or "Technology-Planning" or "Computer-Networks" or "Data-Processing" or "Integrated-Learning-Systems" or "Internet-" or "Local-Area-Networks" or "Communications-Satellites" or "Computer-Mediated-Communication" or "Distance-Education" or "Interactive-Television" or "Online-Courses" or "Open-Universities" or "Telecourses-" or "Virtual-Classrooms" or "Virtual-Universities" or "Web-Based-Instruction".

Web searches were also performed using *Google Scholar* and *Google* search engines. Moreover, manual searches of major journals, including the *Review of Educational Research*, were conducted in addition to branching from major articles and reviews to locate what is known as grey literature. Finally, the *Centre for the Study of Learning and Performance*'s in-house eLEARNing database, compiled as a result of a contract with the *Canadian Council on Learning*, was searched for quantitative reviews and related terms. Searches were updated in November 2008, and results were compiled in a common bibliography.

The search targeted meta-analyses published in the year 1985 onward. The year 1985 was considered as a cut point since it is the time when computer technologies became widely spread and accessible by a vast majority of schools and educational settings (Alessi & Trollip, 2000). Moreover, by that year, although meta-analysis as a procedure was still addressed with scepticism by a group of researchers, it had been established as an acceptable form of quantitative synthesis with clearly specified and systematic procedures. As highlighted by Lipsey and Wilson (2000) this is supported by the fact that the early 1980's witnessed the publication of a variety of books addressing meta-analytic procedures by prominent researchers in the field such as Glass, McGaw, and Smith (1981), Hedges and Olkin (1985), Hunter, Schmidt, and Jackson (1982), Light and Pillemer (1984), and Rosenthal (1984).

Reviewing and Selecting Meta-Analyses

The searches resulted in the location of 429 documents. A variety of approaches have been utilized while reviewing documents for inclusion in a given systematic review. The most comprehensive approach would be to go through the full articles directly which might prove to be extremely time consuming. Examples include the review conducted by Roblyer, Castine, and King (1988) assessing the impact of computer-based instruction and the one conducted by Goldberg, Russell, and Cook (2003) addressing the impact of computers on students' writing. This may also be financially demanding since it necessitates the retrieval of the whole set of documents regardless of whether they are included or not.

Some researchers follow a different approach where they go through the titles of the documents to assess potential relevance and accordingly decide on retrieving articles to be reviewed at full text level. This seems to be more common in the medical and health fields where meta-analysts seem to rely more on such a review process. Examples include a meta-analysis addressing misuse of antibiotic therapies in the community (Kardas, Devine, Golembesky, & Roberts, 2005) and one addressing obesity and asthma incidence in adults (Beuther & Sutherland, 2007). While this is a very efficient approach, it is not highly advisable, particularly in the social science field, because the chances of missing out on a number of relevant articles are high since the ability to judge the relevance of a study based on the title is very questionable.

Another approach to reviewing the literature, which is a compromise between the previous two approaches, entails examining the abstracts to decide on retrieval of a given document as implemented by a variety of researchers including Bernard et al. (2004) in their meta-analysis comparing distance education with classroom instruction. Other examples include the meta-analysis by Blok, Oostdam, Otter, and Overmaat (Blok, Oostdam, Otter, & Overmaat, 2002) addressing computer-assisted instruction in support of beginning reading instruction, and the meta-analysis conducted by Ryan (1991) addressing the effects of microcomputer applications on students' achievement in the elementary classroom. This approach in particular enables the reviewer to have a clearer idea about the research study than the title and allows for a more informed decision about retrieving the full text of the study. Moreover, it permits the reviewer to confidently exclude irrelevant studies; thus, minimizing the overall number of documents to be reviewed at full text level. This will also decrease financial expenses that may ensue from the all-encompassing and more demanding approach of retrieving the full text for the complete set of documents.

For the purpose of this second-order meta-analysis, the abstract review approach was used as a first step for screening identified documents. The review process was conducted by the principal investigator and another colleague. At the time during which the review was conducted, both researchers were PhD candidates at the *Educational Technology Program* at *Concordia University* and have developed ample meta-analytic expertise by being active researchers with the *Systematic Reviews Team* at the *Centre for the Study of Learning and Performance* for five years. To avoid bias and to minimize errors that may lead to overlooking relevant studies, the two researchers worked independently on the abstract review and rated the level of confidence about the decision to retrieve the full texts for the documents using a 5-point scale:

1. Almost definitely unsuitable.
2. Probably unsuitable.
3. Doubtful, but possibly suitable.
4. Most likely suitable.
5. Almost definitely suitable.

Although the abstracts are more informative than titles, on some occasions an abstract did not provide all necessary information. This was augmented by the fact that a variety of terms, such as systematic review and quantitative review, are used synonymously with meta-analysis by some authors. Therefore the deliberate decision was to be widely inclusive during this stage of the project to avoid missing relevant documents. While reviewing the abstracts, the inclusion criteria were considered, and if a document did not meet any of the above inclusion criteria, it was given a "1" or a "2" score and a reason for exclusion was reported as presented in the above provided list of exclusion criteria. Titles whose abstracts were not available were rated "3" to enable further review at full text level to ensure that the decision taken is an informed and reliable one. Ratings by both reviewers were summed for each abstract and those scoring a total of "5" or higher were retrieved for full text review.

Upon reviewing the available abstracts of the 429 identified documents, 158 were labelled for full retrieval. Ratings included disagreements on 62 out of the 429 documents and thus the inter-rater agreement was 85.5% (Cohen's Kappa 0.71).

Once the full texts were retrieved, the decision to include or exclude a given document was relatively straightforward particularly in light of the specificity and clarity of the inclusion/exclusion criteria. Due to limited resources, the full text review was not conducted by the two researchers independently for the whole set of retrieved documents. However, to establish coding reliability both researchers reviewed 15 documents independently, and the inter-rater agreement was 93.3% (Cohen's Kappa = 0.87). The rest of the retrieved full texts were reviewed by the investigator, and in cases where the decision was not straightforward or easy, the second reviewer was consulted for a more confident decision.

From the 158 documents marked for retrieval during the abstract review phase, 12 were not available leaving 146 for full text review, from these 37 met the inclusion criteria and were marked for final inclusion in the second-order meta-analysis.

Examples of Excluded Studies

As presented earlier, studies were excluded based on a variety of criteria. One of the most common reasons marked for excluding documents at full text review level was "REV" indicating that the document was a narrative or qualitative literature review and not a meta-analysis. Examples include the review conducted by Waight, Willging, and Wentling (2002) addressing recurrent themes in e-learning, and the systematic review conducted by Rosenberg, Grad, and Matear (2003) addressing technology use in dental education.

Another common reason was "DUP" indicating that the document at hand was a duplicate of another that was already included. For example, the systematic review conducted by Waxman and his team was published by the Learning Point Associates in two different reports (Waxman, Connell, & Gray, 2002; Waxman, Lin, & Michko, 2003) and only the more recent and comprehensive report was included. An example of studies that were excluded for the technology in both groups include the meta-analysis conducted by Lou, Abrami, and d'Apollonia (2001) addressing small group and individual learning with technology.

A substantive set of meta-analyses were also excluded for addressing students' attitudes and gender differences and not achievement. A number of such meta-analyses were conducted by Liao (Liao, 1999a; 1999b; 2000). Some of the located meta-analyses were excluded because of their focus on distance education settings such as (Bernard et al., 2004; Ungerleider & Burns, 2003). Only one meta-analysis was excluded for its emphasis on special needs students, namely the dissertation completed by Wolf (2006).

Extracting Effect Sizes and Standard Errors

This section presents the procedures that were carried out in this second-order meta-analysis for the extraction of the effect sizes and standard errors from the included meta-analyses.

Effect Sizes

An effect size is the metric introduced by Glass and it represents the difference between the mean of the experimental group and the control group in standardized units. A major advantage is the ability to convert an effect size to a percentile gain of the treatment group compared to the control group. Another benefit is the fact that an effect size is not highly related to sample size, thus one would not get a significant finding based on large sample size only. Furthermore, an aspect that is highly important for meta-analysis is the ability to aggregate effect sizes and subject them to further statistical analyses in order to explain and understand the variation in a population (Lipsey & Wilson, 2000).

The first method for calculating effect sizes was proposed by Glass (1977) and it entails dividing the difference between the experimental group and the control group by the standard deviation of the control group, since it is the untreated condition.

$$\Delta = \frac{\overline{Y}_E - \overline{Y}_C}{SD_C}$$

However, this led to certain overestimation or underestimation when variances were not similar in the two groups. Cohen proposed calculating the effect size by dividing the difference by the pooled standard deviation, which would correct for the first bias. This type of effect size is known as Cohen's *d*.

$$d = \frac{\overline{Y}_E - \overline{Y}_C}{SD_{Pooled}}$$

where pooled standard deviation is calculated by applying the following formula:

$$SD_{Pooled} = \sqrt{\frac{(n_E - 1)SD_E^{\,2} + (n_C - 1)SD_C^{\,2}}{(n_E - 1) + (n_C - 1)}}$$

A further modification was introduced by Hedges to overcome the problem of overestimation with samples smaller than 20 individuals, and his effect size is known as Hedges' *g*. The correction is achieved through the use of a coefficient based on (1- inverse of sample size), so that the larger the sample size the smaller the correction coefficient. With this in mind, Hedges' *g* can be used for large samples since the correction coefficient gets to be closer to 1, and then the number will be the same as Cohen's *d*.

$$g \cong d\left(1 - \frac{3}{4N - 9}\right)$$

For the purpose of this second-order meta-analysis the effect sizes from different meta-analyses were extracted while noting which type of effect size it was. In a perfect situation, where authors provide adequate information, it might have been possible to transform all of the three types of group comparison effect sizes to one type, preferably Hedges g. However, due to reporting limitations, this was hardly possible, particularly when Glass's Δ was used in a given meta-analysis. Keeping in mind that all three (Δ, *d*, *g*) are just variations for calculating effect sizes for differences between two groups, and assuming that the sample sizes were large enough to consider the differences between the three forms to be minimal, it was decided to use the effect sizes in the forms that they were reported in.

However, a few of the included meta-analyses expressed the effect size as a standard correlation coefficient which is not conceptually compatible with either Glass's Δ, Cohen's d or Hedges g. In these instances, the reported effect size was converted to *Cohen's d* by applying the following formula (Bernard & Abrami, 2009):

$$d = \frac{2r_{XY}}{\sqrt{1 - r_{XY}^{2}}}$$

One of the meta-analyses, namely that conducted by LeJeune (2002) included two separate meta-analyses addressing the differences between the use of computer-simulated experiments and traditional learning activities on student achievement outcomes relating to low and high level thinking skills. This allowed for the extraction of two independent effect sizes, and thus, although the overall number of included meta-analyses was 37, the total number of effect sizes was 38.

Moreover, in different meta-analyses, authors reported sub-effect sizes based on various specific study features. For the purpose of this second order meta-analysis it was decided to extract the specific effect sizes pertaining to subject matter, grade level, and type of technology whenever they were reported. This was based on the fact that these features were the most recurring study features in the literature for which individual effect sizes were reported. Throughout this report, these effect sizes will be referred to as specific effect sizes. In particular situations, where authors reported a specific effect size based on less than three studies, that specific effect size was ignored.

Standard Errors

Standard Error and Second-Order Meta-Analysis

One of the issues in a meta-analysis has to do with the fact that effect sizes calculated from larger samples are better estimates than those calculated from studies with smaller samples. If a simple average is used to compute the point estimate, then this allows all effect sizes to contribute equally to the point estimate which is not appropriate given the different levels of reliability that each reflects (Lipsey & Wilson, 2000). This problem may be solved with a simple weighting by sample size, however, Hedges and Olkin (1985) have stressed that the best approach is to use weights based on the standard error of the effect size. The standard error is the "standard deviation of the sampling distribution (the distribution of values we would get if we drew repeated samples of the same size and estimated the statistic for each)" (Lipsey & Wilson, 2000, p. 36). The standard error of g is calculated by applying the following formula:

$$\hat{\sigma}_g = \sqrt{\frac{1}{n_e} + \frac{1}{n_c} + \frac{g^2}{2(n_e + n_c)}}\left(1 - \frac{3}{4(n_e + n_c) - 9}\right)$$

In a regular meta-analysis the application of the formula is quite straight forward, where n_e refers to the number of participants in the experimental group and n_c refers to the number of participants in the control group.

However, in a second-order meta-analysis, things are not as simple since we have two different types of variances, one reflecting the variability at the study level, and one reflecting the variability at the meta-analysis review level. For the study variance, the sample sizes to be used in the standard error computation are clearly the numbers of participants in a given study. This will be reflective of the variability among all the individual effect sizes calculated and incorporated in the collection of included meta-analyses.

In a meta-analysis, the study is the unit of analysis and the sample size is the number of included studies, and the meta-analysis review level variance has to be reflective of the variability based on the number of included studies in each meta-analysis. Therefore the standard error computation should be based on the number of studies included in a given meta-analysis. However, how many control and how many experimental studies is one to consider? Knowing that each study is contributing a control and an experimental group it is logical to use the same number of studies as both experimental and control. Calculations based on this approach will be reflective of the variability among all the average effect sizes (point estimates) calculated in the included meta-analyses.

The question is which standard error should be used for the purpose of the second order meta-analysis; that computed based on sample sizes in original studies or number of studies in meta-analyses. The former approach makes use of the strength of meta-analyses and allows for reliable conclusions, while keeping the enormous variability in the study findings intact thus magnifying the heterogeneity in the findings. On the other hand, the latter approach does not overstate heterogeneity but it ignores the actual strength offered by the individual point estimates from the different meta-analyses.

For the purpose of this second-order meta-analysis, standard errors reflecting both variances were calculated to allow for conducting and comparing analyses with both approaches. To avoid confusion between the two, the standard error based on the use of sample sizes of participants in primary studies will be referred to as *sample-size standard error*, while that based on the number of studies in a given meta-analysis will be referred to as *number-of-studies standard error*.

Sample-Size Standard Error

While extracting information from the included meta-analyses, there were 4 different situations pertinent to the sample size standard error. In the first instance, the authors of a given meta-analysis reported the sample size standard error with the mean effect size. In such cases, the extraction of the information was straightforward and from the total of 37 included meta-analyses, standard error was reported in eight different documents. An example is the meta-analysis conducted by Cohen and Dacanay (1992) addressing computer-based instruction in health education.

In other documents, authors reported the individual effect sizes and the corresponding sample sizes for the included primary studies. Individual standard errors where calculated using the following formula (Bernard & Abrami, 2009):

$$\hat{\sigma}_g = \sqrt{\frac{1}{n_e} + \frac{1}{n_c} + \frac{g^2}{2(n_e + n_c)}}\left(1 - \frac{3}{4(n_e + n_c) - 9}\right)$$

The overall variance was calculated by applying the following formula (Bernard & Abrami, 2009):

$$\hat{\sigma}^2_{g+} = \left(\sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2}\right)^{-1} = \frac{1}{\sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2}} = \frac{1}{\sum_{i=1}^{k} w_i}$$

After that, the standard error of the point estimate was determined by calculating the square root of the overall variance.

The authors generally provided the overall sample size in the primary studies without indicating the experimental versus control group sample sizes. To overcome the missing information problem, it was assumed that the experimental and control groups were equal in size and in the case of an odd overall number of participants it was reduced by one. From the total of 37 included meta-analyses, the standard error was calculated based on individual effect sizes and sample sizes for 12 different meta-analyses. An example is the meta-analysis conducted by Christmann and Badgett (2003) addressing the effects of computer-assisted instruction on elementary students' academic achievement.

Whenever individual effect sizes and corresponding sample sizes were not offered by an author, the confidence interval was looked for. If provided, the standard error was calculated based on either one of the following formulae:

$$Lower = g + -1.96(\hat{\sigma})$$
$$Upper = g + +1.96(\hat{\sigma})$$

From the total of 37 included meta-analyses, the standard error was calculated from confidence intervals for three different meta-analyses. An example is the meta-analysis conducted by Zhao (2003) addressing the development in technology and language learning. The mean effect size was 1.12 and the confidence interval was 0.61 to 1.63, and upon calculation, the standard error was 0.260.

Finally, there were cases where neither individual effect sizes and sample sizes nor confidence intervals were provided and therefore standard error imputation was needed.

The decision was to use the weighted average standard error of the included meta-analyses. For calculating the weighted mean standard error, two approaches were possible, either to weight by number of studies or the number of participants included in each meta-analysis. Knowing that the number of participants was the bases for standard error calculations in the above three cases, and keeping in mind that average effect sizes from different meta-analyses are reflective of the total number of participants rather than the number of studies included the decision was to calculate the number of participants weighted average standard error. Based on the meta-analyses where both number of participants and number of studies were known, it was found that the average number of participants per study was 104 individuals. With this, the missing number of participants in a given meta-analysis was calculated by multiplying the number of studies by the average number of participants per study, namely 104. Having imputed missing values in number of participants it was easy to calculate the number of participants weighted average standard error which was 0.051. From the total of 37 included meta-analyses, the missing standard error was imputed with the weighted average standard error of 0.051 for 14 different studies.

Finally, in a few cases where the known sample sizes were extremely large, they were replaced by a more conservative sample size that was equal to five times the average number of participants per study to avoid their dominating effect on the weighted average effect size. For example, a study included in the meta-analysis conducted by Schenker (2007) had a sample size of 5597 participants, and it was replaced by 520.

Number-of-Studies Standard Error

The number-of-studies standard errors were all calculated by using the information extracted from the included meta-analyses by applying the following formula (Bernard & Abrami, 2009):

$$\hat{\sigma}_g = \sqrt{\frac{1}{n_e} + \frac{1}{n_c} + \frac{g^2}{2(n_e + n_c)}}\left(1 - \frac{3}{4(n_e + n_c) - 9}\right)$$

As presented earlier, knowing that each study is contributing a control and an experimental group the total number of studies included in each meta-analysis was used as both the experimental and control sample size.

Finally, it is important to note that the extraction of all the effect sizes (overall and specific) and standard errors was performed by the two coders independently to ensure reliability. Inter-rater agreement was 98.6 % (Cohen's Kappa = 0.97). Due to limited resources, the calculation of the missing sample-size standard errors and the number-of-studies standard errors was conducted by the principal investigator. To avoid mistakes, random spot checks were done for the data entry into the excel file. Moreover, it is important to note that missing sample-size standard errors were calculated or imputed for overall effect sizes and not for the specific ones due to the vast number of missing information at the specific effect size level.

Developing *a codebook*

Design Process

Any given meta-analysis requires a specific codebook or coding protocol to help in the process of extracting relevant information from the included primary literature. Information collected with the help of the codebook will provide a means for summarising the findings from the included documents as well as explaining part of the variability in the phenomenon under investigation (Lipsey & Wilson, 2000). This should be equally valid in the case of a second-order meta-analysis. The purpose of the codebook in this study is to provide explicit criteria for the process of systematic extraction of sufficient information from the included meta-analysis to allow for: a) synthesizing the findings from the different meta-analyses; b) critically analyzing the quality of the included meta-analyses; and c) explaining the variability in the findings if possible.

To help in the design of the current codebook, three main sources were consulted:

1. Meta-analysis literature pertaining to procedural aspects.
2. Currently published second-order meta-analyses.
3. Available standards or tools for assessing the methodological quality of meta-analyses.

Meta-Analysis Procedures

The literature pertaining to meta-analytic procedures was helpful in highlighting the different phases and steps that should be addressed in our current codebook. As presented in the literature review, a variety of meta-analytic procedures have been established over the years. According to Bernard and Naidu (1990) the steps to be followed in the implementation of a meta-analysis include:

- Specifying the research question.
- Developing and conducting search strategies, which should be as comprehensive and thorough as possible.
- Creating inclusion/exclusion criteria that are related to the research question and help in deciding which documents will be included in a given meta-analysis.
- Reviewing and selecting studies to be included in the meta-analysis which may prove to be complicated and having more than one researcher working simultaneously on this task will prove helpful in avoiding personal bias.
- Extracting and calculating effect sizes with the help of a variety of formulae.
- Developing a codebook which should include demographic, treatment, and design variables and coding study features.
- Conducting statistical analysis and interpretation, this may include in addition to the average effect size the test for homogeneity of effect sizes, moderator analyses and meta-regression.

Throughout the different phases of review and data extraction, it is important to have multiple coders to ensure reliability, and avoid personal bias and unintentional oversights or mistakes (Rosenthal, 1984).

Published Second-Order Meta-Analyses

The examination of the existing collection of second-order meta-analyses was highly informative and offered the chance to take advantage of what has been done so far in this area. A critical review of published second-order meta-analyses revealed that there is a wide variety in the specificity and comprehensiveness of reported codebooks which is not surprising given the fact that this methodology is not widely spread yet. For example, Møller and Jennions (2002) do not give any reference to their codebook while Sipe and Curlette (1997) and Steiner et al. (1991) offer an extensive explanation of their codebooks that are relatively quite lengthy and comprehensive. A summary of the features addressed in each of the published second-order meta-analyses is presented in Table 2.

Methodological Quality of Meta-Analyses: Tools and Standards

As noted in the literature review, some assessment tools have been designed to assess the methodological quality of meta-analyses in the medical and health areas with an absence of anything parallel in the social science field.

One of the most prominent tools used to assess methodological quality of meta-analyses in the medical field is *The Quality of Reporting of Meta-analyses* (QUOROM) which was the outcome of a group conference consisting of 30 clinical epidemiologists, clinicians, researchers, statisticians, and editors. The group's objective was to identify items for a checklist of standards that would be used to assess and evaluate the quality of a meta-analytic report to help in improving the quality of reviews of randomized control trials in the medical field.

The QUOROM checklist reflects the panel's preferences of how a meta-analysis should be reported and focuses on descriptors for the different sections to be included in the report. Sections addressed by the QUOROM include the title, abstract, introduction, methods, results and discussion. For each section there is a set of descriptors for the subsections which the evaluator would assess on a dichotomous yes/no basis to

Table 1. Summary of study features coded in different second-order meta-analyses

Reference	Topic	# MA included	Codes	
Barrick, Mount, & Judge, 2001	Relationship between personality and performance	11	- Performance criteria - Occupational groups	- Effect size - Sample size
Peterson, 2001	The use of college students versus non-college students in social science research	4 28	- Relationship assessed	- Sample size
Luborsky et. al, 2002	Dodo Bird effect of psychotherapy treatments	17	- Number of studies	- Effect size - Sample size
Hammill, 2004	Relation of specific abilities and reading	3	- Time span of studies reviewed - Total studies included - Concurrent studies (coefficients) reviewed - Longitudinal studies (coefficients) reviewed	- Conducted a broad search for articles - Controlled for number of mean effects in study - Investigated moderator variables
Møller & Jennions, 2002	How much variance explained by ecologists and evolutionary biologists	43	No reference to codes	

Lipsey & Wilson, 1993	Efficacy of psychological, educational, and behavioural treatments	290	- number of studies included - methodological quality of included studies - publication bias of included studies	- Small sample bias - Generalized placebo effect
Wilson & Lipsey, 2001	Influence of study method on observed ES (using psychological, behavioural and educational treatments)	319	- ES with variance around grand mean ES - Type of research design - Type of treatment - Type of outcome	- Different type of samples - Number of ES - Number of studies - Number of respondents
Steiner, Lane, Dobbins, & McConnell, 1991	Organizational behaviour and human resources management	35	- # of studies included - # of participants - Rating on theoretical basis - Method of identifying studies - Years covered in literature search - Inclusion of Rosenthal's file drawer calculation - Criteria for eliminating studies - Rating on coding scheme - Calculation of inter-rater agreement for coded variables - Independence of data - Artefacts controlled	- Type of meta-analysis - Inclusion of list of included studies - Inclusion of table of ES - Inclusion of percentage of variance accounted for by effects and artefacts - Tests of moderators conducted - Test of sig of residual variances (hunter et al. procedures) - Procedure used to identify moderators - Rating on discussion of alternative explanations - Rating on discussion of generalizing beyond study

Sipe & Curlette, 1997	Factors related to educational achievement	103	- Background characteristics:
			  - Year of publication
			  - Author
			  - University affiliation
			  - Document type
				- Varying information for total, treatment, and criterion level ES
				  - Construct domain
				  - Treatment domain
				  - Criterion variable

			- Journal of publication - Funding source - Page length - Subject area - Methodological characteristics - Type of meta-analysis - Research question - Meta-analysis procedure used - Sample selection procedures - Control of bias - Discussion section - Extreme scores or outliers - Miscellaneous information - Contextual variables - Age - Gender - Grade level - Level of education - Location - SES - Model-related classification - Characteristics of ES and computations - Characteristics of effect size - Statistical procedures concerning effect size - Further analyses - Effect size metric - Mean ES - Median ES - SD - SE - N of sub-analyses - N of primary studies - N of persons - Varying information for characteristic ES - Characteristics and level - Method of significance testing - Outcome of significance testing - Characteristic mean ES - Characteristic median ES - SD - SE - N of sub-analyses - N of primary studies - N of persons

indicate whether the author of the meta-analysis reported the aspect under investigation. From a practical perspective, one of the challenges in using the QUOROM to assess the methodological quality of a given meta-analysis lies in the fact that the descriptors are loaded and address more than one procedural aspect at the same time. For example in the methods section, the descriptor for the searching subheading, addresses both the resources as well as the search limitations. It is not hard to imagine a meta-analysis where the resources were listed but not the limitations and vice versa. In such a case it will not be easy to assign a yes or a no to that given report, and the reliability of whatever code given gets to be questionable. Nevertheless, the aspects addressed by the QUOROM are highly valid in both the medical and social science meta-analytic arenas.

Another prominent tool in the health science area is that developed by health.evidence.ca investigative team. As part of their mandate to help research consumers become more capable decision makers, they offer free access to their evaluation tool and its dictionary. *The Quality Assessment Tool Review Articles* addresses both qualitative and quantitative literature reviews, and similar to the QUOROM it uses a dichotomous yes/no scoring system for each aspect. The dictionary offers detailed definitions of the terms and an explanation of how to use the assessment tool. It also provides the information for the overall assessment of a given review. The final score is based on the number of "yes" rates a study is given. The total is 10 points, with a score 7 or above reflecting a *Strong quality* review, a score between 5-6 reflecting a *Moderate quality* review, while a score of 4 or less reflects a *Weak quality* review. Although the dichotomous score provides a similar challenge as that of the QUOROM, the dictionary minimizes it by offering clear criteria for the rating process.

Other researchers address the quality of systematic reviews from a global and more general approach than a checklist format. Schlosser (2007) highlights different aspects of a systematic review that should be used as criteria for appraising reviews. The criteria include: a) the presence of a protocol; b) the clarity of the research question; c) the comprehensiveness of the sources to avoid publication bias; d) the scope of the review as reflected by the inclusion/exclusion criteria; e) the presence of temporal or time constraints particularly regarding the start day; f) selection principles used while including the studies; and g) the data extraction procedures from the primary studies, including the reliability of the process.

The Codebook

Due to the complexity of the task at hand, the codebook had to be designed in a way to include a range of codes that will help in attaining the specified objectives. As presented earlier, a variety of resources was reviewed and checked for possible assistance in the design of the current codebook. These included regular procedures used for coding data in primary meta-analyses, the codes presented in the variety of published second-order meta-analysis, and criteria used to assess the methodological quality of systematic reviews.

The overall structure of the developed codebook was highly influenced by the synthesis of meta-analyses conducted by Sipe and Curlette (1997). The guidelines and standards for meta-analytic procedures were helpful in deciding on specific study features to be included in the codebook particularly addressing methodological features. As for the methodological evaluation tools, they were informative regarding aspects that need to be included in the codebook in addition to approaches for creating an overall methodological quality index.

The four main sections of the codebook are: a) study identification; b) contextual features; c) methodological feature; and d) effect size information. Details about each of these sections and their components are presented in the following paragraphs, and the full codebook is attached in Appendix A.

Study Identification

This section addresses descriptive information regarding the authors and the publication venues of the meta-analyses. Categories within this section include:

- Identification number: A unique number created and assigned by the Endnote reference management software for each document to help in easily identifying the documents that are processed throughout the different stages of this project starting with the literature search and ending with the final set of meta-analyses included in the second-order meta-analysis.
- Author: An open code that provides the full reference to the author(s) of the different documents.
- Title: An open code that provides the full title of the document at hand.
- Year of publication: An open code that provides a document's year of publication.
- Type of publication: A code that specifies the type of publication, and it can take any of the following codes:

  1. Journal
  2. Dissertation
  3. Conference Proceedings
  4. Report / Grey literature

Contextual Features

This section addresses descriptive information regarding the settings and participants addressed by a particular meta-analysis. Categories within this section include:

- Research question: An open code that aims at summarizing the main research question to be answered by a meta-analysis.
- Technology addressed: An open code which summarizes the technology addressed in a certain meta-analysis.
- Control group definition/description: An open code that summarizes the definition or the description of the control group in a meta-analysis whenever it is provided by the authors.
- Experimental group treatment definition/description: An open code that summarizes the definition or the description of the experimental group in a meta-analysis whenever it is provided by the authors.
- Grade level: An open code where the grade level(s) addressed by a meta-analysis is/are listed. Based on the extracted codes, categories will be created
- Subject matter: A code that specifies the subject matter addressed by a meta-analysis, and it can take any of the following codes:

    1. Science/Health
    2. Languages
    3. Math
    4. Technology
    5. Social Science
    6. Combination
    7. Information Literacy
    8. Engineering
    9. Not specified

Methodological Features

This section addresses information about the steps followed in the primary meta-analysis. To ensure that the codebook is comprehensive of all the procedural aspects of a meta-analysis, the different phases of conducting a meta-analysis and reporting the findings were targeted in separate sub-sections: a) search phase; b) review phase; c) effect size and study feature extraction phase; d) analysis; and e) reporting aspects.

The study features pertaining to the procedural aspects of the meta-analysis, were specifically designed to capture and reflect the methodological quality and adequacy of a meta-analysis. To make the coding procedure as unambiguous as possible, particularly in relation to methodological features of the included MAs, the levels for each code were designed to be very specific. The five point scale used by Steiner et al. (1991) to rate methodological features of included meta-analyses helped in designing some of the scales used in the current codebook. For each of the included features addressing methodological aspects, the levels were listed from the least to the most methodologically appropriate. The more methodologically adequate procedures were those that reflect a higher level of comprehensiveness, accuracy, and transparency in reporting the procedure and offer higher reliability of the findings. Categories within this section include:

Search phase

- Search time frame: An open code that aims at specifying the search time frame used in a meta-analysis.
- Justification for search time frame: A code that captures whether an author provided a justification for the search time frame whenever it was available. It can take either one of the following two levels:

1. No
2. Yes

- Literature covered: A code that specifies the type of literature covered and included in a meta-analysis and it can take either one of the following two categories:
  1. Published studies only
  2. Published and unpublished studies
- Search strategy: A code that reflects the level of clarity and transparency pertaining to the description of the search strategy, and it can take any of the following levels:
  1. Search strategy not disclosed, no reference to search strategy offered
  2. Minimal description of search strategy with brief reference to resources searched
  3. Listing of resources and databases searched
  4. Listing of resources and databases searched with sample search terms
- Resources used: A code that captures the comprehensiveness of the data resources searched, with more resources reflecting a higher level of comprehensiveness. It can take one or more of the following categories:
  1. Data-base searches
  2. Computerized search of web resources
  3. Hand search of specific journals
  4. Branching

- Databases searched: An open code where databases searched to locate relevant studies are listed.
- Number of data-bases searched: An open code that aims at providing the number of databases searched with the assumption that more databases are usually reflective of a higher level of comprehensiveness.

Review phase

- Inclusion/exclusion criteria: A code that addresses the clarity of the criteria used to include and exclude research studies in a meta-analysis and it can take any of the following levels:

    1. Criteria not disclosed with no description offered
    2. Overview of criteria presented briefly
    3. Criteria specified with enough detail to allow for easy replication

- Included research type: A code that aims at specifying the types of research design included in a meta-analysis. The most restrictive being the least inclusive and the more comprehensive would be the more inclusive approach. It may take any of the following levels:

    1. RCT only
    2. RCT/Quasi
    3. RCT/Quasi/Pre

    999. Not specified

- Article review: A code that aims at capturing the rigour and reliability of the review process for the inclusion of studies in a meta-analysis and it may take any of the following levels:

1. Review process not disclosed
2. Review process by one researcher
3. Rating by more than one researcher
4. Rating by more than one researcher with inter-rater agreement reported

Effect size and study feature extraction phase

- ES Extraction: A code that aims at capturing the rigour and reliability of the ES extraction process in a meta-analysis and it may take any of the following levels:
  1. Extraction process not disclosed, no reference to how it was conducted
  2. Extraction process by one researcher
  3. Extraction process by more than one researcher
  4. Extraction process by more than one researcher with inter-rater agreement reported
- Code Book: A code that addresses the clarity of the codebook used to extract study features in a meta-analysis and it can take any of the following levels:
  1. Codebook not described, no reference to features extracted from primary literature
  2. Brief description of main categories in codebook
  3. Listing of specific categories addressed in codebook
  4. Elaborate description of codebook allowing for easy replication
- Study feature Extraction: A code that aims at capturing the rigour and reliability of the study feature extraction process in a meta-analysis and it may take any of the following levels:
  1. Extraction process not disclosed, no reference to how it was conducted

2. Extraction process by one researcher
3. Extraction process by more than one researcher
4. Extraction process by more than one researcher with inter-rater agreement reported

Analysis

- Independence of data: A code that reflects whether the issue of dependency of ES data was addressed in a meta-analysis or not. It can take either one of the following codes:
  1. No
  2. Yes
- Weighting by number of comparisons: A code that targets a controversial method that was used by a number of meta-analysts to overcome the predicament of studies having higher weights due to extracting effect sizes from multiple non-independent comparisons from the same study such as Waxman, Lin, and Michko's meta-analysis (2007). The literature does not reflect any support for this methodology, which although solves the issue of overweighting some studies rather aggravates the dependency issue. It is also important to note that in all of these meta-analyses there was no mention of any reference that supports such an approach of weighting which only augments the problem. Because of its procedural inappropriateness, if a meta-analysis applied this approach it was considered methodologically less appropriate and was given the code 1 as presented in the following two levels:
  1. Yes

2. No

- Effect size weighted by sample size: A code that reflects whether ES were weighted by sample size in the calculation of the average ES, and it can take either one of the following codes:

1. No
2. Yes

- Homogeneity analysis: A code that specifies whether homogeneity analysis was conducted in a meta-analysis, and it can take either one of the following codes:

1. No
2. Yes

- Moderator analysis: A code that specifies whether moderator analysis was conducted in a meta-analysis, and it can take either one of the following codes:

1. No
2. Yes

- Meta-regression conducted: A code that specifies whether moderator analysis was conducted in a meta-analysis, and it can take either one of the following codes:

1. No
2. Yes

Further reporting aspects

- Inclusion of list of studies: A code that reflects the quality of the report by addressing whether a list of included studies was provided, and it can take either one of the following two levels:

1. No

2. Yes

- Inclusion of ES table: A code that reflects the quality of the report by addressing whether a list of included studies was provided, and it can take either one of the following two levels:

  1. No
  2. Yes

- Time between last study and publication date: An open code addressing the currency and contemporariness of a meta–analysis by specifying the time between the last study included and the publication date. While acknowledging the time needed for a study to move through the publication process, one needs to keep in mind that there is a limit beyond which the research findings may not be representative of the current situation at the time of publication.

Effect Size Information

This section addresses information about the overall effect sizes and the specific effect sizes for different levels of certain variables computed in each meta-analysis.

- Effect size type: A code that specifies which type of ES was utilized by the author(s) of a meta-analysis, and it can take any of the following categories:

  1. Glass
  2. Cohen
  3. Hedges
  4. Others: specify

Total ES

- Mean ES: An open code for reporting the mean ES extracted from a meta-analysis.

- SE: An open code for reporting the SE extracted from a meta-analysis.
- SE extraction: A code to reflect the process by which the SE was extracted for a meta-analysis:
    1. Reported
    2. Calculated from ES and sample size
    3. Calculated from Confidence interval
    4. Imputed with weighted average SE
- Time frame included: An open code to reflect the time frame of the included studies which might be different from the search time frame due to a variety of reasons.
- Number of studies included: An open code that reports the specific number of included studies in a meta-analysis.
- Number of ES included: An open code that reports the specific number of ES included in a meta-analysis.
- Number of participants: An open code that reports the specific number of participants included in a meta-analysis.
- Number of participants extraction: A code that specifies how the number of participants was extracted and it can take either one of the following two categories:
    1. Calculated
    2. Given

Specific Effect Size

- Specific variable: An open code that states the specific variable on which a sub-effect size was based. The main focus of this code was either subject matter, grade level, or type of technology.

- Mean ES: An open code for reporting the mean ES extracted from a meta-analysis.
- Standard error: An open code for reporting the standard error extracted from a meta-analysis.
- Standard error extraction: A code to reflect the process by which the standard error was extracted for a meta-analysis:
    1. Reported
    2. Calculated from effect size and sample size
    3. Calculated from Confidence interval
    4. Imputed with weighted average SE
- Time frame included: An open code to reflect the time frame of the included studies which might be different from the search time frame due to a variety of reasons.
- Number of studies included: An open code that reports the specific number of included studies in a meta-analysis.
- Number of ES included: An open code that reports the specific number of ES included in a meta-analysis.
- Number of participants: An open code that reports the specific number of participants included in a meta-analysis.
- Number of participants extraction: A code that specifies how the number of participants was extracted and it can take either one of the following two categories:
    1. Calculated
    2. Given

Coding Study Features

Since the early stages of establishing procedural guidelines for conducting meta-analyses, the importance of having multiple coders to ensure reliability, and avoid personal bias and unintentional oversights or mistakes throughout the different phases of review and data extraction have been stressed (Rosenthal, 1984). With that in mind, the process of coding study features was conducted by both researchers. A random sample of 3 studies was used to help in training and setting the overall standards and ensuring that both researchers have common understanding of the different study features. Next each researcher extracted full information for each of the included meta-analysis. Inter-rater agreement was 98.7% (Cohen's Kappa = 0.97). After completing the coding independently the two researchers met to resolve any discrepancies.

Designing and Calculating the Methodological Quality Index

As presented in previous sections, while developing the codebook, the different codes were designed to address various methodological aspects pertaining to meta-analysis such as comprehensiveness, scope, contemporariness, accuracy and detail. The ideal situation would be to create a separate index for each of the above methodological aspects of a meta-analysis. Nevertheless, due to the complexity of the process and the overlap between the different constructs, it is not easy to design orthogonal indexes that may be used later to explain variability. Therefore, the overall index approach used by the *Health Evidence Group* was utilized in this second-order meta-analysis while incorporating all the features addressing the different aspects within the overall methodological index. However, to calculate an overall index the codes had to be transformed to a dichotomous yes versus no codes, similar to the *Health Evidence Tool.* The list of 14 items included in the methodological quality index and how the transformation occurred whenever there were more than 2 levels is presented in Table 3. In the dichotomous coding approach level 1 represents the lower methodological quality feature while level 2 represents the higher methodological quality feature. The maximum score a meta-analysis may get for methodological quality is 14. These 14 items may be used in the design and development of a methodological quality tool in a checklist format, and which could be coupled with a dictionary similar to the *Health Evidence Tool* approach. In the dictionary, the original descriptions used in the codebook may be used to help the reader and user decide on what the verdict is for each meta-analysis. The potential use of these items in a future methodological quality tool and its uses will be addressed more elaborately in the discussion section.

Similar to the *Health Evidence Tool,* the score was also changed into a categorical index where:

- A meta-analysis scoring between 1 and 5 is rated as a weak review
- A meta-analysis scoring between 6 and 9 is rated as an average review
- A meta-analysis scoring between 10 and 14 is rated as a strong quality review

For each of the included meta-analyses, the overall score as well as the categorical methodological quality index were calculated for future use in the analyses. The overall and categorical scores for all the included meta-analyses are presented in Table 4.

Table 2. Items included in the methodological quality index and their transformation to dichotomous levels

	Study Feature	**Original Levels**	**Dichotomous Levels**
1.	Literature covered	1. Published studies only 2. Published and unpublished studies	1. Published studies 2. Published studies and unpublished studies
2.	Resources used (may have more than one in a given meta-analysis)	1. Data-base searches 2. Computerized search of web resources 3. Hand search of specific journals 4. Branching	1. One or two different resources were used 2. three or four different resources used
3.	Included research type	1. RCT only 2. RCT/Quasi 3. RCT/Quasi/Pre 999. Not specified	1. RCT only 2. Broader range of designs
4.	Number of data-bases searched	Open code	1. One or two databases 2. Three or more databases
5.	Search strategy	1. Search strategy not disclosed, no reference to search strategy offered 2. Minimal description of search strategy with brief reference to resources searched 3. Listing of resources and databases searched 4. Listing of resources and databases searched	1. Search strategy not disclosed or very brief description 2. Listing of resources and databases searched

		with sample search terms	
6.	Inclusion/exclusion criteria	1. Criteria not disclosed with no description offered 2. Overview of criteria presented briefly 3. Criteria specified with enough detail to allow for easy replication	1. Criteria not disclosed 2. Overview of criteria presented
7.	Article review	1. Review process not disclosed 2. Review process by one researcher 3. Review process by more than one researcher 4. Review process by more than one researcher with inter-rater agreement reported	1. Review process not disclosed or by one researcher 2. rating by more than one researcher
8.	ES Extraction	1. Extraction process not disclosed, no reference to how it was conducted 2. Extraction process by one researcher 3. Extraction process by more 9than one researcher 4. Extraction process by more than one researcher with inter-rater agreement reported	1. Extraction process not disclosed or by one researcher 2. Extraction process by more than one researcher
9.	Code Book	1. Codebook not described, no reference to features extracted from primary literature 2. Brief description of main categories in codebook 3. Listing of specific categories addressed in	1. Absence of or brief description of features 2. Listing and/or description of categories

		codebook 4. Elaborate description of codebook allowing for easy replication	
10.	Study feature Extraction	1. Extraction process not disclosed, no reference to how it was conducted 2. Extraction process by one researcher 3. Extraction process by more than one researcher 4. Extraction process by more than one researcher with inter-rater agreement reported	1. Extraction process not disclosed or by one researcher 2. Extraction process by more than one researcher
11.	Independence of data	1. No 2. Yes	1. No 2. Yes
12.	Weighting by # of comparisons	1. Yes 2. No	1. Yes 2. No
13.	Time between last study included and publication	Open code	1. Four years or more 2. Three years or less
14.	SE calculations	1. Reported 2. Calculated from effect sizes and sample sizes 3. Calculated from Confidence interval 4. Imputed with weighted average standard error	1. Imputed 2. Reported or calculated

Beyond the overall methodological quality index and the categorical rating, and for the purpose of later analyses, the items were grouped into two sets of features based on theoretical similarity and practical overlap, one addressing methodological aspects that impact the comprehensiveness of a meta-analysis and its rigour.

The set targeting comprehensiveness included seven items that addressed the comprehensiveness and the representativeness of a meta-analysis of the overall set of primary studies targeting the area under study. Particularly it included the following items: a) literature covered; b) search strategy; c) resources used; d) number of databases searched; e) inclusion/exclusion criteria; f) included research; and g) the time between last included study and publication date. If a meta-analysis scored high on five or more of the listed items, it was considered to be of high quality regarding comprehensiveness. Alternatively, if a meta-analysis scored high on four or less of the listed items it was considered to be of low quality regarding comprehensiveness.

The set targeting rigour aspects included seven items that addressed the level of rigour applied in the implementation of a given meta-analysis. Particularly it included the following items: a) article review; b) effect size extraction; c) codebook; d) study feature extraction; e) independence of data; f) standard error calculation; and g) weighting by number of comparisons. If a meta-analysis scored high on five or more of the listed items, it was considered to be of high quality regarding the rigour. Alternatively, if a meta-analysis scored high on four or less of the listed items it was considered to be of low quality regarding the rigour.

Table 4. Overall and categorical methodological quality for the included meta-analyses

Meta-Analysis	Methodological quality index	Methodological quality categorical	Quality Rating
Christmann & Badgett (2000a)	5	1	Weak
Niemiec (1987)	5	1	Weak
Rosen & Salomon (2007)	5	1	Weak
Samson et al.(1986)	5	1	Weak
Christmann et al.(1997)	6	2	Average
Lee (1999)	6	2	Average
Liao (1992)	6	2	Average
Zhao (2003)	6	2	Average
Christmann & Badgett (2000b)	7	2	Average
Christmann & Badgett (2003)	7	2	Average
Liao (1998)	7	2	Average
Liao & Chen (2005)	7	2	Average
Liao (2007)	7	2	Average
Liao et al.(2008)	7	2	Average
Roblyer (1988)	7	2	Average
Torgerson & Elbourne (2002)	7	2	Average
Bangert Drowns (1993)	8	2	Average
Fletcher-Flinn & Gravatt (1995)	8	2	Average
Khalili & Shashaani (1994)	8	2	Average
Koufogiannakis & Wiebe (2006)	8	2	Average
Waxman et al.(2003)	8	2	Average
Cohen & Dacanay (1992)	9	2	Average

Hsu (2003)	9	2	Average
Kulik & Kulik (1991)	9	2	Average
LeJeune (2002)	9	2	Average
Ryan (1991)	9	2	Average
Yaakub (1998)	9	2	Average
Bayraktar (2000)	10	3	Strong
Blok et al. (2002)	10	3	Strong
Kuchler (1998)	10	3	Strong
Michko (2007)	10	3	Strong
Soe et al. (2000)	10	3	Strong
Pearson et al. (2005)	11	3	Strong
Schenker (2007)	11	3	Strong
Goldberg et al. (2003)	12	3	Strong
Onuoha (2007)	12	3	Strong
Timmerman & Kruepke (2006)	12	3	Strong

For each of the included meta-analyses, the score for the comprehensiveness and rigour items was calculated and changed to a quality measure for future use in the analyses. Table 5 presents the scores for the comprehensiveness and rigour for all the included meta-analyses.

Identifying Unique Set of Meta-Analyses

After completing the effect size, standard error, and study feature extraction for all the 37 included meta-analyses, the uniqueness of the meta-analyses had to be resolved. A significant problem in meta-analysis is the interdependence issue when the same sample is used in multiple comparisons, either in the same study or across studies. An analogous problem at the second-order meta-analysis would be when the same studies are included in more than one meta-analysis (Sipe & Curlette, 1997).

This issue has been addressed differently in some of the previously published syntheses of meta-analyses. Wilson and Lipsey (2001) excluded one review from each pair of meta-analyses that had more than 25% overlap in primary research addressed, while making judgment calls, when the list of included studies was unavailable. On the other hand, Barrick, Mount, and Judge (2001) conducted two separate analyses, one with the set of meta-analyses that had no overlap in the studies included "independent analysis", and one with the full set of meta-analyses including those with substantial overlap in the studies they include "non-independent analysis". In a combination of both approaches, Sipe and Curlette (1997) considered meta-analyses as unique if they had no overlap or less than three studies in common, otherwise they were considered interdependent. The meta-analysis with the larger number of studies was included, and if both were close in number (not more than 10 studies apart) the more recently published was included. Finally, they conducted analyses for the complete set regardless of inter-dependence, and another set of analyses for the meta-analyses that they considered to be unique.

To address this problem, the first step taken in this second-order meta-analysis was to compile the overall set of primary studies included in the 37 different meta-analyses and specifying the single or different meta-analyses that each study appears in. This was compiled in a master excel file. The overall number of different primary studies that appeared in one or more meta-analyses was 1253.

While checking overlap between the studies covered in the different meta-analyses, articles published by same authors in consecutive years were examined to ensure that it was not a dissertation versus the published article. If that was the case, the two documents were considered to be the same to avoid unwanted overlap and dependency. Keeping in mind that the studies that were included in more than one meta-analysis did not clearly fall into pairs of meta-analyses, it was almost impossible to group the included meta-analyses into clear cut groups of overlapping reviews. Based on that, the decision was to calculate for each meta-analysis the number and frequency of studies that were included in another meta-analysis. Findings of this process are presented in Table 6 in addition to the categorical methodological index for each meta-analysis. For each meta-analysis, the number of studies column reflects the number of included studies, the number of common studies reflects the number of studies that also appeared in other meta-analyses, the percentage overlap reflects the percentage of studies that were

Table 5. Scores and quality for comprehensiveness and rigour aspects for the included meta-analyses

	Comprehensiveness		Rigour	
Reference	**Score**	**Quality**	**Score**	**Quality**
Christmann & Badgett (2000a)	3	Weak	2	Weak
Niemiec (1987)	3	Weak	2	Weak
Rosen & Salomon (2007)	3	Weak	2	Weak
Samson et al.(1986)	4	Weak	1	Weak
Christmann et al. (1997)	4	Weak	2	Weak
Lee (1999)	4	Weak	2	Weak
Liao (1992)	3	Weak	3	Weak
Zhao (2003)	3	Weak	3	Weak
Christmann & Badgett (2000b)	5	Strong	2	Weak
Christmann & Badgett (2003)	5	Strong	2	Weak
Liao (1998)	5	Strong	2	Weak
Liao & Chen (2005)	6	Strong	1	Weak
Liao (2007)	5	Strong	2	Weak
Liao, Cheng, & Chen (2008)	5	Strong	2	Weak
Roblyer (1988)	5	Strong	2	Weak
Torgerson & Elbourne (2002)	3	Weak	4	Weak
Bangert Drowns (1993)	4	Weak	4	Weak
Fletcher-Flinn & Gravatt (1995)	4	Weak	4	Weak
Khalili & Shashaani (1994)	4	Weak	4	Weak
Koufogiannakis & Wiebe (2006)	5	Strong	3	Weak
Waxman et al. (2003)	5	Strong	3	Weak
Cohen & Dacanay (1992)	5	Strong	4	Weak

Hsu (2003)	6	Strong	3	Weak
Kulik & Kulik (1991)	4	Weak	5	Strong
LeJeune (2002)	6	Strong	3	Weak
Ryan (1991)	5	Strong	4	Weak
Yaakub (1998)	6	Strong	3	Weak
Bayraktar (2000)	7	Strong	3	Weak
Blok et al. (2002)	7	Strong	3	Weak
Kuchler (1998)	7	Strong	3	Weak
Michko (2007)	6	Strong	4	Weak
Soe et al. (2000)	5	Strong	5	Strong
Pearson et al. (2005)	7	Strong	4	Weak
Schenker (2007)	7	Strong	4	Weak
Goldberg et al. (2003)	7	Strong	5	Strong
Onuoha (2007)	7	Strong	5	Strong
Timmerman & Kruepke (2006)	7	Strong	5	Strong

common with other meta-analyses, and the quality column reflects the overall methodological quality index.

At this point, it was easy to identify the meta-analysis that had a high percentage of overlap with other documents. The decision was to identify the set of meta-analyses that has the lowest level of overlap in primary studies while retaining the highest percentage of the overall set of primary studies. At the same time, attention was given to the methodological quality in order to retain the high quality meta-analyses based on the methodological quality index and not lose them from the final set of included meta-analyses.

As mentioned before, due to the fact that primary studies included in more than one meta-analysis were not always showing in two particular meta-analyses, the removal of one meta-analysis from the overall set resulted in a change in the frequencies of overlap in more than one meta-analysis. Therefore, the decision was to proceed with the exclusion of highly overlapping meta-analyses one at a time while retaining all meta-analyses that were rated as strong quality until a maximum frequency of 25% overlap was attained for each of the remaining meta-analyses. After the exclusion of any meta-analysis, the frequency of overlap for the remaining set of meta-analyses was calculated again and based on the new frequencies another highly overlapping weak quality or moderate quality meta-analysis was excluded. This process was repeated 12 times, during which the meta-analyses that are highlighted in grey in Table 6 were excluded and resulted in the final set of meta-analyses presented in Table 7.

Table 6. Number of primary studies and percentage of overlap in each of the included meta-analyses

Meta-analysis	# studies	# common studies	% of overlap	Quality
Christmann & Badgett (2000a)	16	4	25.0	Weak
Niemiec (1987)	48	27	56.2	Weak
Rosen & Salomon (2007)	31	12	38.7	Moderate
Samson et al.(1986)	41	18	43.9	Moderate
Christmann et al.(1997)	27	17	63.0	Moderate
Lee (1999)	19	11	57.9	Moderate
Liao (1992)	29	11	37.9	Moderate
Zhao (2003)	9	1	11.1	Moderate
Christmann & Badgett (2000b)	11	9	81.8	Moderate
Christmann & Badgett (2003)	39	18	46.1	Moderate
Liao (1998)	31	4	12.9	Strong
Liao & Chen (2005)	21	4	19.0	Moderate
Liao (2007)	52	19	36.5	Strong
Liao et al.(2008)	48	20	41.7	Moderate
Roblyer (1988)	35	9	25.7	Moderate
Torgerson & Elbourne (2002)	5	2	40.0	Moderate
Bangert Drowns (1993)	19	4	21.0	Moderate
Fletcher-Flinn & Gravatt (1995)	120	36	30.0	Moderate
Khalili & Shashaani (1994)	33	16	48.5	Moderate
Koufogiannakis & Wiebe (2006)	8	1	12.5	Strong
Waxman et al.(2003)	42	6	14.3	Moderate

Cohen & Dacanay (1992)	24	9	37.5	Moderate
Hsu (2003)	25	4	16.0	Moderate
Kulik & Kulik (1991)	239	43	18.0	Moderate
LeJeune (2002)	40	17	42.5	Moderate
Ryan (1991)	40	22	55.0	Moderate
Yaakub (1998)	20	4	20.0	Moderate
Bayraktar (2000)	42	10	23.8	Moderate
Blok et al.(2002)	25	2	8.0	Weak
Kuchler (1998)	65	16	24.6	Strong
Michko (2007)	45	0	0.0	Weak
Soe et al. (2000)	17	2	11.8	Strong
Pearson et al.(2005)	20	2	10.0	Strong
Schenker (2007)	46	11	23.9	Strong
Goldberg et al.(2003)	15	1	6.7	Strong
Onuoha (2007)	35	6	17.1	Strong
Timmerman & Kruepke (2006)	114	30	26.3	Strong

Note: that the greyed columns reflect the meta-analyses that were excluded due to high overlap

Table 7. Unique studies with minimal overlap with number of studies and percentage of overlap

Meta-analysis	# studies	# common studies	% of overlap	Quality
Christmann & Badgett (2000a)	16	4	25.0	Weak
Rosen & Salomon (2007)	31	0	0.0	Weak
Zhao (2003)	9	1	11.1	Moderate
Liao (1998)	31	2	6.4	Moderate
Liao & Chen (2005)	21	2	9.5	Moderate
Liao (2007)	52	2	3.8	Moderate
Roblyer (1988)	35	4	11.4	Moderate
Torgerson & Elbourne (2002)	5	0	0.0	Moderate
Bangert Drowns (1993)	19	1	5.3	Moderate
Fletcher-Flinn & Garavatt (1995)	120	26	21.7	Moderate
Koufogiannakis & Wiebe (2006)	8	1	12.5	Moderate
Waxman et al.(2003)	42	5	11.9	Moderate
Hsu (2003)	25	4	16.0	Moderate
Kulik & Kulik (1991)	239	8	3.3	Moderate
Yaakub (1998)	20	4	20.0	Moderate
Bayraktar (2000)	42	7	16.7	Strong
Blok et al.(2002)	25	2	8.0	Strong
Kuchler (1998)	65	7	10.8	Strong
Michko (2007)	45	0	0.0	Strong
Soe et al. (2000)	17	2	11.8	Strong
Pearson et al.(2005)	20	2	10.0	Strong

Schenker (2007)	46	9	19.6	Strong
Goldberg et al.(2003)	15	1	6.7	Strong
Onuoha (2007)	35	3	8.6	Strong
Timmerman & Kruepke (2006)	114	27	23.7	Strong

An observation that is worth noting is the fact that among the included meta-analyses there were three that addressed the impact of technology on students' achievement in Taiwan (Liao & Chen, 2005; Liao, Chang, & Chen, 2008; Liao, 2007). Although the overlap among the three led to the exclusion of one from the overall set of meta-analyses, none of the primary studies included in the three meta-analyses appeared in any of the other included meta-analyses. This comprised a set of 99 different primary studies conducted in Taiwan and not incorporated in any other meta-analysis.

For the purpose of the task at hand, and due to resource limitations, the following steps were completed by the principal investigator while conducting spot checks to ensure that no mistakes were done. The final number of meta-analyses that were considered to be unique or having acceptable levels of overlap was 25 with none having a frequency of overlapping studies beyond 25%. The overall number of primary studies included in this set was 1055 studies which represents 84.2% of the overall number of primary studies included in the overall set of meta-analyses.

Data for Validation Process

To allow for the validation of the findings of the second-order meta-analysis, the decision was to extract the raw data from the included meta-analyses in order to use them in the calculation of the point estimate which would help in the verification of that calculated through the synthesis of the meta-analyses. Individual effect sizes and sample sizes from the primary studies included in the various meta-analyses were extracted. Knowing that with the givens, particularly the absence of detailed information about each effect size and its source, the dependency issue cannot be totally resolved; the decision was to minimize it as much as possible. The raw data were only extracted from the set of meta-analyses with minimal overlap that were judged to be unique. In the cases where the overall sample size was provided it was assumed that the experimental and control groups were equal in size and in the case of an odd overall number of participants it was reduced by one. However, because these data were to be used for validation purposes, if sample sizes were not given by the authors for the individual effect sizes, no imputations were done.

From the 25 unique studies, 13 offered information allowing for the extraction of 574 individual effect sizes and their corresponding sample sizes, with the overall sample size being 60,853 participants, to be used in the validation process. However, seven meta-analyses offered individual effect sizes but provided no information about sample sizes, four did not give any individual effect sizes, and two provided tables with ranges of effects sizes rather than specifics.

Due to limited resources, the extraction was conducted by the principal investigator only and to avoid mistakes, random spot checks were done for the data entry into the excel file.

CHAPTER 4: ANALYSIS

In this chapter analyses are presented in different sections. The first section presents an overview of the 37 included meta-analyses followed by descriptive analyses regarding their contextual and methodological features. Next. an overview of the methodological quality index for the set of included meta-analyses is offered. After this, the effect size synthesis and moderator analyses are presented followed by the validation through the calculation of the effect sizes from the raw scores. Finally, specific effect sizes pertaining to type of technology, subject matter, and grade level that were extracted from the different included meta-analyses are presented.

Overview of Included Meta-analyses

In total, 38 independent effect sizes were extracted from 37 different meta-analyses involving 1253 different primary studies comparing student achievement in technology enhanced classroom instruction to traditional instruction. The 37 meta-analyses addressed a variety of technological approaches that were used in the experimental conditions to enhance and support face to face instruction. As for the control group it was the traditional or computer free setting in all the included meta-analyses. As presented in the inclusion/exclusion criteria, if the comparison group in a given meta-analysis incorporated the use of technology, it was excluded for the purpose of this second-order meta-analysis. Table 8 presents the list of included meta-analyses with the main research question for each.

Table 3. Included meta-analyses with the main research questions

Reference	Research Objective or Question
*Christmann & Badgett (2000a)	"Are there differences between the academic achievement levels of college-level students who were exposed to computer-assisted instruction, and those who were not exposed to this instruction?" (p. 94)
*Bangert Drowns (1993)	"Does word processing, especially in the context of writing instruction, facilitate the enhancement of writing skill?" (p. 69)
Cohen & Dacanay (1992)	"The purpose of the present study is to conduct a meta-analysis on the CBI literature in health professions education" (p.261)
*Kulik & Kulik (1991)	"Need to determine whether the record of effectiveness of CBI has changed with the development of new kinds of computers" (p. 77)
*Roblyer (1988)	"The purpose of this review is to provide as complete and up-to-date a picture as possible of trends in both research topics and findings, and of the contributions of instructional computing applications to educational effectiveness" (p.15)
*Waxman et al. (2003)	"The purpose of the present study is to synthesize recent research on the effects of teaching and learning with technology on student outcomes" (p. 6)
Ryan (1991)	" This meta-analysis was undertaken to clarify existing knowledge of the academic achievement effects of current microcomputer use in elementary schools" (p.162)
*Hsu (2003)	"How effective is the use of computer-assisted instruction (CAI) in enhancing the statistical learning of college students as compared with non-computer instructional techniques" (p.18)
Christmann & Badgett (2000b)	"What differences exist between the academic achievement levels of science students who are exposed to computer-assisted instruction, and those who were not exposed to this instruction within the academic areas of general science, biology, chemistry, and physics within three demographic settings...urban, rural, and suburban

	demographic setting" (p.119)
*Timmerman & Kruepke (2006)	"Addresses the overall effectiveness of CAI in relation to traditional instruction" (p.74)
*Pearson et al.(2005)	"… the impact of digital technology tools on middle-school students in the following areas: strategy use, meta-cognition, reading motivation, reading engagement, and reading comprehension" (p.6)
*Zhao (2003)	The main objective was "assessing the overall effectiveness of uses of technology in language education through meta-analysis" (p. 8)
*Torgerson & Elbourne (2002)	"The aim of the current paper is to undertake a systematic review of such robust evidence as it exists about the effectiveness of ICT in the teaching and learning of spelling in English language" (p.131)
*Soe et al. (2000)	"How effective is computer-assisted instruction in teaching students to read?" (p. 11)
Lee (1999)	"… analyze evidence concerning the effectiveness of two modes of simulation, presentation and practice" (p.72)
Christmann et al.(1997)	"what differences exist among academic achievement levels of students in eight subject areas who were exposed to computer-assisted instruction and those who were not exposed to this instruction" (p.283)
*Fletcher-Flinn & Gravatt (1995)	"The aim of this study was to provide an updated meta-analysis on the learning effect of (CAI) over a broad range of study features with particular attention focused on the effectiveness debate" (p.219)
Liao (1992)	" … investigate the effects of CAI on students' cognitive performance" (p.2)
Niemiec (1987)	"Seeks to gauge the overall effectiveness of CBI in the elementary school and to identify those groups of students who are most likely to benefit from exposure to CBI" (p.85)
*Yaakub (1998)	"What is the overall effectiveness of CAI on achievement as compared to traditional instruction?" (p. 1)

*Kuchler (1998)	"How effective is the use of computers in enhancing the mathematical achievement (i.e. the development of mathematical concepts, computational skills, and programming skills) of secondary school students as compared with non-computer instructional techniques?" (p.9)
LeJeune (2002)	"What are the differences between the use of computer-simulated experiments and traditional learning activities on student achievement outcomes relating to low level thinking skills …. and high level thinking skills" (p.6)
*Bayraktar (2000)	"Investigate the effectiveness of computer-assisted instruction (CAI) on student achievement in secondary and college science subject areas in the United States" (p.5)
*Goldberg et al.(2003)	"Does word processing impact k-12 student writing? If so, in what ways (i.e. is quality or quantity of student writing impacted)?" (p.5)
*Koufogiannakis & Wiebe (2006)	"Which library instruction methods are most effective for improving the information skills of students at an introductory, undergraduate level, using cognitive outcomes" (p.5) with one comparison between CAI and traditional instruction
Khalili & Shashaani (1994)	"Although the present research employs the same techniques and procedures used in the previous research to review studies on the effectiveness of computer applications in school, it reviews more recent studies" (p.3)
*Liao & Chen (2005)	"To synthesize existing research comparing the effects of computer simulation instruction (CSI) versus traditional instruction (TI) on students' achievement in Taiwan" (p.1)
Christmann & Badgett (2003)	"Concentrates on currently available research that has primarily examined the effectiveness of microcomputer-based software on the academic achievement of elementary school students" (p. 93)
Samson et al.(1986)	"summarize the available research relating to the effects of computer use on student achievement in junior and senior high school" (p.313)
*Blok (2002)	"The effectiveness of CAI programs for beginning reading instruction" (p. 107)

*Liao (1998)	"Synthesize existing research comparing effects of hypermedia and traditional instruction on students' achievement" (p.1)
*Rosen & Salomon (2007)	To test the differential effects of technology intensive constructivist oriented mathematics instruction on student achievement as compared to traditional ones.
*Michko (2007)	"To synthesize recent quantitative research on the effects of teaching and learning with technology on student outcomes in undergraduate engineering education" (p. 22)
*Schenker (2007)	"Examine effectiveness of using technology to enhance statistics instruction" (p.23)
*Onuoha (2007)	"Was computer based laboratory more effective in raising students' science academic achievements compared with the traditional hands-on laboratory in college and precollege level science education" (p.14"
*Liao et al.(2008)	"Synthesize existing research comparing the effects of computer applications (i.e. computer-assisted instruction, computer simulations, and web-based learning) versus traditional instruction on elementary students' achievement in Taiwan" (p. 43)
Liao (2007)	"Synthesize existing research comparing the effects of computer-assisted instruction (CAI) versus traditional instruction (TI) on students' achievement in Taiwan" (p.216)

* The studies marked with an asterisk are the unique studies with minimal overlap

For each of the meta-analyses, and as presented in the Codebook section, whenever provided by the authors, the definitions for both the experimental and the control group were extracted. In general, the control group received minimal attention from the authors. In 10 meta-analyses representing 27.1% of the included set, no explicit reference was given to the control group condition, with implicit indications that it was a traditional or computer free setting. Furthermore, 15 meta-analyses representing 40.5% of the included meta-analyses used the term "traditional instruction" to define the control group. Finally, 12 meta-analyses representing 32.4% of the included set specified that the control group was the computer free group, where the technology term used was that implicated with the experimental condition.

As for the experimental group, some of the meta-analyses offered a brief overview of the definition for their experimental group while others did not mention anything while still others provided a list of the included technologies. For this study feature it was very difficult to calculate frequencies of each since the differences between the different approaches are not clear cut and offering a label for each would not be adequately reliable. However, it is important to note that very few meta-analyses offered a detailed and clear description of the experimental group. Moreover, most of the meta-analyses that did offer definitions were particularly dissertations and not journal publications such as that completed by Schenker (2007) addressing effectiveness of technology use in teaching statistics at higher academic levels.

Descriptives

This section presents the descriptive analyses for the general study information in addition to contextual and methodological features pertaining to the 37 different included meta-analyses. For the purpose of these analyses, the *Statistical Package for Social Sciences (SPSS)* data analysis software was used.

General Study Information

Two of the general study information are reported, namely type and year of publication. Regarding type of publication, the meta-analyses included in this second-order meta-analysis where journal publications, dissertations, conference proceedings, or reports. The most frequent type was the journal publication where 24 of the included meta-analyses were of this type representing 64.9% of the included set of documents. The frequency distribution is presented in Table 9.

As for year of publication, the included meta-analyses were published between the years 1985 and 2008. The year 2003 witnessed the largest number of meta-analyses where five different ones were published. The frequency distribution is presented in Table 10.

Regarding the year of publication, when the years were grouped into five year time periods, the frequencies reflected an increasing trend in the number of meta-analyses published within each time frame. The frequency distribution within the five year time frames is presented in Table 11.

Table 9. Frequency distribution of type of publication

Type of Publication	Frequency	Relative %
Journal	24	64.9
Dissertation	8	21.6
Conference proceeding	1	2.7
Report	4	10.8
Total	37	100

Table 10. Frequency distribution of year of publication

Year of Publication	Frequency	Relative %
1986	1	2.7
1987	1	2.7
1988	1	2.7
1991	2	5.4
1992	2	5.4
1993	1	2.7
1994	1	2.7
1995	1	2.7
1997	1	2.7
1998	3	8.1
1999	1	2.7
2000	4	10.8
2002	3	8.1
2003	5	13.5
2005	2	5.4
2006	2	5.4
2007	5	13.5
2008	1	2.7
Total	37	100

Table 11. Frequency distribution of time frame of publication

Year of Publication	Frequency	Relative %
1985-1990	3	8.1
1991-1995	7	18.9
1996-2000	9	24.3
2001-2005	10	27.0
2006-present	8	21.6
Total	37	100

Contextual Features

Three contextual features were extracted, namely the technology addressed, the grade level and the subject matter. Regarding the technology addressed, the most frequently addressed approach was computer-assisted instruction with 17 out of the 37 included meta-analyses targeted computer-assisted instruction representing 45.9% of the overall set. The frequency distribution is presented in Table 12.

Table 12. Frequency distribution of technology addressed in the meta-analyses

Technology Addressed	Frequency	Relative %
CAI	17	45.9
CBI	5	13.5
CSI	1	2.7
Digital media	1	2.7
Educational technology	1	2.7
Hypermedia	1	2.7
ICT	2	5.4
Math program	1	2.7
Microcomputer	1	2.7
Simulations	3	8.1
Technology	2	5.4
Word processor	2	5.4
Total	37	100

As for grade level, the included meta-analyses focused on a specific category of grade levels (elementary, secondary, or post secondary), included more than one category, or were inclusive of all grade levels. The post-secondary category and the all inclusive were the most frequent with nine meta-analyses addressing. The frequency distribution is presented in Table 13.

Table 13. Frequency distribution of grade level addressed in the meta-analyses

Grade Level	Frequency	Relative %
Elementary	6	16.2
Secondary	5	13.5
Post-secondary	9	24.3
Elementary and Secondary	5	13.5
Secondary and Post-secondary	3	8.1
All inclusive	9	24.3
Total	37	100

Considering subject matter, most frequently the meta-analyses addressed a combination of subject matter areas with 19 meta-analyses representing 51.4% of the included meta-analyses addressing a combination of subjects. Nevertheless, the specific subject matter that received the highest attention among the included meta-analyses was language, followed by science and health. On the other hand, the subjects receiving the least attention were engineering, technology, and information literacy. The frequency distribution is presented in Table 14. For a full list of technologies addressed, grade levels included, and subject matter incorporated in each meta-analysis check Table 15.

Table 14. Frequency distribution of subject matter addressed in the meta-analyses

Subject matter	**Frequency**	**Relative %**
Science and Health	5	13.5
Language	6	16.2
Math	3	8.1
Technology	1	2.7
Combination	19	51.4
Information literacy	1	2.7
Engineering	2	5.4
Total	37	100

Methodological features

The frequencies for the methodological features are presented in subsections, with each one addressing methodological aspects pertaining to a specific phase in the implementation of a meta-analysis. The subsections include: a) search phase; b) review phase; c) effect size and study feature extraction phase; d) analysis phase; and e) further reporting issues.

Search phase

As presented earlier, the search phase represents one of the most important phases in a meta-analysis. From the 37 included meta-analyses, 26 specified the search time-frame used, representing 70.3% of all the included meta-analyses. From these 26 studies only 11 meta-analyses justified the used search time-frame, representing 42.3% of those studies that reported the time-frame from the first place.

Table 4. List of technologies addressed, grade levels included, and subject matter incorporated in each meta-analysis

Meta-analysis	Technology	Grade level	Subject matter	Quality
Christmann & Badgett (2000a)	CAI	3	6	Weak
Niemiec (1987)	CBI	1	6	Weak
Rosen & Salomon (2007)	Math program	4	4	Moderate
Samson et al.(1986)	CBI	2	6	Moderate
Christmann et al.(1997)	CAI	2	6	Moderate
Lee (1999)	Simulations	6	6	Moderate
Liao (1992)	CAI	6	8	Moderate
Zhao (2003)	ICT	3	2	Moderate
Christmann & Badgett (2000b)	CAI	2	1	Moderate
Christmann & Badgett (2003)	CAI	1	6	Moderate
Liao (1998)	Hypermedia	6	6	Strong
Liao & Chen (2005)	CSI	4	6	Moderate
Liao (2007)	CAI	6	6	Strong
Liao et al.(2008)	CAI	1	6	Moderate
Roblyer (1988)	CBI	6	6	Moderate
Torgerson & Elbourne (2002)	ICT	1	2	Moderate
Bangert Drowns (1993)	Word processor	6	6	Moderate
Fletcher-Flinn & Gravatt (1995)	CAI	3	6	Moderate
Khalili & Shashaani (1994)	Technology	6	6	Moderate

Koufogiannakis & Wiebe (2006)	CAI	3	7	Strong
Waxman et al.(2003)	Technology	4	6	Moderate
Cohen & Dacanay (1992)	CBI	3	1	Moderate
Hsu (2003)	CAI	3	3	Moderate
Kulik & Kulik (1991)	CBI	6	6	Moderate
LeJeune (2002)	Simulations	6	1	Moderate
Ryan (1991)	Microcomputer	1	6	Moderate
Yaakub (1998)	CAI	5	6	Moderate
Bayraktar (2000)	CAI	5	1	Moderate
Blok et al.(2002)	CAI	1	2	Weak
Kuchler (1998)	CAI	2	3	Strong
Michko (2007)	Educational technology	3	8	Weak
Soe et al. (2000)	CAI	4	2	Strong
Pearson et al.(2005)	Digital media	2	2	Strong
Schenker (2007)	CAI	3	3	Strong
Goldberg et al.(2003)	Word processor	4	2	Strong
Onuoha (2007)	Simulations	5	1	Strong
Timmerman & Kruepke (2006)	CAI	3	6	Strong

For subject matter: 1= Science and health; 2= Language; 3= Math; 4= Technology; 5= Social science; 6= Combination; 7= Information literacy; 8= Engineering.

For grade level: 1= Elementary; 2= Secondary; 3= Post-secondary; 4= Elementary & secondary; 5= Secondary & post-secondary; 6= inclusive of all

As for the literature included, 31 meta-analyses representing 83.8% of the included documents addressed both published and unpublished studies while six meta-analyses representing 16.2% of the included documents addressed only published primary studies.

Considering the search strategy, the majority of the meta-analyses (60.5%) listed the resources and databases searched while offering sample search terms. The frequency distribution of characteristics of the search strategy reporting by the different meta-analyses is presented in Table 16.

Table 16. Frequency distribution of characteristics of the search strategy reporting

Search Strategy	**Frequency**	**Relative %**
Not disclosed, no reference to search strategy	2	5.4
Minimal description with brief reference to resources searched	2	5.4
Listing of resources and databases searched	11	29.7
Listing of resources and databases with sample search terms	22	59.5
Total	37	100

Considering the different search venues used in the included meta-analyses, and keeping in mind that most meta-analyses used more than one search venue, the most frequently used were database searches with 86.5% of the included meta-analyses utilizing them. This was followed by branching where 64.9% of the included meta-analyses used it. The frequency distribution of the different search venues in the different meta-analyses is presented in Table 17.

Table 17. Frequency distribution of search venues

Search venue	**Frequency**	**Relative %**
Database searches	32	86.5
Computerized searches	4	10.8
Hand search	14	37.8
Branching	24	64.9

Among all the included meta-analyses, the number of databases searched ranged between one and nineteen, with two being the most frequent where 11 meta-analyses, representing 29.7% of the meta-analyses, used two databases; followed by three where nine meta-analyses, representing 24.3% of the meta-analyses, used three databases.

Review phase

The review phase includes the criteria used and process implemented by each meta-analysis in order to decide on which primary studies to include. In general, the majority of the meta-analyses offered an overview of the criteria used as reflected by the 29 meta-analyses that were coded as such. The frequency distribution of the different levels of clarity in the report regarding inclusion and exclusion criteria is presented in Table 18.

Table 18. Frequency distribution of inclusion/exclusion criteria

Inclusion/Exclusion Criteria	Frequency	Relative %
Criteria not disclosed with no description offered	0	0.0
Overview of criteria presented briefly	29	78.4
Criteria specified in detail allowing for easy replication	8	21.6
Total	37	100

The next aspect of the inclusion phase was the research types incorporated in each meta-analysis. The highest frequency was for the all inclusive approach that included randomized control trials, quasi-experimental, and experimental designs. The frequency distribution of the included research types is presented in Table 19.

Table 19. Frequency distribution of included research types

Included Research Types	Frequency	Relative %
RCT	1	2.7
RCT/Quasi	5	16.7
RCT/Quasi/pre	24	64.9
Missing	7	18.9
Total	37	100

As for the review process itself, 31 meta-analyses representing 83.8% of the included set did not refer to it at all. The frequency distribution of the methodological feature addressing the article review process is presented in Table 20.

Table 20. Frequency distribution of article review process

Article Review	Frequency	Relative %
Review process not disclosed	31	83.8
Review process by one researcher	4	10.8
Rating by more than one researcher	2	5.4
Rating by more than one researcher with inter-rater agreement provided	0	0.0
Total	37	100

Moreover, the number of years included in the meta-analyses ranged between 4 and 36, while number of studies ranged between 5 and 248.

Effect Size and Study Feature Extraction Phase

This phase includes three different features the effect size extraction, the codebook, and the study feature extraction. For the effect size extraction the majority of the included meta-analyses, 31 meta-analyses representing 83.8% of included meta-analyses, did not disclose any information about how it was conducted. The frequency distribution of the methodological feature addressing the effect size extraction process is presented in Table 21.

Table 21. Frequency distribution of effect size extraction process

Effect Size Extraction Process	Frequency	Relative %
Not disclosed, no reference to how it was conducted	31	83.8
By one researcher	3	8.1
By more than one researcher	3	8.1
By more than one researcher with inter-rater agreement provided	0	0.0
Total	37	100

As for the codebook, the larger percentage of meta-analysts, 15 meta-analyses representing 40.5% of the included meta-analyses, listed the specific categories addressed in their reviews. The frequency distribution of the methodological feature addressing the clarity of the codebook is presented in Table 22.

Table 22. Frequency distribution of codebook

Codebook	Frequency	Relative %
Not described, no reference to extracted features	7	18.9
Brief description of main categories in codebook	9	24.3
Listing of specific categories addressed in codebook	15	40.1
Elaborate description of codebook allowing for easy replication	6	16.2
Total	37	100

Finally within this subsection, the majority of the meta-analyses, 16 out of the included 37 representing 43.2% of the overall collection, did not refer to the study feature extraction process. The frequency distribution of the study feature extraction process is presented in Table 23.

Table 23. Frequency distribution of the study feature extraction process

Study Feature Extraction Process	Frequency	Relative %
Not disclosed, no reference to how it was conducted	16	43.2
By one researcher	3	8.1
By more than one researcher	11	29.7
By more than one researcher with inter-rater agreement provided	7	18.9
Total	37	100

Analysis Phase

Concerning the analytical aspects within each meta-analysis, the study features extracted included the independence of data, weighting effect sizes by sample size, homogeneity of variance, moderator analysis, and meta-regression analysis. The numbers and frequencies of meta-analyses that applied each of the listed approaches are presented in Table 24.

Table 24. Frequency distribution of studies implementing different analytical approaches

Analysis	Frequency	Relative %
Independence of Data	18	48.6
Effect sizes weighted by sample size	15	40.5
Homogeneity of variance conducted	12	32.4
Moderator Analysis conducted	14	37.8
Meta-regression Analysis conducted	2	5.4

Moreover, the type of effect size reported in each meta-analysis was also coded for. The most frequent type was Glass's Δ, followed by Hedges g. The frequency distribution of the type of effect size is presented in Table 25.

Table 25. Frequency distribution of the type of effect Size

Effect size type	Frequency	Relative %
Glass	16	43.2
Cohen	4	10.8
Hedges	11	29.7
Missing	6	16.2
Total	37	100

Concerning the extraction of standard error for the second-order meta-analysis, and as presented in the methodology section, there were four different situations in the included meta-analyses: a) reported in the meta-analysis; b) calculated from effect sizes and sample sizes; c) calculated from confidence intervals; and d) missing then imputed. The numbers and frequencies of meta-analyses to which each of the listed situations apply are presented in Table 26.

Table 26. Frequency distribution of standard error calculation process

Standard error	Frequency	Relative %
Reported	8	21.6
Calculated (effect size and sample size)	12	32.4
Calculated (Confidence interval)	3	8.1
Missing- imputed	14	37.8
Total	37	100

Finally, as noted in the codebook section, reviewing the included meta-analyses revealed a controversial method that was used by a number of meta-analysts to overcome the predicament of studies having higher weights due to extracting effect sizes from multiple non-independent comparisons from the same study. This was applied in seven out of the 37 studies representing 18.9% of all the included meta-analyses. The meta-analyses that applied this approach are: Samson, Niemiec, Weinstein, and Walberg (1986), Niemiec (1987), Liao (1998), Waxman, Lin, and Michko (2003), Liao and Chen (2005), Liao (2007), and Liao, Chang and Chen (2008). It is also important to note that in all of these meta-analyses there was no mention of any reference that supports such an approach of weighting which only augments the problem.

Further Reporting Issues

Beyond the methodological quality features, the codebook included a few study features that address particular reporting issues. Regarding the inclusion of a table summarizing the individual effect sizes, the coding procedure revealed that 31 studies provided such a table representing 83.8% of the included meta-analyses. As for the number of participants addressed in a given meta-analysis, 10 meta-analyses representing 27.0% of the included meta-analyses gave the overall number of participants while 13 meta-analyses representing 35.1% of the included meta-analyses gave information allowing for the calculation of the overall number of participants. Unfortunately, 14 meta-analyses representing 37.8% of the included set did not offer any information about the sample size included in their analysis.

As for the time difference between the last included primary study and the publication date of a given meta-analysis, it ranged between zero and five years. The most frequent time period was two years with 11 meta-analyses representing 29.7% of the included studies reflecting such a situation. However, it was strange that 12 meta-analyses representing 32.4% of the included set had a time frame of four years or more between the last included study and the publication date. The frequency distribution of the different time periods is presented in Table 27.

Table 27. Frequency distribution of time period in years between last included study and publication date

Difference between last study and publication date	Frequency	Relative %
0	2	5.4
1	6	16.2
2	11	29.7
3	6	16.2
4	6	16.2
5	6	16.2
Total	37	100

Methodological Quality Index

As presented earlier, a methodological quality index was calculated for each included meta-analysis based on a set of 14 study features that reflect different methodological aspects of the meta-analyses. The specific codes for each meta-analysis for the different items are presented in Table 28. Moreover, the table presents the overall methodological quality score for each meta-analysis along with the categorical one.

The total score was meant to reflect the number of methodological quality aspects addressed by a given meta-analysis, and was calculated by counting the number of twos for each meta-analysis. As presented in the section addressing the methodological quality index, the score for each meta-analysis was categorized into a three level methodological quality score as follows:

- A meta-analysis scoring between 1 and 5 is rated as a weak review
- A meta-analysis scoring between 6 and 9 is rated as an average review
- A meta-analysis scoring between 10 and 14 is rated as a strong quality review

Table 5. Specific codes for each meta-analysis on the different items included in the methodological quality index

Reference	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total	Category
Christmann & Badgett (2000a)	2	1	99	99	2	2	1	1	1	1	1	1	2	2	**5**	**1**
Niemiec (1987)	2	2	1	1	2	2	1	1	2	1	1	1	1	1	**5**	**1**
Rosen & Salomon (2007)	1	2	1	1	2	2	1	1	1	2	99	1	1	2	**5**	**1**
Samson (1986)	2	2	1	1	2	2	1	1	2	1	1	1	1	1	**5**	**1**
Christmann et al. (1997)	2	1	99	99	2	2	1	1	1	1	1	2	2	2	**6**	**2**
Lee (1999)	2	2	1	1	2	2	1	1	1	1	1	1	2	2	**6**	**2**
Liao (1992)	2	2	1	1	2	99	1	1	1	2	2	1	1	2	**6**	**2**
Zhao (2003)	1	2	1	1	2	99	1	1	1	1	2	2	2	2	**6**	**2**
Christmann & Badgett (2000b)	2	2	1	2	2	2	1	1	1	1	1	1	2	2	**7**	**2**
Christmann & Badgett (2003)	2	2	1	2	2	2	1	1	1	99	1	1	2	2	**7**	**2**
Liao (1998)	2	2	1	1	2	2	1	1	2	2	1	2	1	1	**7**	**2**
Liao & Chen (2005)	2	2	1	2	2	2	1	1	1	2	1	2	1	1	**7**	**2**
Liao (2007)	2	2	1	2	2	2	1	1	2	2	1	1	1	1	**7**	**2**
Liao, Cheng, & Chen (2008)	2	2	1	2	2	2	1	1	2	2	1	1	1	1	**7**	**2**
Roblyer (1988)	2	2	2	1	2	99	1	1	1	1	2	2	1	2	**7**	**2**
Torgerson & Elbourne (2002)	1	2	1	2	2	1	1	2	1	1	2	1	2	2	**7**	**2**
Bangert Drowns (1993)	2	2	1	1	2	99	1	1	2	1	2	2	2	2	**8**	**2**
Fletcher-Flinn & Gravatt (1995)	2	1	1	1	2	2	1	1	2	2	2	2	1	2	**8**	**2**
Khalili & Shashaani (1994)	1	2	1	1	2	2	1	1	2	2	2	2	1	2	**8**	**2**
Koufogiannakis & Wiebe (2006)	2	2	1	2	2	99	1	1	1	1	2	2	2	2	**8**	**2**

Waxman et al. (2003)	1	2	2	1	2	2	1	1	2	2	1	2	2	1	**8**	**2**
Cohen & Dacanay (1992)	2	1	2	2	2	99	1	1	1	2	2	2	2	2	**9**	**2**
Hsu (2003)	2	2	2	2	2	99	1	1	2	1	1	2	2	2	**9**	**2**
Kulik & Kulik (1991)	2	2	1	1	2	2	1	1	2	2	2	1	2	2	**9**	**2**
LeJeune (2002)	2	2	1	2	2	2	1	1	1	1	2	2	2	2	**9**	**2**
Ryan (1991)	2	2	1	1	2	2	1	1	2	1	2	2	2	2	**9**	**2**
Yaakub (1998)	2	2	1	2	2	2	1	1	1	2	2	2	1	2	**9**	**2**
Bayraktar (2000)	2	2	2	2	2	2	1	1	2	2	1	2	1	2	**10**	**3**
Blok et al. (2002)	2	2	2	2	2	2	1	1	2	1	1	2	2	2	**10**	**3**
Kuchler (1998)	2	2	2	2	2	2	1	1	2	1	2	2	1	2	**10**	**3**
Michko (2007)	1	2	2	2	2	2	1	2	2	1	1	2	2	2	**10**	**3**
Soe, Koki, & Chang (2000)	2	2	1	1	2	2	1	1	2	2	2	2	2	2	**10**	**3**
Pearson et al. (2005)	2	2	2	2	2	2	2	1	2	1	1	2	2	2	**11**	**3**
Schenker (2007)	2	2	2	2	2	2	1	1	2	2	1	2	2	2	**11**	**3**
Goldberg et al. (2003)	2	2	2	2	2	2	1	1	2	2	2	2	2	2	**12**	**3**
Onuoha (2007)	2	2	2	2	2	2	1	1	2	2	2	2	2	2	**12**	**3**
Timmerman & Kruepke (2006)	2	2	2	2	2	2	1	2	1	2	2	2	2	2	**12**	**3**

In the title row: 1= Literature covered; 2= Search strategy; 3= Resources used; 4= Number of databases searched; 5= Inclusion/exclusion criteria; 6= Included research; 7= Article review; 8= Effect size extraction; 9= Codebook; 10= Study feature extraction; 11= Independence of data; 12= Time between last included study and publication date; 13= Standard error calculation; 14= Weighting by number of comparisons
Within the cells, 1 reflects low methodological quality, 2 reflects high methodological quality, and 99 refers to missing information

Considering the overall set of meta-analyses, the minimum score calculated for the methodological quality was 5 and the maximum was 12, with the median being 8.00 and the average score being 8.16 with a standard deviation of 2.0. The most frequent score was seven with eight different meta-analyses attaining it representing 21.6% of the overall set of included meta-analyses. The frequency distribution of the methodological quality index is presented in Table 29.

Table 29. Frequency distribution of methodological quality index

Score	Frequency	Relative %
5	4	10.8
6	4	10.8
7	8	21.6
8	5	13.5
9	6	16.2
10	5	13.5
11	2	5.4
12	3	8.1
Total	37	100

As for the frequencies of the categorical evaluation, the majority of the meta-analyses were of moderate quality with 23 meta-analyses representing 62.2% of the included meta-analyses scoring within the average methodological quality range. The frequency distribution of the categorical methodological quality index is presented in Table 30.

As noted in the section addressing the design of the methodological quality index, two sub scores were calculated, one addressing the comprehensiveness of a meta-analysis and one addressing its rigour.

Table 30. Frequency distribution of categorical methodological quality index

Categorical Quality	Frequency	Relative %
Weak	4	10.8
Moderate	23	62.2
Strong	10	27.0
Total	37	100

For comprehensiveness, the minimum score calculated was 3 and the maximum was 7, with the median being 5 and the average score being 5.08 with a standard deviation of 1.34 The most frequent score was 5 with 11 different meta-analyses attaining it representing 29.7% of the overall set of included meta-analyses. The frequency distribution of the comprehensiveness quality score is presented in Table 31. When the scores were changed into the categorical format, 13 studies representing 35.1% of the included meta-analyses were deemed to be weak and 24 studies representing 64.9% of the included meta-analyses were deemed to be strong.

Table 31. Frequency distribution of comprehensiveness quality score

Comprehensiveness Quality Score	Frequency	Relative %
3	5	13.5
4	8	21.6
5	11	29.7
6	5	13.5
7	8	21.6
Total	37	100

For the rigour aspect, the minimum score calculated was 1 and the maximum was 5, with the median being 3 and the average score being 3.08 with a standard deviation of 1.19. The most frequent scores were 2 and 3 were each had 10 different meta-analyses attaining it representing 27.0% of the overall set of included meta-analyses each. The frequency distribution of the methodological quality score targeting the rigour is presented in Table 32. When the scores were changed into the categorical format, 32 studies representing 86.5% of the included meta-analyses were deemed to be weak and 5 studies representing 13.5% of the included meta-analyses were deemed to be strong.

Table 32. Frequency distribution of the rigour quality score

Rigour Quality Score	Frequency	Relative %
1	3	8.1
2	10	27.0
3	10	27.0
4	9	24.3
5	5	13.5
Total	37	100

In order to examine the relationship between time of publication and overall methodological quality of the included meta-analyses, a Pearson Product Moment correlation was conducted using the continuous methodological quality index. Results revealed a significant positive correlation of moderate strength ($r=.35$, $p<0.05$) between publication date and the overall methodological index indicating that with time, overall methodological quality seems to be improving.

Furthermore, to investigate whether both aspects of quality are correlated with date of publication, the specific scores for the comprehensiveness and the rigour were correlated with publication date. Results of the Pearson Product Moment correlation revealed a significant positive correlation of moderate strength ($r=.47$, $p<0.01$) between publication date and comprehensiveness score, and a non-significant positive correlation of week strength ($r=.14$, $p>0.05$) between publication date and methodological score addressing rigour. The results give an indication that with time, the comprehensiveness quality of meta-analyses seems to be improving but not the rigour aspects.

Effect Size Synthesis

For the purpose of the effect size synthesis, the effect size, standard error, methodological quality indexes, and scores for the extracted study features for each of the 37 different meta-analyses were entered into *Analysis*™ 2.0 (Borenstein, Hedges, Higgins, & Rothstein, 2005).

A list of the effect size type, the effect size numerical value, the sample-size standard error (standard error based on the sample sizes corresponding to the individual effect sizes in the included meta-analyses), and the number-of-studies standard error (standard error based on the number of studies included in each meta-analysis) for each of the included meta-analyses is presented in Table 33. Figure 1 presents the forest plot for the overall set of effect sizes when sample-size standard error was used, while Figure 2 presents the forest plot for the overall set of effect sizes when number-of-studies standard error was used.

Table 6. List of the of effect size value and type with the standard errors

Reference	Mean ES	ES type	Sample size SE	Meta-analysis SE
* Christmann & Badgett (2000b)	0.127	Missing	0.045	0.343
* Bangert Drowns (1993)	0.270	Missing	0.110	0.325
Cohen and Dacanay (1992)	0.410	Glass Δ	0.120	0.234
* Kulik & Kulik (1991)	0.300	Glass Δ	0.029	0.090
* Roblyer et al. (1988)	0.310	Hedges *g*	0.051	0.163
* Waxman et al. (2003)	0.448	Glass Δ	0.141	0.264
Ryan (1991)	0.309	Glass & Hedges *g*	0.020	0.224
* Hsu (2003)	0.430	Hedges *g*	0.033	0.285
Christmann & Badgett (2000)	0.266	Glass Δ	0.420	0.428
* Timmerman & Kruepke (2006)	0.242	R	0.029	0.130
* Pearson et al. (2005)	0.489	Hedges *g*	0.112	0.319
* Zhao (2003)	1.120	Hedges *g*	0.260	0.503
* Torgerson & Elbourne (2002)	0.370	Cohen *d*	0.160	0.638
* Soe et al. (2000)	0.264	Hedges *g* & R	0.051	0.344
Lee (1999)	0.410	Glass Δ	0.023	0.327
Christmann et al. (1997)	0.209	Glass Δ	0.027	0.273
* Fletcher Flinn & Gravatt (1995)	0.240	Glass Δ	0.051	0.129
Liao (1992)	0.480	Glass Δ	0.051	0.256
Niemiec (1987)	0.450	Glass Δ	0.051	0.205
* Yaakub (1998)	0.350	Glass Δ & Hedges *g*	0.051	0.310

* Kuchler (1998)	0.440	Hedges *g*	0.051	0.176
LeJeune (2002)	0.340	Hedges *g*	0.042	0.318
LeJeune (2002)	0.380	Hedges *g*	0.035	0.303
* Bayraktar (2000)	0.273	Cohen *d*	0.051	0.219
* Goldberg et al. (2003)	0.410	Hedges *g*	0.070	0.368
* Koufogiannakis & Wiebe (2006)	-0.090	Hedges *g*	0.194	0.500
Khalili & Shashaani (1994)	0.380	Glass Δ	0.051	0.237
* Liao (2005)	0.517	Glass Δ	0.051	0.304
Christmann & Badgett (2003)	0.342	Glass Δ	0.022	0.227
Samson et al. (1986)	0.320	Glass Δ	0.051	0.216
* Blok et al. (2002)	0.254	Hedges *g*	0.056	0.283
* Liao (1998)	0.480	Glass Δ	0.051	0.241
* Rosen & Salomon (2007)	0.460	Hedges *g*	0.051	0.252
* Michko (2007)	0.433	Hedges *g*	0.067	0.212
* Schenker (2007)	0.239	Cohen *d*	0.020	0.209
* Onuoha (2007)	0.260	Cohen *d*	0.035	0.240
Liao (2007)	0.449	Glass Δ	0.051	0.205
* Liao et al. (2008)	0.552	Glass Δ	0.051	0.198

Figure 1. Forest plot for the overall set of 38 effect sizes when sample-size standard error was used

Study name	Statistics for each study						
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
1.000	0.127	0.045	0.002	0.039	0.215	2.822	0.005
2.000	0.270	0.110	0.012	0.054	0.486	2.455	0.014
3.000	0.410	0.120	0.014	0.175	0.645	3.417	0.001
4.000	0.300	0.029	0.001	0.243	0.357	10.345	0.000
5.000	0.310	0.051	0.003	0.210	0.410	6.078	0.000
6.000	0.448	0.141	0.020	0.172	0.724	3.177	0.001
7.000	0.309	0.020	0.000	0.270	0.348	15.450	0.000
8.000	0.430	0.033	0.001	0.365	0.495	13.030	0.000
9.000	0.266	0.420	0.176	-0.557	1.089	0.633	0.527
10.000	0.242	0.029	0.001	0.185	0.299	8.345	0.000
11.000	0.489	0.112	0.013	0.269	0.709	4.366	0.000
12.000	1.120	0.260	0.068	0.610	1.630	4.308	0.000
13.000	0.370	0.160	0.026	0.056	0.684	2.313	0.021
14.000	0.264	0.051	0.003	0.164	0.364	5.176	0.000
15.000	0.410	0.023	0.001	0.365	0.455	17.826	0.000
16.000	0.209	0.027	0.001	0.156	0.262	7.741	0.000
17.000	0.240	0.051	0.003	0.140	0.340	4.706	0.000
18.000	0.480	0.051	0.003	0.380	0.580	9.412	0.000
19.000	0.450	0.051	0.003	0.350	0.550	8.824	0.000
20.000	0.350	0.051	0.003	0.250	0.450	6.863	0.000
21.000	0.440	0.051	0.003	0.340	0.540	8.627	0.000
22.000	0.340	0.042	0.002	0.258	0.422	8.095	0.000
23.000	0.380	0.035	0.001	0.311	0.449	10.857	0.000
24.000	0.273	0.051	0.003	0.173	0.373	5.353	0.000
25.000	0.410	0.070	0.005	0.273	0.547	5.857	0.000
26.000	-0.090	0.194	0.038	-0.470	0.290	-0.464	0.643
27.000	0.380	0.051	0.003	0.280	0.480	7.451	0.000
28.000	0.517	0.051	0.003	0.417	0.617	10.137	0.000
29.000	0.342	0.022	0.000	0.299	0.385	15.545	0.000
30.000	0.320	0.051	0.003	0.220	0.420	6.275	0.000
31.000	0.254	0.056	0.003	0.144	0.364	4.536	0.000
32.000	0.480	0.051	0.003	0.380	0.580	9.412	0.000
33.000	0.460	0.051	0.003	0.360	0.560	9.020	0.000
34.000	0.433	0.067	0.004	0.302	0.564	6.463	0.000
35.000	0.239	0.020	0.000	0.200	0.278	11.950	0.000
36.000	0.260	0.035	0.001	0.191	0.329	7.429	0.000
37.000	0.449	0.051	0.003	0.349	0.549	8.804	0.000
38.000	0.552	0.051	0.003	0.452	0.652	10.818	0.000
	0.330	0.007	0.000	0.317	0.343	50.242	0.000

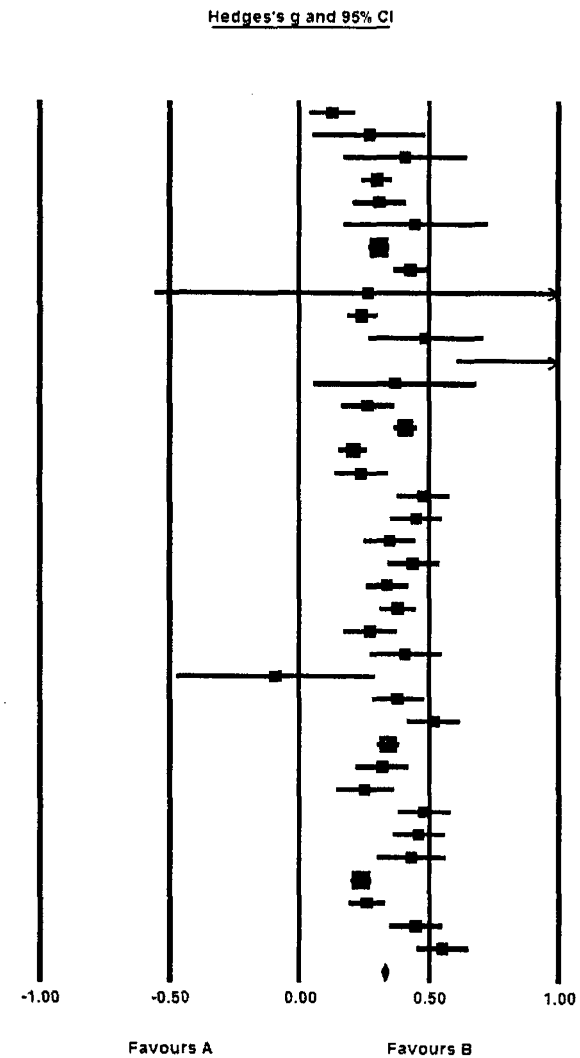
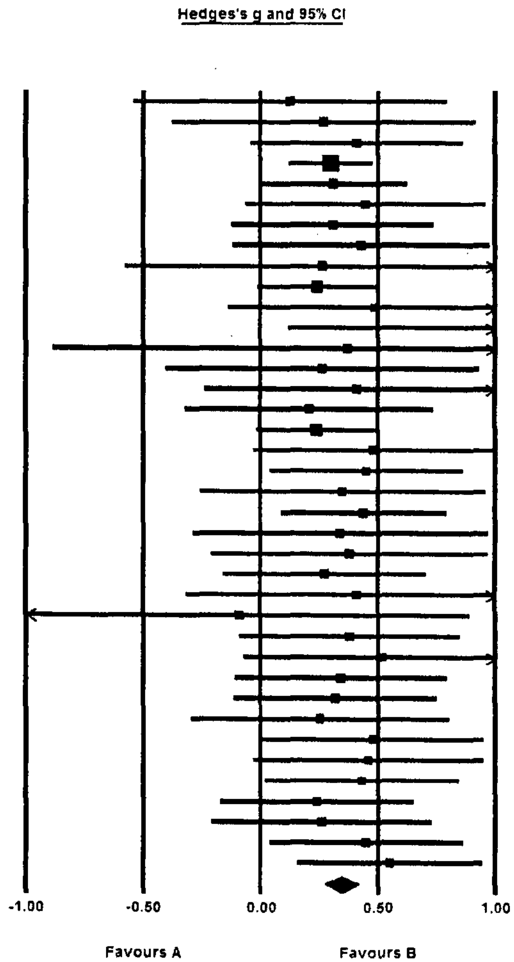


Figure 2. Forest plot for the overall set of 38 effect sizes when number-of-studies standard error was used

Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
1.000	0.127	0.340	0.116	-0.539	0.793	0.374	0.709
2.000	0.270	0.330	0.109	-0.377	0.917	0.818	0.413
3.000	0.410	0.230	0.053	-0.041	0.861	1.783	0.075
4.000	0.300	0.090	0.008	0.124	0.476	3.333	0.001
5.000	0.310	0.160	0.026	-0.004	0.624	1.938	0.053
6.000	0.448	0.260	0.068	-0.062	0.958	1.723	0.085
7.000	0.309	0.220	0.048	-0.122	0.740	1.405	0.160
8.000	0.430	0.280	0.078	-0.119	0.979	1.536	0.125
9.000	0.266	0.430	0.185	-0.577	1.109	0.619	0.536
10.000	0.242	0.130	0.017	-0.013	0.497	1.862	0.063
11.000	0.489	0.320	0.102	-0.138	1.116	1.528	0.126
12.000	1.120	0.510	0.260	0.120	2.120	2.196	0.028
13.000	0.370	0.640	0.410	-0.884	1.624	0.578	0.563
14.000	0.264	0.340	0.116	-0.402	0.930	0.776	0.437
15.000	0.410	0.330	0.109	-0.237	1.057	1.242	0.214
16.000	0.209	0.270	0.073	-0.320	0.738	0.774	0.439
17.000	0.240	0.130	0.017	-0.015	0.495	1.846	0.065
18.000	0.480	0.260	0.068	-0.030	0.990	1.846	0.065
19.000	0.450	0.210	0.044	0.038	0.862	2.143	0.032
20.000	0.350	0.310	0.096	-0.258	0.958	1.129	0.259
21.000	0.440	0.180	0.032	0.087	0.793	2.444	0.015
22.000	0.340	0.320	0.102	-0.287	0.967	1.063	0.288
23.000	0.380	0.300	0.090	-0.208	0.968	1.267	0.205
24.000	0.273	0.220	0.048	-0.158	0.704	1.241	0.215
25.000	0.410	0.370	0.137	-0.315	1.135	1.108	0.268
26.000	-0.090	0.500	0.250	-1.070	0.890	-0.180	0.857
27.000	0.380	0.240	0.058	-0.090	0.850	1.583	0.113
28.000	0.517	0.300	0.090	-0.071	1.105	1.723	0.085
29.000	0.342	0.230	0.053	-0.109	0.793	1.487	0.137
30.000	0.320	0.220	0.048	-0.111	0.751	1.455	0.146
31.000	0.254	0.280	0.078	-0.295	0.803	0.907	0.364
32.000	0.480	0.240	0.058	0.010	0.950	2.000	0.046
33.000	0.460	0.250	0.063	-0.030	0.950	1.840	0.066
34.000	0.433	0.210	0.044	0.021	0.845	2.062	0.039
35.000	0.239	0.210	0.044	-0.173	0.651	1.138	0.255
36.000	0.260	0.240	0.058	-0.210	0.730	1.083	0.279
37.000	0.449	0.210	0.044	0.037	0.861	2.138	0.033
38.000	0.552	0.200	0.040	0.160	0.944	2.759	0.006
	0.344	0.036	0.001	0.273	0.414	9.564	0.000



Outlier Analysis and Publication Bias

With both the overall and unique sets of meta-analyses, and with both sample size and number-of-studies standard error approaches, outlier analysis through the "One study removed" approach revealed that effect sizes fell within the 95th confidence interval of the average effect size. Therefore, with all the approaches, all the effect sizes were considered to fall within an acceptable range around the average effect and there was no need to exclude any. Figures 3, 4, 5, and 6 present the full "One study removed" analysis for each of the two sets of meta-analyses and with each of the two standard error approaches.

As for the standard error by Hedges' *g* funnel plots for the effect sizes, for both sets of meta-analyses, and with both standard error approaches, they revealed almost symmetrical distributions around the mean effect size in each case with no need for imputations indicating the absence of any obvious publication bias. Figures 7, 8, 9, and 10 present the funnel plots for each of the two sets of meta-analyses and with each of the two standard error approaches.

Figure 3. One study removed for overall set of 38 effect sizes with sample-size standard error approach

Study name	Point	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
	Statistics with study removed						
1.000	0.334	0.007	0.000	0.321	0.347	50.369	0.000
2.000	0.330	0.007	0.000	0.317	0.343	50.185	0.000
3.000	0.330	0.007	0.000	0.317	0.343	50.130	0.000
4.000	0.331	0.007	0.000	0.318	0.345	49.176	0.000
5.000	0.330	0.007	0.000	0.317	0.343	49.874	0.000
6.000	0.330	0.007	0.000	0.317	0.342	50.148	0.000
7.000	0.332	0.007	0.000	0.319	0.346	47.820	0.000
8.000	0.326	0.007	0.000	0.313	0.339	48.621	0.000
9.000	0.330	0.007	0.000	0.317	0.343	50.238	0.000
10.000	0.335	0.007	0.000	0.321	0.348	49.641	0.000
11.000	0.329	0.007	0.000	0.316	0.342	50.072	0.000
12.000	0.329	0.007	0.000	0.316	0.342	50.149	0.000
13.000	0.330	0.007	0.000	0.317	0.343	50.189	0.000
14.000	0.331	0.007	0.000	0.318	0.344	49.991	0.000
15.000	0.323	0.007	0.000	0.309	0.336	47.113	0.000
16.000	0.337	0.007	0.000	0.324	0.351	49.856	0.000
17.000	0.331	0.007	0.000	0.318	0.344	50.052	0.000
18.000	0.327	0.007	0.000	0.314	0.340	49.441	0.000
19.000	0.328	0.007	0.000	0.315	0.341	49.518	0.000
20.000	0.330	0.007	0.000	0.317	0.342	49.772	0.000
21.000	0.328	0.007	0.000	0.315	0.341	49.543	0.000
22.000	0.330	0.007	0.000	0.317	0.343	49.586	0.000
23.000	0.328	0.007	0.000	0.315	0.341	49.076	0.000
24.000	0.331	0.007	0.000	0.318	0.344	49.968	0.000
25.000	0.329	0.007	0.000	0.316	0.342	49.912	0.000
26.000	0.330	0.007	0.000	0.317	0.343	50.286	0.000
27.000	0.329	0.007	0.000	0.316	0.342	49.696	0.000
28.000	0.327	0.007	0.000	0.314	0.340	49.347	0.000
29.000	0.329	0.007	0.000	0.315	0.342	47.780	0.000
30.000	0.330	0.007	0.000	0.317	0.343	49.849	0.000
31.000	0.331	0.007	0.000	0.318	0.344	50.055	0.000
32.000	0.327	0.007	0.000	0.314	0.340	49.441	0.000
33.000	0.328	0.007	0.000	0.315	0.341	49.492	0.000
34.000	0.329	0.007	0.000	0.316	0.342	49.848	0.000
35.000	0.341	0.007	0.000	0.327	0.354	49.036	0.000
36.000	0.332	0.007	0.000	0.319	0.346	49.731	0.000
37.000	0.328	0.007	0.000	0.315	0.341	49.520	0.000
38.000	0.326	0.007	0.000	0.313	0.339	49.259	0.000
	0.330	0.007	0.000	0.317	0.343	50.242	0.000

Hedges's g (95% CI) with study removed

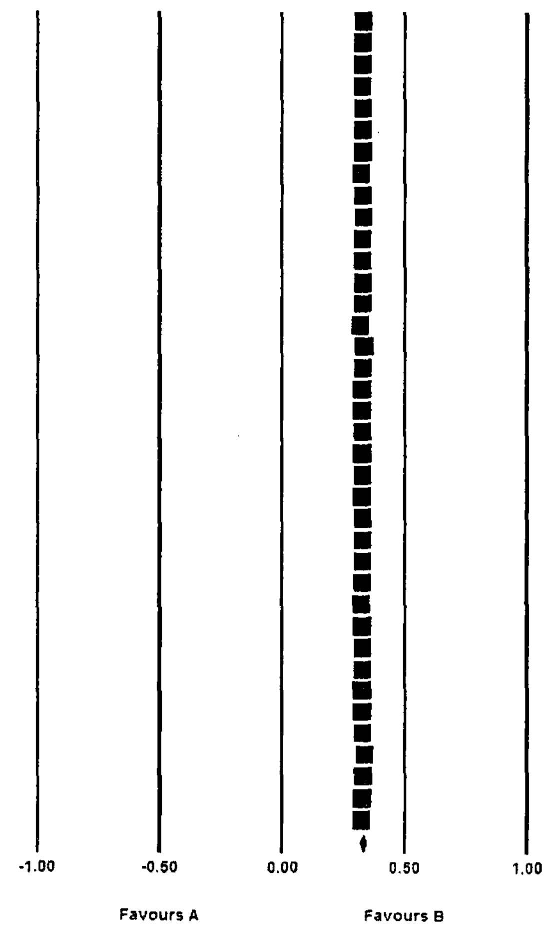


Figure 4. One study removed for the unique set of 25 effect sizes with sample-size standard error approach

Study name	Statistics with study removed						
	Point	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
1.000	0.323	0.009	0.000	0.305	0.342	34.663	0.000
2.000	0.316	0.009	0.000	0.298	0.333	34.430	0.000
4.000	0.317	0.010	0.000	0.298	0.336	32.932	0.000
5.000	0.315	0.009	0.000	0.297	0.334	33.975	0.000
6.000	0.315	0.009	0.000	0.297	0.333	34.381	0.000
8.000	0.306	0.010	0.000	0.287	0.324	32.165	0.000
10.000	0.323	0.010	0.000	0.304	0.342	33.596	0.000
11.000	0.314	0.009	0.000	0.296	0.332	34.273	0.000
12.000	0.314	0.009	0.000	0.296	0.332	34.384	0.000
13.000	0.315	0.009	0.000	0.297	0.333	34.439	0.000
14.000	0.317	0.009	0.000	0.299	0.335	34.139	0.000
17.000	0.318	0.009	0.000	0.300	0.336	34.225	0.000
20.000	0.314	0.009	0.000	0.296	0.332	33.832	0.000
21.000	0.311	0.009	0.000	0.293	0.329	33.511	0.000
24.000	0.317	0.009	0.000	0.298	0.335	34.107	0.000
25.000	0.314	0.009	0.000	0.296	0.332	34.041	0.000
26.000	0.316	0.009	0.000	0.298	0.334	34.575	0.000
28.000	0.309	0.009	0.000	0.290	0.327	33.236	0.000
31.000	0.317	0.009	0.000	0.299	0.335	34.233	0.000
32.000	0.310	0.009	0.000	0.292	0.328	33.368	0.000
33.000	0.310	0.009	0.000	0.292	0.329	33.440	0.000
34.000	0.313	0.009	0.000	0.295	0.331	33.950	0.000
35.000	0.335	0.010	0.000	0.315	0.355	32.662	0.000
36.000	0.319	0.009	0.000	0.301	0.338	33.745	0.000
38.000	0.307	0.009	0.000	0.289	0.326	33.113	0.000
	0.315	0.009	0.000	0.297	0.333	34.514	0.000

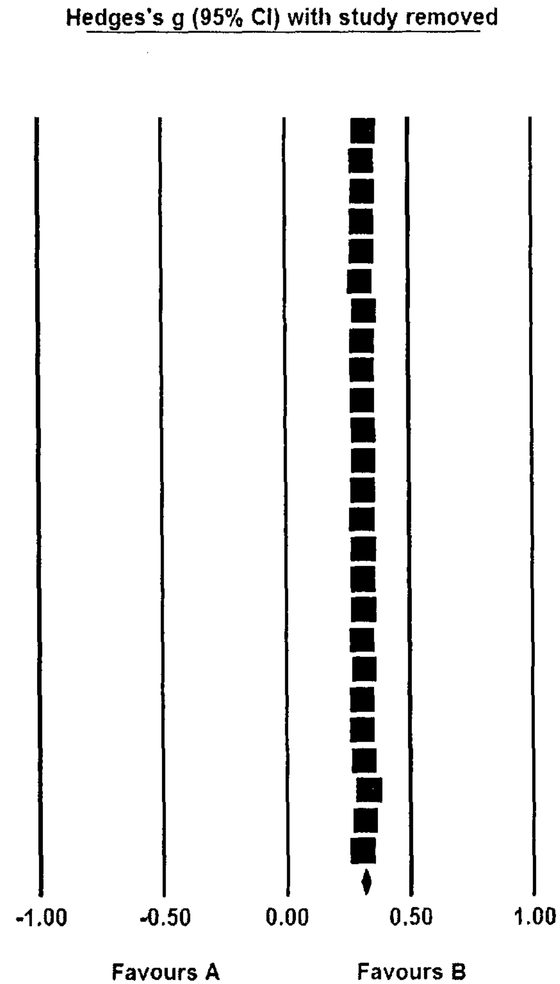


Figure 5. One study removed for overall set of 38 effect sizes with number-of-studies standard error approach

Study name	Point	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
	Statistics with study removed						
1.000	0.346	0.036	0.001	0.275	0.417	9.579	0.000
2.000	0.344	0.036	0.001	0.274	0.415	9.532	0.000
3.000	0.342	0.036	0.001	0.271	0.413	9.401	0.000
4.000	0.352	0.039	0.002	0.275	0.429	8.980	0.000
5.000	0.345	0.037	0.001	0.273	0.418	9.369	0.000
6.000	0.342	0.036	0.001	0.270	0.413	9.417	0.000
7.000	0.345	0.036	0.001	0.273	0.416	9.462	0.000
8.000	0.342	0.036	0.001	0.271	0.413	9.445	0.000
9.000	0.344	0.036	0.001	0.273	0.415	9.546	0.000
10.000	0.352	0.037	0.001	0.279	0.425	9.417	0.000
11.000	0.342	0.036	0.001	0.271	0.413	9.453	0.000
12.000	0.340	0.036	0.001	0.269	0.410	9.433	0.000
13.000	0.343	0.036	0.001	0.273	0.414	9.547	0.000
14.000	0.344	0.036	0.001	0.274	0.415	9.536	0.000
15.000	0.343	0.036	0.001	0.272	0.414	9.486	0.000
16.000	0.346	0.036	0.001	0.275	0.417	9.546	0.000
17.000	0.352	0.037	0.001	0.279	0.425	9.421	0.000
18.000	0.341	0.036	0.001	0.270	0.412	9.400	0.000
19.000	0.340	0.036	0.001	0.269	0.412	9.335	0.000
20.000	0.343	0.036	0.001	0.273	0.414	9.498	0.000
21.000	0.340	0.037	0.001	0.268	0.411	9.263	0.000
22.000	0.344	0.036	0.001	0.273	0.414	9.505	0.000
23.000	0.343	0.036	0.001	0.272	0.414	9.481	0.000
24.000	0.346	0.036	0.001	0.274	0.417	9.489	0.000
25.000	0.343	0.036	0.001	0.272	0.414	9.502	0.000
26.000	0.346	0.036	0.001	0.275	0.416	9.602	0.000
27.000	0.343	0.036	0.001	0.272	0.414	9.434	0.000
28.000	0.341	0.036	0.001	0.270	0.412	9.426	0.000
29.000	0.344	0.036	0.001	0.272	0.415	9.448	0.000
30.000	0.344	0.036	0.001	0.273	0.416	9.454	0.000
31.000	0.345	0.036	0.001	0.274	0.416	9.527	0.000
32.000	0.340	0.036	0.001	0.269	0.412	9.371	0.000
33.000	0.341	0.036	0.001	0.270	0.412	9.398	0.000
34.000	0.341	0.036	0.001	0.269	0.412	9.350	0.000
35.000	0.347	0.036	0.001	0.275	0.418	9.510	0.000
36.000	0.345	0.036	0.001	0.274	0.417	9.509	0.000
37.000	0.340	0.036	0.001	0.269	0.412	9.336	0.000
38.000	0.337	0.037	0.001	0.265	0.408	9.219	0.000
	0.344	0.036	0.001	0.273	0.414	9.564	0.000

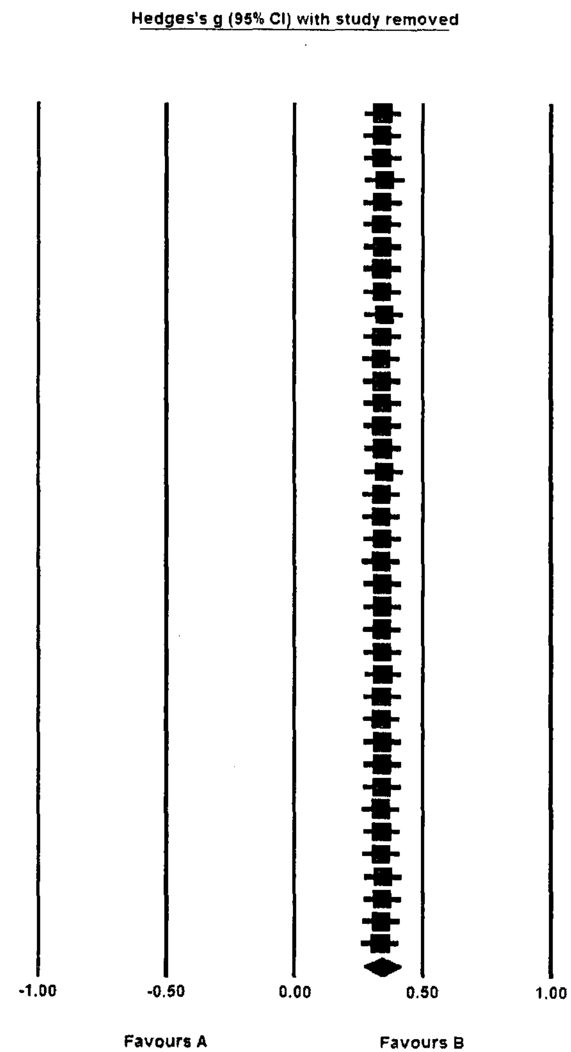


Figure 6. One study removed for unique set of 25 effect sizes with number-of-studies standard error approach

Study name	Statistics with study removed							Hedges's g (95% CI) with study removed
	Point	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	
1.000	0.336	0.042	0.002	0.253	0.419	7.952	0.000	
2.000	0.334	0.042	0.002	0.251	0.417	7.897	0.000	
4.000	0.342	0.047	0.002	0.249	0.435	7.215	0.000	
5.000	0.334	0.043	0.002	0.249	0.420	7.698	0.000	
6.000	0.330	0.042	0.002	0.246	0.413	7.761	0.000	
8.000	0.330	0.042	0.002	0.247	0.414	7.795	0.000	
10.000	0.343	0.044	0.002	0.256	0.430	7.751	0.000	
11.000	0.330	0.042	0.002	0.247	0.413	7.804	0.000	
12.000	0.327	0.042	0.002	0.245	0.410	7.783	0.000	
13.000	0.333	0.042	0.002	0.250	0.415	7.916	0.000	
14.000	0.334	0.042	0.002	0.251	0.417	7.901	0.000	
17.000	0.343	0.044	0.002	0.257	0.430	7.756	0.000	
20.000	0.332	0.042	0.002	0.249	0.415	7.856	0.000	
21.000	0.327	0.043	0.002	0.242	0.411	7.576	0.000	
24.000	0.335	0.043	0.002	0.251	0.419	7.844	0.000	
25.000	0.332	0.042	0.002	0.249	0.414	7.862	0.000	
26.000	0.336	0.042	0.002	0.253	0.418	7.980	0.000	
28.000	0.329	0.042	0.002	0.246	0.412	7.772	0.000	
31.000	0.335	0.042	0.002	0.251	0.418	7.890	0.000	
32.000	0.328	0.043	0.002	0.245	0.412	7.706	0.000	
33.000	0.329	0.043	0.002	0.246	0.412	7.738	0.000	
34.000	0.329	0.043	0.002	0.245	0.412	7.680	0.000	
35.000	0.337	0.043	0.002	0.253	0.420	7.868	0.000	
36.000	0.335	0.043	0.002	0.252	0.418	7.869	0.000	
38.000	0.323	0.043	0.002	0.239	0.407	7.526	0.000	
	0.333	0.042	0.002	0.251	0.415	7.937	0.000	

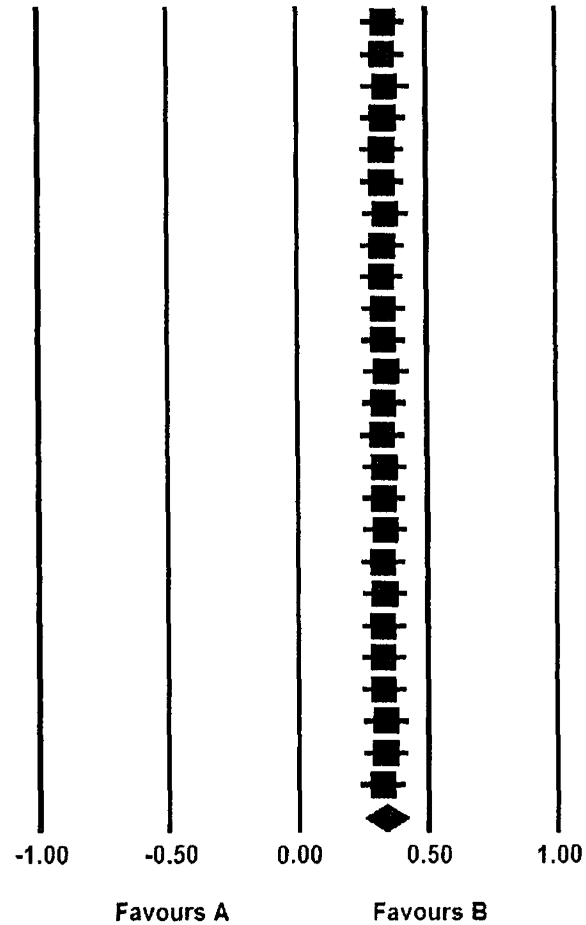


Figure 7. Funnel plot for the overall set of 38 effect sizes with sample-size standard error approach

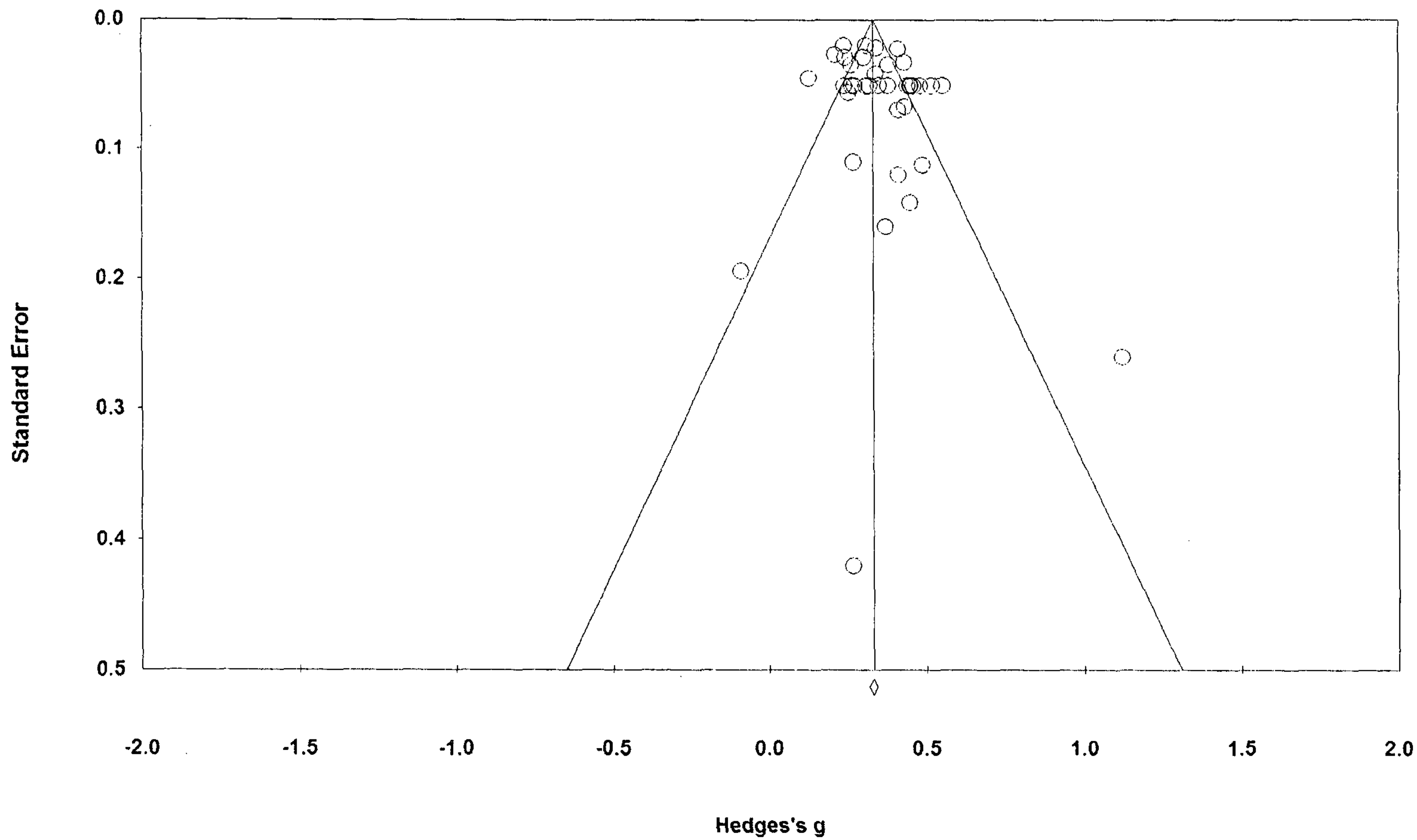


Figure 8. Funnel plot for the unique set of 25 effect sizes with sample-size standard error approach

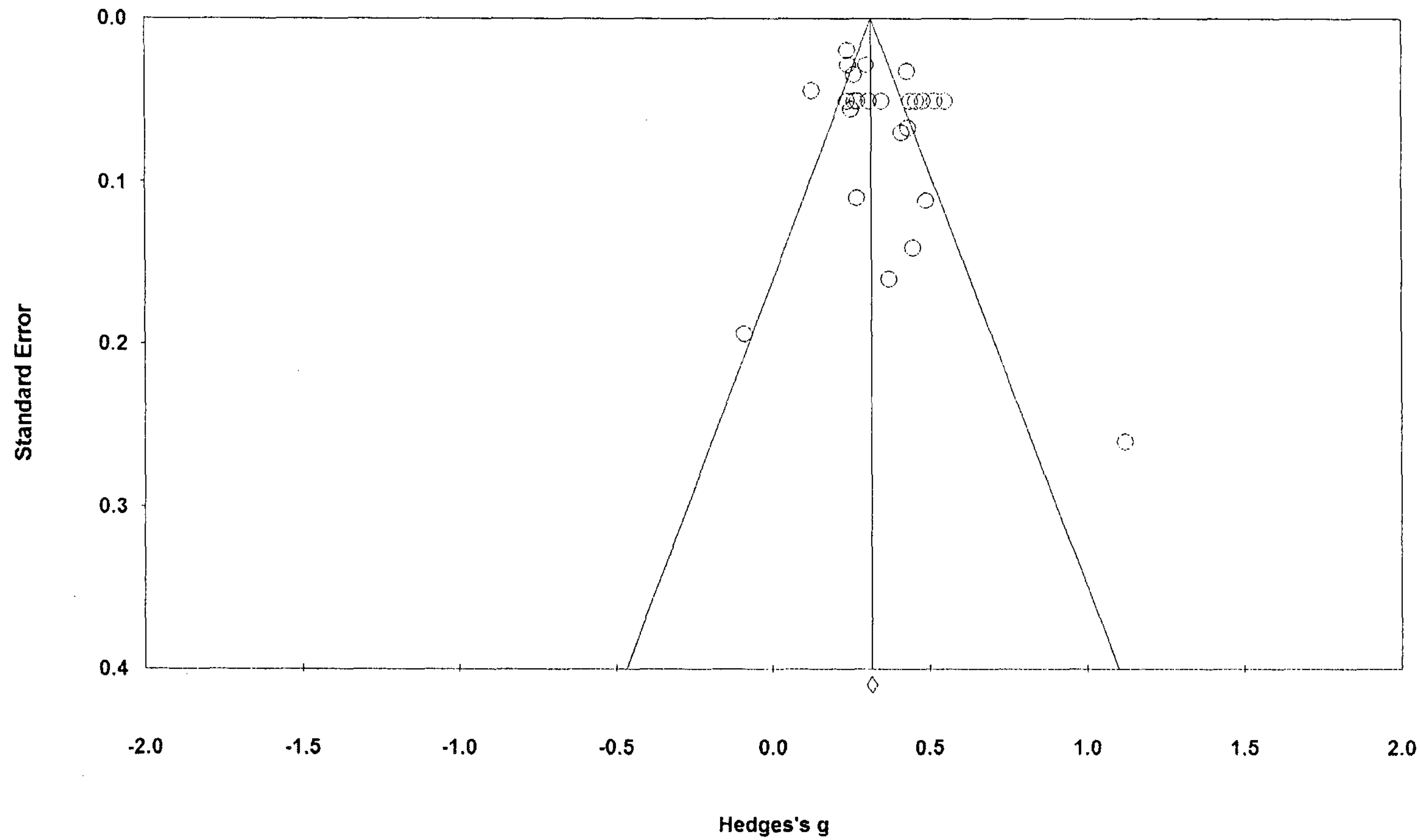


Figure 9. Funnel plot for overall set of 38 effect sizes with number-of-studies standard error approach

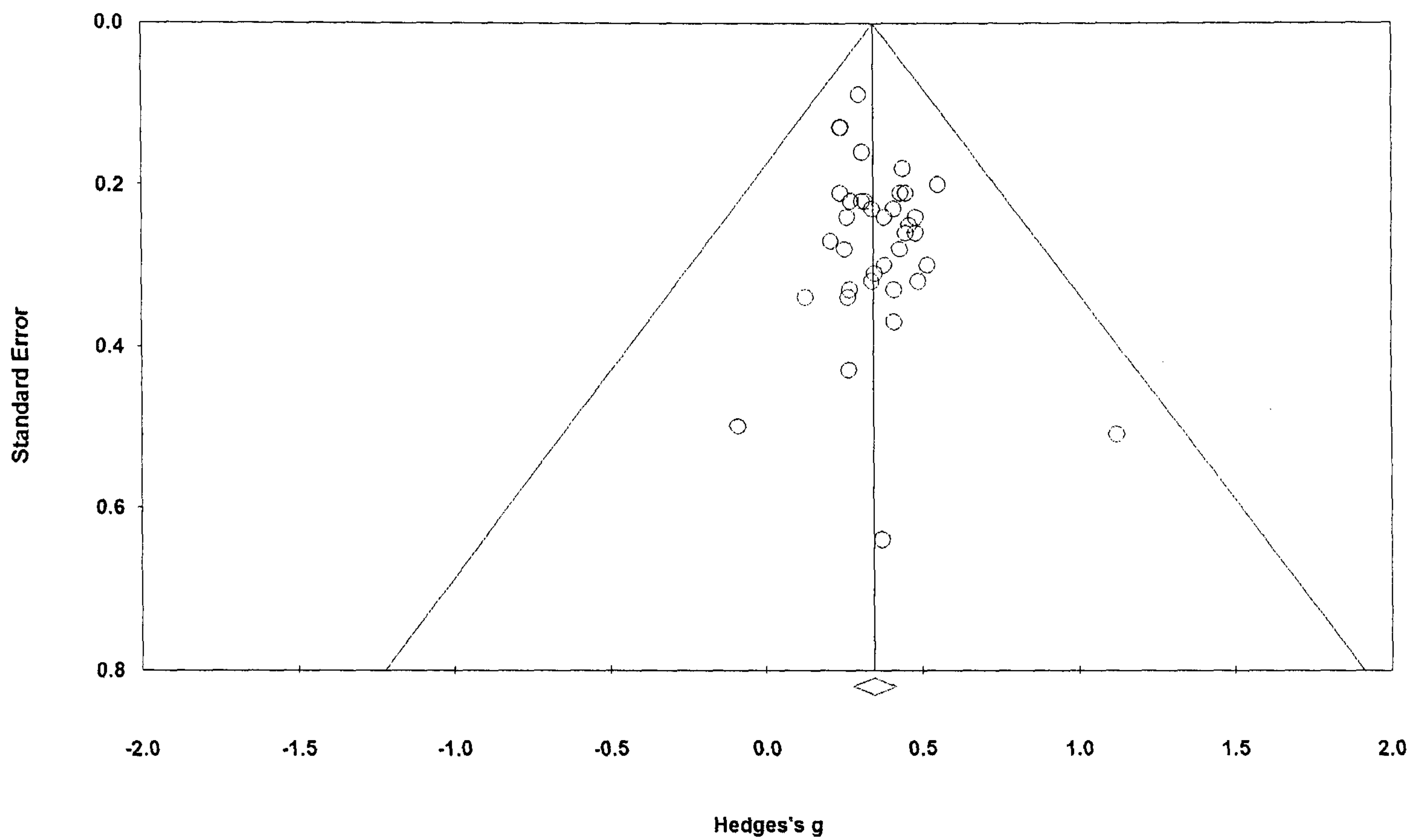
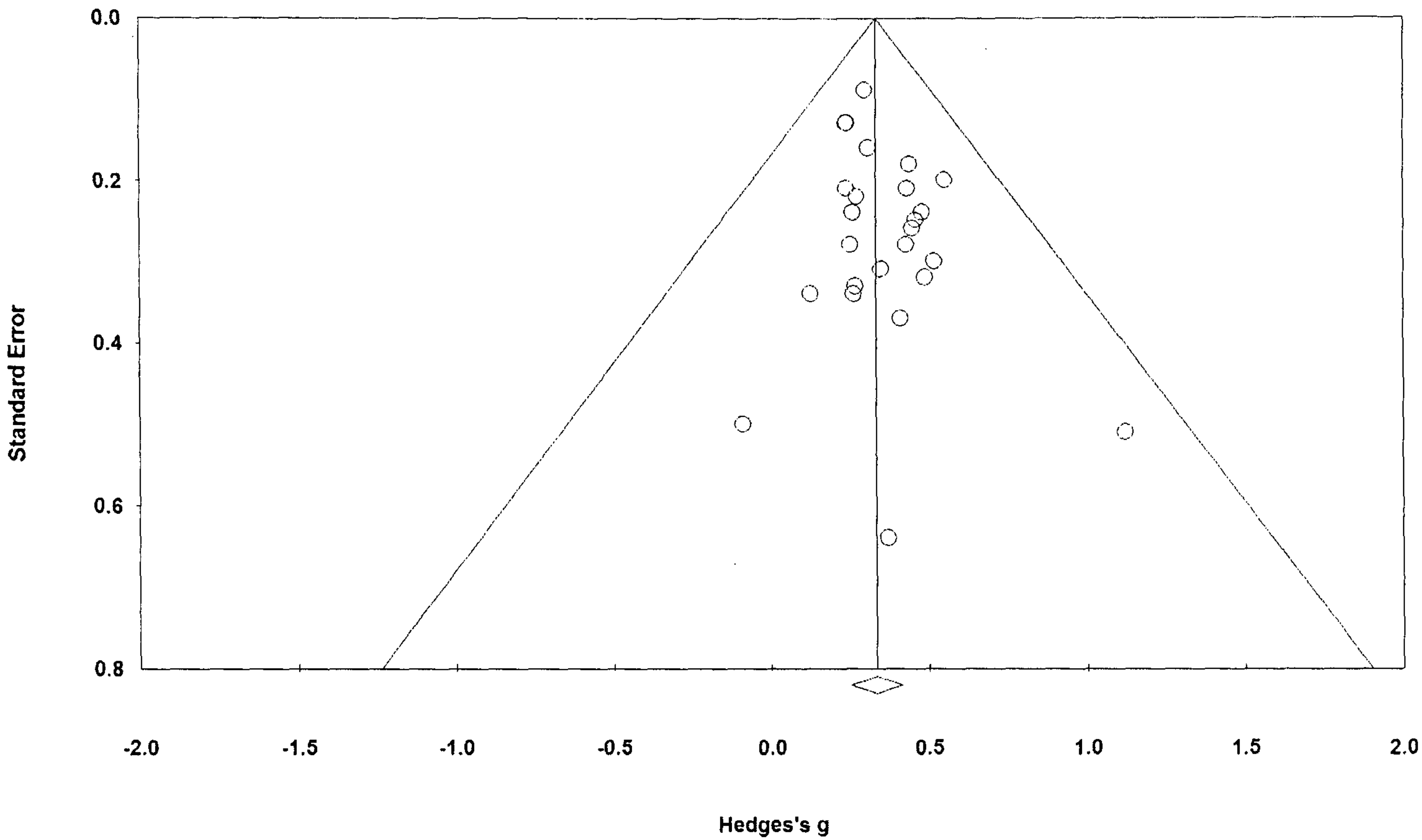


Figure 10. Funnel plot for the unique set of 25 effect sizes with number-of-studies standard error approach



Effect Size Synthesis

The effect size synthesis was conducted twice, once for the 38 effect sizes from the overall 37 included meta-analyses and once for the 25 effect sizes from the set of 25 meta-analyses with minimal overlap which were considered to be unique. This was done using the two approaches, the sample-size standard error approach and the number-of-studies standard error approach. In the current second-order meta-analysis, due to the high level of inclusivity of the primary studies, it is safe to assume that the collection of included studies is not a random sample of the population. Therefore it is appropriate to use a fixed effects model and not a random effects model while synthesizing the effect sizes, and hence, findings are reported based on the fixed effects model. However, it is important to note that results for both models, fixed and random effects, reflected findings that are extremely consistent while using sample-size standard error and identical ones when using number-of-studies standard error.

When the sample-size standard error was used, the weighted mean effect size was significantly different from zero for the overall set of meta-analyses as well as the unique ones. The point estimate of 0.330 for the overall set of effect sizes was significantly different from zero, z (38) = 50.241, p< .01, and significantly heterogeneous, Q_T (38) = 202.285, p< .01. Similarly the point estimate of 0.315 for the 25 effect sizes from the unique meta-analyses, it was also significantly different from zero, z (25) = 34.514, p< .01, and significantly heterogeneous, Q_T (25) = 142.882, p< .01. The relatively high Q value reflects the high variability in the effect sizes at the study level which was retained through this approach of using the sample-size standard error. Table 34 presents the weighted mean effect size for the overall set of 38 effect sizes and the unique set of 25 effect sizes when the sample-size standard errors were used.

Table 34. Point estimate with confidence intervals for the overall set and unique set of studies with sample-size standard error used

	Effect Size and Confidence Interval				Heterogeneity	
	k	*g+*	*SE*	95% *CI*	*Q*-value	I^2
Overall set of 37 meta-analyses	38	0.330*	0.006	0.316/0.342	202.28*	81.708
Unique set of 25 meta-analyses	25	0.315*	0.000	0.297/0.333	142.881*	83.203

$*p < 0.01$; $\chi^2_{crit}(37, \alpha = 0.01) = 59.89$; $\chi^2_{crit}(24, \alpha = 0.01) = 42.98$

When the number-of-studies standard error was used, the weighted mean effect size was significantly different from zero for the overall set of meta-analyses as well as the unique ones. The point estimate of 0.343 for the overall set of effect sizes was significantly different from zero, z (38) = 9.564, p< .01, and highly homogeneous, Q_T (38) = 9.864, p= 1. Similarly the point estimate of 0.333 for the 25 effect sizes from the unique meta-analyses, it was also significantly different from zero, z (25) = 7.936, p< .01, and highly homogeneous, Q_T (25) = 8.534, p= 1. The lower variance in the effect sizes at the meta-analysis level is reflected in the relatively small *Q* value. Table 35 presents the weighted mean effect size for the overall set of 38 effect sizes and the unique set of 25 effect sizes when the number-of-studies standard errors were used.

With both sets of meta-analyses, and with the two types of standard errors, the average effect sizes ranging between 0.315 and 0.354 reflect a medium strength effect size according to Cohen (1988), favouring the utilization of technology in the experimental condition. However, the sample-size standard error approach reveals heterogeneity in effect sizes while the number-of-studies standard error reveals homogeneity. The heterogeneity in the findings based on the sample-size standard error indicates the need to run moderator analyses in an attempt to explain the variability. All further analyses were conducted based on the set of unique 25 meta-analyses for two main reasons. First, the use of the 25 effect sizes from the unique set of meta-analyses is the less problematic outcome particularly regarding the data dependency issue and using results from the same study multiple times. Second, with both weighting approaches the mean effect size for the 25 unique meta-analyses was smaller than that for the 36 overlapping ones.

Table 35. Point estimate with confidence intervals for the overall set and unique set of studies with number-of-studies standard error used

	Effect Size and Confidence Interval				Heterogeneity	
	k	*g+*	*SE*	95% *CI*	*Q*-value	I^2
Overall set of 37 meta-analyses	38	0. 343*	0.036	0.273/0.414	9.864	0.000
Unique set of 25 meta-analyses	25	0.333*	0.042	0.250/0.415	8.534	0.000

*$p < 0.01$; $\chi^2_{crit}(37, \alpha = 0.01) = 59.89$; $\chi^2_{crit}(24, \alpha = 0.01) = 42.98$

Moderator Analysis

Knowing that the point estimate for the 25 effect sizes from the unique meta-analyses was 0.315 when the sample-size standard error was used, while being significantly heterogeneous, moderator analyses were conducted in an attempt to identify features that may explain this variability. Keeping in mind that moderator analyses based on small number of effect sizes cannot be meaningfully interpreted because of insufficient statistical power, in the instances where less than five effect sizes belonged to a particular category the decision was taken to combine them with effects from the category that is conceptually closest. Whenever this was the case, the logic behind the combination is presented in the section corresponding to the specific moderator analysis.

Methodological Quality

Starting with methodological quality, the categorical scale was used in moderator analyses. Due to the fact that only two meta-analyses were rated as weak on methodological quality, they were combined with the meta-analyses rated as moderate. Results reflected that there was a significant difference with the strong methodological meta-analyses having a smaller point estimate than the weak/moderate methodological ones with $Q_B\,(25) = 24.635, p < .001$. Table 36 presents the results for categorical methodological quality moderator analysis.

Beyond the moderator analysis for the overall methodological quality index, the specific methodological quality indexes, namely the comprehensiveness and the rigour of a given meta-analysis were used in moderator analysis. Results reflected that the comprehensiveness aspect was not a significant moderator of the effect size as presented in Table 37. On the other hand, results revealed that the rigour aspect was a significant moderator of effect size with the more rigorous meta-analysis offering a smaller point estimate as presented in Table 38.

Table 36. Moderator analysis for methodological quality

Methodological Quality	Effect size and Confidence Interval			Heterogeneity	
	k	$g+$	95% CI	Q-value	I^2
Weak and Moderate	15	0.364*	0.338/0.391	89.995*	84.444
Strong	10	0.273*	0.249/0.298	28.292*	68.189
Total within				118.287*	
Total between				24.635*	
Overall	25	0.315*	0.297/0.333	142.923*	83.208

*$p < 0.01$

Table 37. Moderator analysis for methodological quality index for comprehensiveness

Methodological Quality for Comprehensiveness	Effect size and Confidence Interval			Heterogeneity	
	k	$g+$	95% CI	Q-value	I^2
Low	7	0.287*	0.248/0.325	35.756*	83.220
High	18	0.323*	0.303/0.343	104.485*	83.730
Total within				140.241*	
Total between				2.681	
Overall	25	0.315*	0.297/0.333	142.923*	83.208

*$p < 0.01$

Table 38. Moderator analysis for methodological quality index for rigour

Methodological Quality for Rigour	Effect size and Confidence Interval			Heterogeneity	
	k	$g+$	95% *CI*	Q-value	I^2
Low	20	0.334*	0.312/0.355	128.145*	85.173
High	5	0.275*	0.244/0.307	5.989	33.211
Total within				134.134*	
Total between				8.789*	
Overall	25	0.315*	0.297/0.333	142.923*	83.208

*$p < 0.01$

Type of Publication

Keeping in mind that the included meta-analyses were either published or unpublished meta-analyses, it was of interest to investigate whether the type of publication was a moderating variable for the effect size. Moderator analysis was conducted for journal published meta-analyses versus dissertations and reports. Analysis revealed no significant difference among the two sets indicating that type of publication is not a moderating variable for effect size as presented in Table 39.

Grade Level

The included meta-analyses had various emphases concerning grade level. While some focused on a specific range of grade levels such as elementary, secondary, or post-secondary grade levels, others addressed a combination of grade levels or were even all inclusive of all grade levels within formal educational contexts. For moderator analyses, the meta-analyses were grouped into two sets; one that included those focusing on a specific range of grade levels, and another that included a combination. Results revealed that there was a significant difference among the mean effect sizes for the two groups of meta-analyses with the mean effect size for the more specific meta-analyses being smaller than the more inclusive ones as presented in Table 40.

Table 39. Moderator analysis for type of publication

Type of publication	Effect size and Confidence Interval			Heterogeneity	
	k	g+	95% *CI*	Q-value	I^2
Journal Publications	14	0.312*	0.285/0.338	83.517*	84.434
Dissertations and Reports	11	0.318*	0.294/0.342	59.281*	83.131
Total within				142.798*	
Total between				0.125	
Overall	25	0.315*	0.297/0.333	142.923*	83.208

*$p < 0.01$

Table 40. Moderator analysis for grade focus of meta-analysis

Focus of meta-analysis	Effect size and Confidence Interval			Heterogeneity	
	k	g+	95% *CI*	Q-value	I^2
Specific grade range	12	0.281*	0.257/0.306	71.890*	84.699
Combination	13	0.356*	0.329/0.383	54.506*	77.984
Total within				126.396*	
Total between				16.527*	
Overall	25	0.315*	0.297/0.333	142.923*	83.208

*$p < 0.01$

Furthermore, analysis for the moderating effect of the different ranges of grade levels was conducted with the 12 meta-analyses that addressed a particular range of grade levels. Because these were specific sub-analyses, the minimum k=5 rule was not applied. Results revealed that the grade level had a significant moderating effect with the average effect size for secondary grade levels being the largest while those for elementary and post-secondary were identical as presented in Table 41. However, due to the very small k's the results are not highly reliable and should be addressed with caution.

Table 41. Moderator analysis for specific range of grade levels

Type of publication	Effect size and Confidence Interval			Heterogeneity	
	k	$g+$	95% *CI*	Q-value	I^2
Elementary	2	0.267*	0.163/0.370	0.468	0.000
Secondary	2	0.448*	0.357/0.539	0.159	0.000
Post-secondary	8	0.267*	0.243/0.295	57.329*	87.790
Total within				57.956*	
Total between				13.934*	
Overall	12	0.281*	0.257/0.306	71.890*	84.699

*$p < 0.01$

Included Literature

The integrated meta-analyses were of two types; those that included only published primary studies, and those that included both published and unpublished primary studies. Analysis revealed that this study feature significantly moderated the average effect size with the mean effect size for the more comprehensive approach being smaller one as presented in Table 42.

Table 42. Moderator Analysis Included Literature

Included Literature	Effect size and Confidence Interval			Heterogeneity	
	k	$g+$	95% *CI*	Q-value	I^2
Only published	5	0.459*	0.386/0.533	6.930*	42.277
Published and unpublished	20	0.306*	0.288/0.324	120.238*	84.198
Total within				127.168*	
Total between				15.755*	
Overall	25	0.315*	0.297/0.333	142.923*	83.208

*$p < 0.01$

Type of Effect Size

Finally, analysis was conducted to investigate whether the type of effect size was a moderator variable on not. Although the majority of the meta-analyses used Glass's Δ, Cohen's *d,* or Hedge's *g,* some meta-analyses did not specify or give any indication of what type of effect size they used, while others used more than one and still others reported the effect size in correlational format. Therefore this analysis was conducted with the meta-analyses that used one of the three main types (Glass's Δ, Cohen's *d,* or Hedge's *g*). Due to the theoretical correspondence between Cohen's *d* and Hedge's which use the pooled standard deviation in the calculation of the effect size as compared to Glass's Δ where the standard deviation of the control group is used, it was decided to consider two groups for this analysis, namely, Glass's Δ versus Cohen's *d* and Hedges' *g.* Results revealed that the type of effect size was a significant moderating variable with the Cohen's *d* and Hedges' *g* being smaller than the Glass's Δ as presented in Table 43.

Table 43. Moderator Analysis for Type of Effect Size

Types of effect size	Effect size and Confidence Interval			Heterogeneity	
	k	$g+$	95% *CI*	Q-value	I^2
Glass Δ	6	0.384*	0.347/0.421	37.725*	86.746
Cohen's *d* & Hedges' *g*	13	0.318*	0.293/0.343	65.193*	81.593
Total within				102.918*	
Total between				8.458*	
Overall	19	0.338*	0.318/0.359	111.376*	83.839

$^*p < 0.01$

Specific Effect Sizes

As presented earlier, specific effect sizes pertaining to type of technology, subject matter, and grade level were extracted from the different included meta-analyses. This was based on the fact that these features were the most recurring study features in the literature for which individual effect sizes were reported. In particular situations, where authors reported a specific effect size based on less than three studies, that specific effect size was ignored. In a perfect situation where authors provide adequate information, it might have been possible to integrate the effect sizes addressing a specific feature and conduct moderator analyses to understand more about moderating variables. However, and due to reporting limitations, this was hardly possible, particularly in the absence of specific standard errors and information that might be helpful in their calculation or imputation. The analysis for the specific effect sizes will be limited to their presentation and general reflections.

In this section, the specific effect sizes reported in the different meta-analyses will be presented in tabular form. Table 44 presents the specific effect sizes for grade levels,

Table 45presents the specific effect sizes for different subject matter, and Table 46 presents the specific effect sizes for the different technology tools or utilizations. Within each of the categories of the three study features, the specific effect sizes are listed in decreasing order of overall methodological index. In certain cases, various independent average effect sizes that belong to one of the subcategories selected for this process were reported in the same meta-analysis and these were extracted and reported individually in the corresponding table.

Table 7. Specific effect sizes for grade levels reported in the different meta-analyses

	Meta-analysis Reference	# of studies	# of ES	Mean ES	Comprehen siveness	Rigour	Overall Quality
Preschool	Fletcher-Flinn & Gravatt (1995)	6	6	0.55	4	4	8
Elementary	Fletcher-Flinn & Gravatt (1995)	27	27	0.26	4	4	8
	Khalili & Shashaani (1994)	9	9	0.34	4	4	8
	Liao (1998)	8	8	0.30	5	2	7
	Roblyer (1988)	44	44	0.32	5	2	7
	Rosen & Salomon (2007)	23	23	0.42	3	2	5
Secondary	Onuoha (2007)	19	19	0.24	7	5	12
	Bayraktar (2000)	missing	53	0.27	7	3	10
	Yaakub (1998)	missing	4	0.11	6	3	9
	Fletcher-Flinn & Gravatt (1995)	20	20	0.20	4	4	8
	Khalili & Shashaani (1994)	7	7	0.62	4	4	8
	Khalili & Shashaani (1994)	6	6	0.11	4	4	8
	Liao (1998)	4	4	0.09	5	2	7
	Roblyer (1988)	22	22	0.19	5	2	7
	Rosen & Salomon (2007)	9	9	0.58	3	2	5
Elementary	LeJeune (2002)	9	9	0.14	6	3	9
Secondary	LeJeune (2002)	9	9	0.42	6	3	9
Post-secondary	Onuoha (2007)	17	17	0.21	7	5	12
	Timmerman & Kruepke (2006)	78	78	0.28	7	5	12

Bayraktar (2000)	missing	55	0.27	7	3	10
Hsu, Yung Chen	3	4	0.43	6	3	9
Hsu, Yung Chen	3	4	0.53	6	3	9
LeJeune (2002)	11	11	0.49	6	3	9
LeJeune (2002)	13	13	0.35	6	3	9
Yaakub (1998)	missing	14	0.45	6	3	9
Fletcher-Flinn & Gravatt (1995)	48	48	0.20	4	4	8
Khalili & Shashaani (1994)	12	12	0.45	4	4	8
Liao (1998)	15	15	0.44	5	2	7
Roblyer (1988)	10	10	0.57	5	2	7

Table 8. Specific effect sizes for the different subject matter reported in the different meta-analyses

	Meta-analysis Reference	# of studies	# of ES	Mean ES	Comprehen siveness	Rigour	Overall Quality
Art/music	Liao (1998)	3	3	-0.15	5	2	7
computer sc/tech	Kuchler (1998)	11	11	0.35	7	3	10
	Khalili & Shashaani (1994)	3	3	0.28	4	4	8
language	Fletcher-Flinn & Gravatt (1995)	23	23	0.12	4	4	8
	Khalili & Shashaani (1994)	6	6	0.17	4	4	8
	Liao (1998)	11	11	0.60	5	2	7
	Roblyer (1988)	4	4	-0.06	5	2	7
	Roblyer (1988)	33	33	0.26	5	2	7
	Christmann et al.(1997)	3	4	-0.42	4	2	6
	Christmann et al.(1997)	3	4	0.26	4	2	6
	Niemiec (1987)	missing	9	0.24	3	1	5
math	Fletcher-Flinn & Gravatt (1995)	24	24	0.32	4	4	8
	Khalili & Shashaani (1994)	18	18	0.52	4	4	8
	Liao (1998)	5	5	0.12	5	2	7
	Roblyer (1988)	35	35	0.36	5	2	7
	Christmann et al.(1997)	13	15	0.18	4	2	6
	Niemiec (1987)	missing	28	0.28	3	2	5
problem solving	Kuchler (1998)	11	11	0.24	7	3	10
	Khalili & Shashaani (1994)	11	11	0.41	4	4	8

	Liao (1992)	16	missing	0.20	3	3	6
	Niemiec (1987)	missing	20	0.12	3	2	5
	Samson et al.(1986)	missing	20	0.12	4	1	5
science	Timmerman & Kruepke (2006)	64	64	0.18	7	5	12
	Timmerman & Kruepke (2006)	21	21	0.32	7	5	12
	Fletcher-Flinn & Gravatt (1995)	31	31	0.26	4	4	8
	Khalili & Shashaani (1994)	7	7	0.12	4	4	8
	Liao (1998)	4	4	0.15	5	2	7
	Liao (1998)	5	5	0.89	5	2	7
	Roblyer (1988)	4	4	0.49	5	2	7
	Christmann et al.(1997)	6	9	0.64	4	2	6
	Christmann & Badgett (2000a)	3	5	0.17	3	2	5
social science	Liao (1998)	7	7	0.72	5	2	7
	Timmerman & Kruepke (2006)	10	10	0.06	7	5	12
	Timmerman & Kruepke (2006)	22	22	0.37	7	5	12

Table 9. Specific effect sizes for the different technology tools or utilizations reported in the different meta-analyses

	Meta-analysis Reference	# of studies	# of ES	Mean ES	Comprehensiveness	Rigour	Overall Quality
Drill	Bayraktar (2000)	missing	22	-0.11	7	3	10
	Kuchler (1998)	15	15	0.51	7	3	10
	Fletcher-Flinn & Gravatt (1995)	46	46	0.23	4	4	8
	Khalili & Shashaani (1994)	4	4	0.11	4	4	8
	Roblyer (1988)	18	18	0.27	5	2	7
	Liao (1992)	6	missing	0.15	3	3	6
	Niemiec (1987)	missing	146	0.47	3	2	5
	Samson et al.(1986)	missing	51	0.26	4	1	5
Expert system	Hsu (2003)	3	4	0.99	6	3	9
offline instruction	Samson et al.(1986)	missing	45	0.20	4	1	5
Simulation	Bayraktar (2000)	missing	43	0.39	7	3	10
	Hsu (2003)	3	5	0.48	6	3	9
	Fletcher-Flinn & Gravatt (1995)	25	25	0.25	4	4	8
	Khalili & Shashaani (1994)	4	4	0.79	4	4	8
	Liao (1998)	7	7	0.48	5	2	7
	Liao (2008)	6	6	0.70	5	2	7
	Liao (1992)	6	missing	0.31	3	3	6

	Samson et al.(1986)	missing	5	0.58	4	1	5
tutorial	Bayraktar (2000)	missing	26	0.37	7	3	10
	Kuchler (1998)	12	12	0.54	7	3	10
	Khalili & Shashaani (1994)	8	8	0.26	4	4	8
	Liao (1998)	2	2	0.97	5	2	7
	Liao (1992)	3	missing	2.38	3	3	6
	Niemiec (1987)	missing	16	0.34	3	2	5
	Samson et al.(1986)	missing	23	0.64	4	1	5
Web-based	Hsu (2003)	4	4	0.23	6	3	9
Word processing	Fletcher-Flinn & Gravatt (1995)	24	24	0.22	4	4	8
	Roblyer (1988)	8	8	0.23	5	2	7
managed instruction	Samson et al.(1986)	missing	10	0.32	4	1	5

Validation of Average Effect Size

One of the objectives of this research project was to validate the findings of the second order meta-analysis particularly regarding the average effect size pertaining to the impact of technology use on students' achievement. To allow for the validation of the findings of the second-order meta-analysis, the raw data from the included meta-analyses were extracted in order to use them in the calculation of the point estimate in a process similar to a regular meta-analysis and not a second order one. From the 25 unique meta-analyses 13 offered information allowing for the extraction of 574 individual effect sizes and their corresponding sample sizes, with total number of participants being 60,853, to be used in the validation process. With the fixed effects model, the weighted mean effect size of 0.304 for the 574 individual effect sizes was significantly different from zero, z (574) = 37.13, p< .01, and highly heterogeneous, Q_T (574) = 2927.87, p< 0.01. The random effects model revealed highly similar results with the weighted mean effect size for the 574 individual effect sizes being 0.327 and significantly different from zero, z (574) = 16.55, p< .01. Table 47 presents the weighted mean effect size for the overall set of 574 individual effect sizes extracted from 13 different meta-analyses, with both fixed and random effect models.

Table 47. Point estimate with confidence interval for the set of 574 individual effect sizes

Model	Effect Size and Confidence Interval				Heterogeneity	
	k	*g+*	*SE*	95% *CI*	*Q*-value	I^2
Fixed	574	0. 304*	0.008	0.288/0.320	2927.869*	80.429

$*p < 0.01$, $\chi^2_{crit}(573, \alpha = 0.01) = 654.68$

CHAPTER 5: DISCUSSION

With the ongoing interest in the impact of technology on learning, and the growing attention to and increasing number of published meta-analyses this dissertation had two main components. The first component was methodological aiming at: a) designing an approach to assess the methodological quality of meta-analyses in the social science field; b) piloting a second-order meta-analysis procedure that takes into consideration methodological quality; and c) validating the results of the second order meta-analysis. The second component aimed at answering substantive questions related to meta-analyses addressing the impact of computer technology on student achievement in formal educational contexts through implementing the second-order meta-analysis methodology. The objectives of the second component were to: a) critically examine the meta-analyses addressing the impact of computer technology on learning; b) synthesize the findings of meta-analyses addressing technology integration and student achievement through a second-order-meta-analysis; and c) explain the variance in the effect sizes if possible. Beyond the methodological aspects, particular research questions addressed in this dissertation were:

1. Does technology use enhance student achievement in formal face-to-face classroom settings as compared to traditional settings? If so, to what extent?
2. What features, if any, moderate the overall effects of technology use on students' achievement?

Through the discussion chapter the results and findings of this study will be addressed in two major sections, one pertaining to the methodological aspects related to the second-order meta-analysis, and one targeting the findings from the implementation of the proposed methodology with the technology integration meta-analysis literature. These will be followed by a section in which reflections on the strengths, limitations, and future directions are presented and discussed. Finally, general conclusions and recommendations are offered.

Second-Order Meta-Analysis Procedure

As presented earlier, one of the main components of this dissertation was methodological and procedural in nature. It entailed the development and pilot testing of a systematic methodology for a second-order meta-analysis which could help in offering reliable answers to general questions by making use of already available meta-analyses. In addition, the goal was to design and implement an approach to assess the methodological quality of meta-analyses.

Before reflecting on the suggested and implemented second-order meta-analysis, it is important to present the suggested procedure that was piloted in this study. The overall approach was based on the general guidelines and procedures followed in a regular meta-analysis with some modifications to achieve the specified objectives. Throughout the whole process, measures should be taken to ensure clarity, comprehensiveness and reliability of the review. The steps suggested and followed in this study were:

1. Specifying the research question.

The first and foremost stage in a second-order meta-analysis is specifying the research question to be answered. This is of extreme importance because the entire review process will be guided by its scope and boundaries. Although the research question should be a general one to allow for the location of various meta-analyses that address it, it should be focused enough to allow for some common grounds among the included meta-analyses and help in reaching results which could be interpreted meaningfully. For example, one of the very first attempts at synthesizing meta-analytic findings was conducted by Lipsey and Wilson (1993) to assess the efficacy of psychological, educational, and behavioural treatments. Their review was rather too general leading to harsh criticisms to the approach, particularly by Eysenck (1995) who argued that "a method that averages apples, lice, and killer whales (here psychological, educational, and behavioural treatments) can hardly command scientific respect" (p.110). With this in mind, one needs to focus the question to avoid the apples and oranges controversy at an even higher level than a regular meta-analysis.

2. *Creating inclusion/exclusion criteria.*

The criteria for including a meta-analysis in a given second-order meta-analysis should be developed based on the research question and the focus of the review. As with the current study, attention should be given to the specificity and clarity of the inclusion and exclusion criteria as much as possible. The criteria should address the various aspects based on which a meta-analysis will be included in the second-order meta-analysis or not. Similar to all reviews, the specificity and clarity of the criteria will help in: a) setting the scope of the second-order meta-analysis and determining the population to which generalizations will be possible based on the limitations and boundaries set by the inclusion/exclusion criteria; b) designing and implementing the most adequate search strategy to identify and retrieve the most pertinent literature to the research question; and c) minimising bias in the review process for inclusion of meta-analyses. Inclusion criteria should include aspects that deal with contextual features related to the research question, as well as criteria targeting publication features such as date of publication and type of publication.

3. Developing and implementing search strategies.

Similar to regular meta-analyses, this step is extremely important and has a major impact on the comprehensiveness and contemporariness of a given second-order meta-analysis. The attention devoted to the design of the search strategy and the use of a variety of resources including databases, branching, hand searches, and web searches will help in the location of the most relevant meta-analyses with minimizing the risk of missing pertinent documents. Generally speaking and due to the specificity of the meta-analytic approach, the search parameters for meta-analyses to be included in a second-order meta-analysis are more specific, and the relevant documents may be easier to locate. However, one should not be limited to the term "meta-analysis" because of the presence of various other terms that may be used by authors such as "systematic review" and "quantitative synthesis". In addition, for a more comprehensive and representative set of meta-analyses, and to avoid publication bias and the file drawer effect, the search should not be limited to published literature but should include unpublished documents such as conference papers, dissertations, and reports.

4. Reviewing and selecting meta-analyses.

The review and selection process is a crucial phase during which the documents located by the literature search are screened for inclusion based on the pre-set criteria. Although it may seem to be rather straight forward, it may be influenced by the reviewer's personal biases or may be liable for some errors or oversights and therefore measures should be taken to minimize subjectivity in the process. As with a regular meta-analysis, the specificity of the inclusion/exclusion criteria usually helps in making the process more straight forward. However, and based on the current piloting of the procedure, the review and selection process for a second-order meta-analysis resulted in a smaller number of discrepancies among the reviewers than that for a regular meta-analysis. This may be easily explained by the relatively smaller number of documents to review and the specificity of the meta-analytic documents as compared to primary studies to be reviewed for inclusion in a regular one.

5. Extracting effect sizes and standard errors.

Considering that the effect size is the common metric used in all meta-analyses, other than having to note which type of effect size was used, the extraction process of the effect sizes should prove to be relatively straight forward. Even with the cases where the effect sizes do not reflect the group mean differences such meta-analyses that express the effect sizes as standard correlation coefficients, the conversion processes are not hard to implement.

In a second-order meta-analysis, things are not as simple with standard errors since there are two different types of variances, one reflecting the variability at the study level, and one reflecting the variability at the meta-analysis review level. For the study variance, the sample sizes to be used in the standard error computation are clearly the numbers of participants in a given study. This will be reflective of the variability among all the individual effect sizes calculated and incorporated in the collection of included meta-analyses.

However, in a meta-analysis, the study is the unit of analysis with the sample size being the number of included studies. Therefore, the meta-analysis review level variance has to be reflective of the variability based on the number of included studies in each meta-analysis. With this variance, the standard error computation should be based on the number of studies included in a given meta-analysis. Knowing that each study is contributing a control and an experimental group it is logical to use the same number of studies as both experimental and control. Calculations based on this approach will be reflective of the variability among all the average effect sizes (point estimates) calculated in the included meta-analyses. Moreover, this approach makes sense when study features associated with the methodological aspects of the meta-analyses under review (issues related to comprehensiveness and rigour) are to be addressed in the moderator analyses.

As mentioned earlier, there are other advantages and disadvantages to both approaches of standard error calculation. The sample-size standard error approach makes use of the strength of meta-analyses and allows for reliable conclusions, while keeping the enormous variability in the study findings intact thus magnifying the heterogeneity in the findings. With this approach, we might be running a higher risk of type I error and finding false positives. On the other hand, the number-of-studies standard error approach does not overstate the heterogeneity, but it ignores the actual strength offered by the individual point estimates from the different meta-analyses, and increases the chances of committing type II error and finding false negatives. Currently, and due to the novelty of the approach, there is no valid support for excluding either one of the two approaches, and for the purpose of this second-order meta-analysis, standard errors reflecting both variances were calculated to allow for conducting and comparing analyses with both approaches.

6. Developing a codebook.

Similar to a regular meta-analysis, each second-order meta-analysis depends on a specific codebook to help in the extraction and collection of information from the included studies. Specificity is highly important when designing the codebook in order to ensure the clarity of the codes and thus minimise ambiguity that may lead to disagreements among the reviewers' decisions. The codebook should be tailored to the particular objectives of the second-order meta-analysis and its area of focus. For example, aspects related to the research question would have to be designed with the particular objective of the second-order meta-analysis. Nevertheless, some of the features presented in the codebook used for the current second-order meta-analysis may be used either as is, or with some slight modifications or adaptations in relation to the objectives of each particular second-order meta-analysis. An example would be the items used to assess the methodological quality of a given meta-analysis to be discussed shortly after.

7. Coding study features.

Although one may tend to believe that once the codebook is set, the process of coding study features is relatively easy, reality reflects a different situation. This step may prove to be one of the most challenging steps in any regular or second-order meta-analysis. The reason for this is the high variability in the information that authors do include in their reports and the enormous amount of missing information that is left out by many. The researchers coding study features have to be very keen at picking details without allowing their personal information and biases to influence their interpretation of the presented information in a document.

8. Designing and calculating the methodological quality index.

With the variability in the methodological quality of published meta-analyses, it is important to address this aspect when reviewing and synthesizing meta-analyses, which was not addressed by the previous authors who conducted second-order meta-analyses of quantitative reviews of meat-analyses. For this second order meta-analysis, the codebook included methodological features which allowed for the calculation of a methodological quality index. The items included 14 different study features that address various methodological aspects pertaining to meta-analysis such as comprehensiveness, scope, contemporariness, accuracy and detail. The items were:

1. Literature covered
2. Resources used
3. Included research type
4. Number of data-bases searched
5. Search strategy
6. Inclusion/exclusion criteria
7. Article review
8. Effect size extraction
9. Codebook
10. Study feature extraction
11. Independence of data
12. Weighting by number of comparisons
13. Time between last study included and publication
14. Standard error calculation

For a full list of 14 items included in the methodological quality index and how the transformation occurred whenever there were more than 2 levels refer to Table 3.

9. Identifying unique set of meta-analyses.

A major difference between a second-order meta-analysis and a regular one is the need to resolve the overlap in the primary studies included in each meta-analysis. This requires the identification of the set of meta-analyses with the least overlap and which could be considered as the set of unique meta-analyses. Based on the literature review of previously conducted second-order meta-analyses or quantitative syntheses of meta-analyses, there seems to be a consensus that there should be a cut point for the maximum overlap in primary literature among any two meta-analyses. With the current second-order meta-analysis approach, two other aspects played an important role in the decision. While assessing the overlap among the included meta-analyses the decision was to identify the set of meta-analyses that had the lowest level of overlap in primary studies while retaining the highest percentage of the overall set of primary studies. At the same time, attention was given to the methodological quality in order to retain the high quality meta-analyses based on the methodological quality index and not lose them from the final set of included meta-analyses.

10. Conducting statistical analyses.

The proposed and implemented analysis phase went beyond what was applied by previous researchers who experimented with second-order meta-analysis or quantitative reviews of meta-analyses. None of the previous attempts at synthesizing meta-analytic findings used a weighting process to calculate the mean effect size or conducted moderator analyses of any sort. Rather, the majority presented an overall average effect size with standard deviation in addition to specific average effect sizes and standard deviations for particular subsets based on specific features (e.g., Butler, Chapman, Forman, & Beck, 2006; Hammill, 2004; Lipsey & Wilson, 1993; Sipe & Curlette, 1997). With the current approach, the analysis phase was not limited to offering an average effect size. Tests of homogeneity were conducted with two different approaches, one using the sample-size standard error and one using the number-of-studies standard error. In addition, moderator analyses were conducted for substantive features as well as methodological quality features with the goal of explaining the variance in the extracted effect sizes when heterogeneity was found with the sample-size standard error approach.

11. Interpretation.

Once the various phases of document review, effect size extraction, study feature coding, synthesis, and analysis are over, the interpretation of the results takes center stage. With meta-analyses, there is always the fear that too much confidence is placed in the findings with the temptation to overgeneralize them without being fully aware of the inherent limitations that may influence their reliability and credibility (Preiss & Allen, 1995). This concern is all the more valid at the second-order meta-analysis level for various reasons. With the novelty in the implemented procedures one needs to keep in mind the advantages and disadvantages of using each of the two approaches (the sample-size standard error and the number-of-studies standard error) in synthesizing the effect sizes. Other aspects that one needs to consider while interpreting the results include the high level of detachment from the original data, the quality of meta-analyses included in the second-order meta-analysis, and the assumptions made throughout the process regarding certain statistical aspects.

Finally, it is important to note that to ensure reliability and to avoid personal bias the various steps should be conducted by two researchers working independently with the reporting of the inter-rater agreement to reflect the level of confidence in the decisions reached. As with regular meta-analyses, the researchers would meet to resolve any discrepancies and finalize decisions once the review or extraction process is completed. If resources are limited and this is not possible, a less demanding process may be employed where a sample of the documents is reviewed by two researchers to allow for rating comparison and establishing inter-rater agreement.

Technology Integration: Second-Order Meta-Analysis

The application of the proposed methodology with meta-analyses addressing the impact of technology on students' achievement in formal face-to-face educational contexts resulted in the review of more than 400 documents. From these, 37 distinct meta-analyses involving 1253 different primary studies comparing student achievement in technology enhanced classroom settings to traditional instruction were included in this second-order meta-analysis. The number of participants included on these sets of meta-analyses was approximately 130,300. These 37 meta-analyses addressed a variety of technological approaches that were used in the experimental conditions to enhance and support face-to-face instruction, such as computer assisted instruction, computer based instruction, simulations, word processors, and computer mediated communication.

The overlap of primary studies in the 37 meta-analyses was checked to identify the set with the lowest level of overlap in primary studies while retaining the highest percentage of the overall set of primary studies and preserving the high methodological quality meta-analyses. The final number of meta-analyses that were considered to be unique or having acceptable levels of overlap was 25 with none having a frequency of overlapping studies beyond 25%. The overall number of primary studies included in this set was 1055 studies which represents 84.2% of the overall number of primary studies included in the overall set of meta-analyses. The number of participants included in these sets of meta-analyses was approximately 109,700.

In the following sections, the findings regarding the critical examination of the overall set of 37 meta-analyses will be discussed followed by the synthesis of the 25 unique ones.

Critical Examination of the Included Meta-Analyses

The included set of 37 meta-analyses reflected the high level of attention given to the meta-analytic technique as well as to the impact that technology integration has on students' achievement. The three relevant meta-analyses published between 1985 and 1990 were followed by seven between the years 1991 and 1995. The number increased to nine meta-analyses in the time period between 1996 and 2000 only to increase further to 10 newer ones published between 2001 and 2005. Finally, there were eight new meta-analyses addressing technology integration in formal education as compared to traditional teaching environments between the years 2006 and 2008 only which will probably increase in the future. The increasing trend in the number of meta-analyses conducted over the years in this area is an attestation to the ongoing interest in the role played by technology in the learning process and the growing belief in meta-analysis as a viable technique for research synthesis. Moreover, the massive number of 1253 different primary studies integrated in the meta-analyses is a further confirmation of the continued interest in technology's impact on students' achievement.

Besides proposing a methodology for conducting a second-order meta-analysis, this study aimed at answering substantive questions related to meta-analyses addressing the impact of computer technology on student achievement through implementing the proposed methodology. One of the specific objectives was to critically examine the meta-analyses addressing the impact of computer technology on learning by reviewing and analyzing various methodological aspects. The findings will be discussed in the following four sections: a) contextual features; b) methodological features; c) analysis phase; and d) further reporting aspects.

Contextual Features

The contextual features of the included meta-analyses reflected a high level of variability in their focus concerning the grade level, the addressed technology, and the subject matter.

First: Grade level

Although the level which stood out as receiving particular attention was the post-secondary with nine meta-analyses focusing on it, most of the included meta-analyses, particularly 17 out of the 37, targeted a combination of age groups. This indicates the continuing interest in technology's impact on students' achievement throughout various academic levels. These findings were highly expected in light of the overall attention to technology's role in the educational context as a whole.

Second: Addressed technology

Reviewed meta-analyses targeted a wide variety of technologies with the highest attention given to a general category of *Computer Assisted Instruction* where 18 out of the 37 included meta-analyses focused on it. However, with the problem that we have regarding terminology and definitions in the field, one cannot confidently assume that all the meta-analyses addressing computer assisted instruction are actually targeting the same technological tools or the same pedagogical use of a given tool. The review reflected a very high level of variability at the terminology level and findings with this regard should be taken cautiously. One reason why computer assisted instruction may be the one targeted the most could be its inclusivity of a wide range of tools and teaching approaches, and its continuing usage over the years.

Third: Subject matter

Although the majority of the meta-analyses addressed a combination of topics, language seemed to be of highest interest followed by science and health. This is not surprising given the variety of technological tools targeting the development of language skills, and those that address science and health instruction through a variety of techniques including simulations.

The odd finding was the lack of attention given to both technology and information literacy as content areas. Each of these two topics was addressed by one meta-analysis only which may indicate a minor interest for technology use in this area and may be a reflection of the scarcity of primary studies in the field. Another explanation may be the idea that technology is so central in both subject areas to the extent where it is considered so transparent and common place that it is not an important factor to be studied independently. In both cases, this is quite alarming since technology is fundamental in both technology studies and information literacy particularly with the advent of the World Wide Web and the wide-ranging use of the internet. Moreover, the ubiquitous use of computer technology in the 21st century renders an understanding of both technology studies and information literacy indispensable to our students in most of their future endeavours. Therefore, it is important to have a deeper understanding of the role technology may play in the instruction of technology and information literacy in order to better prepare the students for future challenges.

Methodological Features

While designing the codebook for this second-order meta-analysis, a lot of consideration was given to methodological aspects where a substantive number of the study features addressed the procedural aspects of the included literature. It is important to keep in mind that the actual codes and the coding procedure were dependent on and limited by the quality of the report at hand. Nevertheless, the process proved to be of extreme importance in the understanding of the quality of the included meta-analyses. The results reflect some of the general variations and developments that meta-analytic procedures have witnessed, in addition to revealing some strengths and weaknesses in the set of included meta-analyses and will be discussed in the following subsections.

First: Conceptual issues

The codebook included open codes to address the definitions for both the experimental and control groups. In an area where terms such as *Computer Assisted Instruction, Computer Based Instruction,* and *Traditional Instruction,* are so vague and unclear, one would expect researchers to be adamant on defining their terms to avoid ambiguity. Interestingly this was anything but the case in the literature under investigation. This reflects negatively on the ability to understand the focus as well as the findings of a given meta-analysis. Moreover, this minimizes the possibility of explaining the variance in effect sizes due to the inability to code for a variety of contextual features pertaining to the use of technology and instructional design. More attention should be given to this issue by authors as well as reviewers, and a greater level of scrutiny should be applied regarding the inclusion of definitions in the report.

The lack of attention to definitions was also coupled with the absence of theoretical models or pedagogical approaches pertaining to technology use. Only one meta-analysis conducted by Rosen and Salomon (2007) addressed technology use from a constructivist perspective focusing on technology-intensive learning environments. Such a deficiency would have been acceptable if there was a shortage of theoretical frameworks in this area; however, this is definitely not the case. If we examine only the third edition of *The Handbook of Research on Educational Communications and Technology* (Spector, Merrill, Van Merrienboer, & Driscoll, 2008) we find that the editors dedicate a section that includes seven chapters focusing on instructional and learning strategies. Another section that includes 11 chapters dedicated to models with a focus on issues pertaining to various theoretical frameworks and approaches to learning and how they relate to instructional design and technology use in different educational contexts. Examples include generative learning, inquiry learning, collaborative activity, cooperation and technology use, cognitive apprenticeship model, problem-based learning, and resource based learning, to name just a few. Among other influential frameworks is the learner-centered approach for e-learning advocated by McCombs and Vakili (2005) based on the APA learner-centered principles (APA, 1997). With such an abundance of frameworks and approaches, one would have expected that they would be taken into consideration in framing the research questions to be answered by the meta-analyses, or during the design of the codebooks.

Moreover, keeping in mind that the technology debate is centered on pedagogical uses of technology (Clark, 1994; Kozma, 1994) and the confounding issues between the two, and with the continuous calls by researchers to focus on the pedagogical aspects of technology use, one would have expected to find some reference to *how* technology is utilized rather than *if* technology is present. The absence of any reference to the pedagogical uses of technology made it impossible to take such aspects into consideration in this second-order meta-analysis. Therefore the call for focusing on instructional design aspects of technology use in empirical research equally applies to meta-analyses.

A similar lack of attendance to theoretical frameworks was found in the meta-analyses addressing organizational behavior and human resource management by Steiner et al. (1991). This is another indication that higher emphasis is given to the numerical value of the effect size rather than investigating or validating theoretical models or frameworks, which has been highlighted by researchers such as Bangert-Drowns (1995) as one of the misunderstandings about meta-analysis. Particularly in our field, there is a strong need for meta-analysts to move beyond the focus on the average effect size and start working on testing theoretical frameworks and models of technology integration to help in the advancement of the field.

Second: Methodological quality indexes

The overall methodological quality index reflects the variability in the quality of the included meta-analyses where 10 were ranked as strong, 23 as moderate, and 4 as weak. Although overall, there seems to be an adequate number of strong meta-analyses, a more careful and critical analysis of the codes provides some good news and some bad news.

The good news is that, in general, most of the meta-analyses seem to be adequate and strong regarding their comprehensiveness. In particular, 64.9% of the included meta-analyses were considered strong in this area. The bad news, on the other hand, is that the majority of the meta-analyses were judged as weak on the rigour aspects. Specifically, 86.5% of the included meta-analyses were deemed weak in this area. This is rather alarming because it brings to question the reliability of the findings presented in a given meta-analyses.

The correlation analyses between publication date and methodological quality revealed that there is a significant moderate positive relationship between the publication date and comprehensiveness. This may be explained by the fact that with time the body of literature is increasing and demanding a higher level of adequacy in the search process and a more inclusive approach to the variety of available resources. A more reasonable explanation has to do with the development in technology that allows for the design and implementation of more adequate and far reaching search and retrieval strategies. The technological tools currently available for knowledge retrieval make the variety of processes entailed with searching, locating, and retrieving studies less demanding than the situation in the 1980's. With the general trend for technology growth and development, it appears that the process is guaranteed to improve further with time.

However, there was no significant relationship between time and methodological quality pertaining to rigour aspects which is in agreement with Steiner et al.'s findings. With all the advancements that the meta-analytic approach has witnessed, one would have expected a certain level of improvement over time concerning procedural aspects which is not reflected in the findings from the current analysis.

Furthermore, and unfortunately, the methodological quality indexes did not reveal any meta-analysis that may be considered the gold standard that we strive for. Although 10 were considered strong, none received a perfect score of 14. Had there been any of such a standard one may argue that the corresponding results would be most reliable and informative, however, this was not the case. Moreover, even if we want to consider the strongest ones that scored 12 out of 14, we will find that they are relatively focused and not comprehensive, and therefore not capable of giving recommendations or implications beyond the specific populations that they are targeting. Particularly speaking, Timmerman & Kruepke (2006) focused on computer assisted instruction with post-secondary students only, Onuoha (2007) focused on computer-based laboratory in science instruction at the secondary and post-secondary levels, and Goldberg et al. (2003) focused on the use of word processing with K-12 students.

Third: Specific methodological quality features

Upon checking specific features more information regarding the methodological quality of the meta-analyses is revealed. Regarding the search phase, the majority of the authors seem to cover the basic ground regarding the most important approaches such as database searches, branching, and hand searches while reporting their methodology. The major databases such as ERIC, Dissertation Abstracts International, and PsycINFO are listed in a large number of the included meta-analyses. Some of the aspects that were overlooked in a relatively small number of meta-analyses included the search time frame and its justification, and the terms used in the search strategy. Nevertheless, in most cases, a reader of most of these meta-analyses would be generally aware of what was done during the search phase. This also applies to the case of the included literature and research designs.

A disturbing observation was that a number of researchers in the field tended to confuse quasi-experimental and pre-experimental research designs. On more than one occasion, researchers specifically noted that their meta-analysis included RCTs and quasi-experimental designs while their study features and effect size calculation formulae clearly indicated that pre-experimental designs were included. This should be considered a serious mistake, and may be a reflection or indication of an underlying weakness in the programs preparing educational researchers. In our field of study, researchers' ability to conduct randomized control trials is highly limited by a variety of constraints. Hence we are fully aware that often there is a need to resort to quasi-experimental designs and on some occasions even to pre-experimental ones. However, this does not justify not knowing the difference between pre-experimental and quasi-experimental designs particularly since the two differ highly regarding threats to internal validity. Moreover, ever since the publication of Campbell and Stanley's 1963 work on experimental and quasi-experimental designs, there has been no shortage in literature and guidance on this topic (Campbell & Stanley, 1963; Shadish, Cook, & Campbell, 2002).

Another disturbing observation that relates highly to the weak methodological score on the rigour aspect has to do with the review of studies for inclusion in a meta-analysis and the effect size extraction procedures. In both cases, none of the included meta-analyses provided information reflecting that these two steps were conducted by independent researchers with the provision of inter-rater agreement levels.

Particularly, 32 meta-analyses gave no reference to the review process at all, with 4 indicating that one researcher reviewed the documents, while only two meta-analyses had the review or rating process conducted by more than one researcher. Although the reviewer usually uses the set of inclusion and exclusion criteria, this is not an adequate approach. Since the process may be influenced by the reviewer's own personal biases as well susceptibility to random human error, the best way to overcome it is through having multiple reviewers. This is true even with the most specific and clear set of inclusion and exclusion criteria due to the nature of the studies reviewed where each one presents a new situation. As a researcher who has been involved in a variety of meta-analyses, and based on first-hand experience with such review procedures, I can confidently say that no inclusion/exclusion list can eliminate personal biases or prevent unintentional mistakes. It is only through working independently with another researcher to resolve discrepancies one can ensure acceptable reliability of the review or coding processes.

The picture was not much brighter with the effect size extraction where also 32 meta-analyses gave no reference to the extraction process at all. In three meta-analyses it was indicated that one researcher extracted the effect sizes and in another three that the extraction process was conducted by more than one researcher. In this situation, personal bias tends to have an impact, especially when a given study includes multiple comparisons and there is a need to decide on which one or ones to include and which to ignore. Another problem that a researcher may risk by limiting the extraction phase to one researcher has to do with potential calculation errors of the effect size that may go unnoticed if not compared with another researcher who is working independently. This also brings to question the reliability of the findings in a given meta-analysis.

Nevertheless, there was one extraction process that received attention by a number of researchers and that was the study feature extraction. For this phase, eleven of the included meta-analyses reported that it was implemented by more than one researcher while seven reported the same with the provision of inter-rater agreement. Still, in three meta-analyses the study feature extraction was done by one researcher while 16 meta-analyses did not offer any information about the process. This is another aspect that reflects negatively on the quality of a meta-analysis particularly its reliability. The situation is aggravated by the facts that this phase can be highly subjective and that a lot of information is missing in many primary studies. Unconsciously, a reviewer tends to assume certain things when extracting study features from a given primary study under the influence of his/her own background and understanding of a given area. Also there is a high chance of missing out on information during this phase. Therefore having more than one reviewer extracting study features is essential for ensuring the adequate quality of any meta-analysis.

Failure to address rigour with multiple reviewers for more than one phase may have two explanations. The first could be related to the researchers' lack of awareness about the importance of reliability in coding, and the best methods of minimizing bias in meta-analytic procedures and ensuring a high level of reliability. There could be the misconception or myth that a meta-analysis is inherently objective and unbiased due to its quantitative nature which is not true. Similar to all research methodologies and statistical approaches, proper implementation is one of key aspect to its success. If this is the case, then a higher level of clarity in the training for meta-analytic procedures and the dissemination of standards is required. More emphasis should be placed on the role of inter-rater agreement at the different phases in a meta-analysis and in presenting meta-analytic procedures whether in researcher or student training.

Another reason for this could be linked to the need for extra man power and resources which entails a certain financial commitment which might not be easily achievable for many researchers. If this is the case, there are some less costly solutions such as conducting the review and extraction procedures by more than one researcher for a sample of the documents and reporting the inter-rater agreement for that sample. Such an approach will serve as a pilot run which will help establish the inter-rater agreement and allow the researchers to address and discuss some of their personal biases and determine the best way of approaching and overcoming them.

Analysis Phase

The most used type of effect size was Glass's Δ with 17 meta-analyses using it, followed by Hedges' g with 11 meta-analyses. Cohen's d was used in four meta-analyses. As for analysing effects, the coding procedure revealed that homogeneity of variance was implemented with 32.4% of the meta-analyses while 37.8% conducted various moderator analyses and meta-regression analyses were performed with only 5.3%. This is rather understandable keeping in mind that statistical procedures for the calculation of the effect size have changed and developed over time and that there are different theoretical underpinnings for each type of calculation which would have its own advocates.

However, specific study features addressing the analysis phase reflected certain report-related concerns and some more significant problems related to statistical procedures.

First: Effect size type

Among the 37 included meta-analyses six neither reported the type of effect sizes used nor provided any indication that could help in inferring this information. What is especially problematic is the fact that out of these six meta-analyses, four were published in peer reviewed journals. It is alarming to find that a number of meta-analyses passed through the peer review process and reached publication level with the absence of some basic information such as the type of effect size used not being noticed.

Second: Independence of data

The analysis revealed that almost half of the included meta-analyses addressed the independence of data issue, with slightly more than half having not resolved it. As for weighting by sample size, 15 of the included meta-analyses comprising 39.5% of the included set used this approach in their calculation.

However, an unexpected and alarming finding had to do with a controversial method of weighting by number of comparisons which was used in eight of the included meta-analyses. According to the authors, this was done to overcome the predicament of studies having higher weights due to extracting effect sizes from multiple non-independent comparisons from the same study such as Waxman, Lin, and Michko's meta-analysis (2003). It is important to note here that these were not cases were multiple groups were compared leading to some dependency if the same group was used in more than one comparison. These were cases where effect sizes were calculated from any set of separate results (grouped by different variables such as race, grade level, gender, ability) that were reported in a given primary study. Specifically, Waxman, Lin, and Michko (2003) extracted 27 different effect sizes from a single study (with a total of 282 effect sizes being calculated from 42 studies). Another example is the meta-analysis conducted by Niemiec, Samson, Weinstein, and Walberg (1987) where 224 effect sizes from 48 primary studies. The literature does not reflect any support for this methodology which aggravates the dependency issue even if it solves the problem of overweighting some studies. It is also important to note that in all of these meta-analyses there was no reference that would support such an approach of weighting. This finding once again brings to question the review process which did not succeed in capturing such an unsupported methodology, and heightens the need to work harder on disseminating and promoting appropriate meta-analytic procedures.

Third: Standard errors

There was the issue related to missing information particularly regarding standard errors associated with the reported effect sizes. While coding, it was found that eight meta-analyses reported the effect sizes with the corresponding standard errors and 15 meta-analyses offered some information that allowed for the calculation of the standard error. In contrast, 15 meta-analyses provided an average effect size while offering no information whatsoever that could help in calculating the standard error. Another finding had to do with the number of participants included in each meta-analysis. The coding procedure revealed that 15 meta-analyses representing 39.5% of the included set did not offer any information about the number of participants.

This failing may be explained by the fact that meta-analysis requires a set of skills that are not common among many researchers in the social science area who may be serving as reviewers on review boards of some educational journals. To overcome this shortcoming, journal editors need to ensure that reviewers who deal with meta-analyses are skilled in the field, which calls for more focus to be given to meta-analytic techniques in researcher training. Nonetheless, one finds that such faults are not limited to the peer review of meta-analyses. The close inspection of a random sample of published research studies will reveal a wide assortment of flawed analyses or missing information that succeeded in passing through the review process without being detected. A practical approach that could be easily implemented without having to resort to long term training is providing the reviewers with a checklist or set of guidelines, similar to the one used in this second-order meta-analysis, that allows for a more straightforward and standardized review process of meta-analyses.

Further Reporting Aspects

The codebook included a few other study features that went beyond contextual and methodological features and analyses, namely the provision of a list of included studies and a list of effect sizes, and the time period between the last included study and the publication date.

The list of included studies was not a problem where all the included meta-analyses provided one. However, on more than one occasion where a meta-analysis included achievement outcomes in addition to other outcomes such as attitudes, the authors did not specify which studies were relevant to each outcome. As for the table or list of individual effect sizes from the different meta-analyses, it was offered by 31 meta-analyses representing 81.6% of the included set which is rather remarkable.

Considering the time period between the last included study and the publication date which reflects the contemporariness of a given meta-analyses, it ranged between zero and three for 26 of the included meta-analyses. Considering the time frame for peer review procedures, this is neither surprising nor inadequate. What is alarming, however, is the finding that six meta-analyses had a time period of four years and another six had a time period of five years. Even with the delays one may encounter with the formalities of peer review and publication procedures, a four or five year time delay is highly unacceptable particularly in an area of study that is continuously changing and evolving such as technology integration. This speaks to both the quality of the meta-analyses and the peer review process. A meta-analysis that is four years old may not be reflective of the current situation and should be updated before it appears in a particular educational journal.

Summarizing the critical examination of the included meta-analyses, findings reflected a high level of variability in the application of the meta-analytic procedures, indicating a few strengths but revealing various weaknesses that need attention. Overall, the present findings regarding the methodological quality of the meta-analyses addressing technology integration in educational contexts are in agreement with those reached are by Steiner et al. in their empirical assessment of meta-analyses in organizational behavior human resources management (1991). Although their work was not a second-order meta-analysis from the perspective of synthesizing findings, they systematically analysed the methodological quality of the meta-analyses addressing the research question of interest. Many of our particular findings pertaining to the methodological quality of the included meta-analyses are highly comparable to their results. Our findings are in line with theirs in a different area of interest; which is rather alarming due to the fact that after more than 15 years, the picture regarding the methodological quality of meta-analyses has not changed drastically. Such a finding is an indication of the dire situation we are in vis-à-vis the implementation of the meta-analytic procedures.

Overlap in Primary Literature

One of the phases of the proposed meta-analysis procedure entailed the examination of the overlap in the primary literature included in the reviewed meta-analyses. This was one of the most demanding stages in the procedure due to the massive number of primary studies to be dealt with. After compiling the lists of primary studies from the 37 meta-analyses and cancelling duplicates, the overall number of different primary studies that appeared in one or more meta-analysis was 1253.

The process of compiling the list, resolving overlap, and deciding on the set of meta-analyses with minimal overlap, revealed some aspects about the meta-analyses that are worth noting and reflecting upon. First of all, it helped in the detection of a variety of documentation errors. In most of the cases the mistakes were discovered through detecting various discrepancies in referencing of the same study in different meta-analyses. For example, a dissertation by Ash (1985) appeared in two meta-analyses (Kulik & Kulik, 1991; Ryan, 1991) and was improperly dated as a 1986 dissertation in the Kulik and Kulik meta-analysis. Unfortunately, this particular type of mistake was noted with at least 3 primary studies in the Kulik and Kulik (1991) meta-analysis, and with a variety of other studies in other meta-analyses. Researchers should attend closely to the referencing task to minimize such mistakes as much as possible.

On a different note, one finding that was highly interesting had to do with three meta-analyses conducted by Liao and colleagues (Liao & Chen, 2005; Liao, Chang, & Chen, 2008; Liao, 2007) and addressing the impact of technology on student achievement in Taiwan. Although there was a certain level of overlap in their primary studies, eventually leading to the exclusion of one of the three from the final set of unique studies with minimal overlap, none of the studies appeared in any of the other meta-analyses (97 different studies conducted in Taiwan did not appear in any other meta-analysis). This finding is indicative of the immeasurable body of literature that is out of our reach due to language barriers and accessibility issues. If in one country there were 97 studies that might have been relevant to the technology integration question, one may confidently assume that there is a substantial number of studies that are conducted in other countries and are not available to us. The gravity of this issue is highlighted by the fact that out of the 37 included meta-analyses, only one had explicit reference to including studies published in a language other than English. The meta-analysis conducted by Pearson et al. (2005) included English and Spanish primary studies.

This aspect relates to the comprehensiveness of a given meta-analysis and the generalizability of its findings. However, there is no easy or practical solution for this problem. Any attempt at fully resolving it would need human resources and financial commitments that are unrealistic and beyond any review team's reach. A team would have to include skilful members with various language backgrounds in order to access the different databases from different countries, and review them reliably, which might not be easily attainable. One solution may be to try and incorporate some of the literature from other languages but this should be supported by a conceptual reasoning for the selection of the language not just merely due to the presence of a team member who is skilled in a particular language. Another solution is based on the hope that a representative sample of studies from around the world gets to find its way to English peer reviewed journals. However, with this we are at risk of increasing the chances of contributing to publication bias and over-relying on significant findings only. With the current givens, it seems that this is one of the shortcomings that we have to deal with and accept as not being able to overcome.

Synthesis of Effect Sizes

Besides critically examining the included meta-analyses, this study aimed at synthesizing their findings and explaining the variability in the effect sizes if possible. The particular research questions were:

1. Does technology use enhance student achievement in formal face-to-face classroom settings as compared to traditional settings? If so, to what extent?
2. What features, if any, moderate the overall effects of technology use on students' achievement?

Average Effect Size

To answer the research question, the proposed second-order meta-analysis was implemented and due to the novelty of the proposed methodology, two approaches were used, one based on sample-size standard error and one based on number-of-studies standard error. Furthermore, to validate the findings the weighted average effect size was also calculated from 574 individual effect sizes and their corresponding sample sizes extracted from 13 unique meta-analyses.

Results offer a clear answer to the first question and provide some insights for the second. Findings of the synthesis for the 38 effect sizes from the overall 37 included meta-analyses and for the 25 effect sizes from the set of 25 meta-analyses with minimal overlap which were considered to be unique provide an answer to the first question and settle the controversy about whether technology is helpful or not. In both cases, the weighted mean effect size was significantly different from zero. What makes the findings all the more trust-worthy and reliable is the fact that both approaches, namely the sample-size standard error and the number-of-studies standard error produced virtually the same results (with the overall and unique set of effect sizes). In addition, the average effect size calculated directly from available original primary studies' effect sizes validated the results further by offering similar findings. With all three approaches the point estimate is in the order of 0.3 while being significantly different from zero at the $p < 0.01$ level as evident in Table 48. These findings indicate an average effect size of medium strength according to Cohen (1988), favouring the utilization of technology in the experimental condition over traditional instruction in the control group.

Table 48. Weighted average effect sizes, standard errors, and confidence intervals

Approach	*k*	*g+*	*SE*	95% *CI*
Sample-size SE	38	0.330*	0.006	0.316/0.342
	25	0.315*	0.000	0.297/0.333
Number-of-studies SE	38	0. 343*	0.036	0.273/0.414
	25	0.333*	0.042	0.250/0.415
Raw effect size	574	0. 304*	0.008	0.288/0.320

$^{*}p < 0.01$

Such findings are overwhelming by the sheer number of primary studies that are incorporated in them, and the large number of participants that are thus included. With the synthesis of the effect sizes from the whole set of 37 meta-analyses, the results are amalgamating outcomes from 1253 primary studies with an overall sample size of approximately 130,300 participants. To stay on a more conservative side (i.e. to avoid counting participants more than once) the set of 25 meta-analyses that were considered to be unique are synthesizing results from 1055 primary studies with an overall sample size of approximately 109,700 participants. As for the calculation from the raw effect size scores, provided by meta-analysts, it is a rapid emulation of a regular meta-analysis with 574 effect sizes and an overall sample size of 60,853 participants.

Having included all meta-analyses with different time frames and targeting different technologies, one may say that the findings seem to disqualify the potential explanation which attributes technology's positive impact to the novelty aspect. It is obvious with our findings that technology in the classroom has passed or endured the test of time with a moderate positive average effect size.

One needs to keep in mind the population which the average effect size may be generalizable to. It is limited by the boundaries of the inclusion/exclusion criteria, and the findings are applicable to comparisons between technology enhanced classrooms versus more traditional settings where technology is not used to enhance the learning process. The earlier meta-analyses may have included studies with the control group being completely technology-free which was possible in the 1980s. However, the more recent meta-analyses cannot claim to compare technology enhanced classrooms with technology free ones, since most classrooms are currently equipped with some technological tools or students are provided access to computer labs. There, the comparisons were rather between experimental settings that were using technology as an active part within the instructional design of the targeted course and a control group where technology was not used to enhance the learning process. Moreover, the findings are limited to regular formal educational contexts and do not apply to situations with special needs or exceptional students, nor to on-the-job training or ongoing professional development

Homogeneity and Moderator Analyses

Similar to a regular meta-analysis, the average effect size is not the only focus or objective of the synthesis procedure. The variability in the effect sizes and the variables that moderate it are also highly important in allowing for a more thorough answer to the research question at hand. Unfortunately, currently and due to the novelty of the approach, the findings in this regard are not conclusive. Although the two implemented approaches, namely the sample-size standard error, and the number-of-studies sample error gave very consistent results regarding to the point estimate and its significance level, the tests of homogeneity with the two approaches, as expected, produced drastically different results as evident in Table 49.

Table 49. Weighted average effect sizes with homogeneity statistics

Approach	k	$g+$	Q-value	I^2
Sample-size SE	38	0.330*	202.28*	81.708
	25	0.315*	142.881*	83.203
Number-of-studies SE	38	0. 343*	9.864	0.000
	25	0.333*	8.534	0.000
Raw effect size	574	0. 304*	2927.869*	80.429

*$p< 0.01$

While the sample-size standard error approach indicates heterogeneity, the number-of-studies approach reveals full homogeneity; however each could only be interpreted within the context of its source of variance. The sample-size approach reflects the variance at the individual participants level (more than 100,000 participants) while the number-of-studies approach reflects the variance at the meta-analysis level (approximately 1000 studies). It should also be considered in relation to the type of study features that are being addressed by the moderator analysis, namely study level contextual features, more reflective of variability in population, or meta-analysis methodological features associated with each review procedural aspects.

The results of the raw effect size calculation with the heterogeneous outcome, offers backup for the sample-size approach; however, it does not offer conceptual and theoretical support for either one of the two approaches. Due to the heterogeneity in the results from the first approach and not the second one, the moderator analyses were only conducted with the findings from the sample-size standard error approach. It is important to stress that we need to be very cautious in drawing conclusions from these moderator analyses, and consider the results based on them as indicators of potential moderator variables.

The perfect validation would have been conducting moderator analyses with the results from the 574 individual effect sizes, but the missing information made it impossible. A solution would have been to contact the authors and collect their own data files to allow for the conduction of homogeneity testing and moderator analyses with the full set of effect sizes, however, this does not seem to be a practical or a feasible approach. Even if one wants to be optimistic and expect a high level of cooperation from all the authors, there are two main problems that will make the process highly challenging. For one, some of the meta-analyses were published too long ago and there is no guarantee that the authors still have their data files. On the other hand, the codebooks of the different meta-analyses are so varied, that it will not be easy to compile them into one common file and allow for adequate analysis. Therefore, for the purpose of the current project, this option was not followed through.

With the current findings, it seems that the second-order meta-analysis has proven its potential for synthesizing effect sizes and estimating the average effect size in relation to a specific phenomenon. However, it may not be capable of offering adequate answers for the homogeneity issue. This will be discussed further in the strengths and limitations section. Nevertheless, the findings from the moderator analyses pertaining to the results do offer some indication of variables that seem to have a relationship with the effect size.

First: Substantive contextual features

Unfortunately, as much as one would like to answer particular substantive questions regarding which technologies work best and under which conditions; currently, a second-order meta-analysis does not seem to be capable of answering them. After all, when conducting a second-order meta-analysis, a researcher is highly distant from the data and contextual aspects of the primary studies. This is mostly the case when a researcher is conducting a review of any sort, and it gets to be amplified with a second-order meta-analysis since the main source of information is a meta-analysis that has already conducted a selective filtration of information from the included primary studies. In other words, with a second-order meta-analysis, we are introducing another degree of separation from the original data. What makes it more challenging is the fact that not all meta-analysts report similar information or study features thus making the available pieces of information more of a mosaic that was not planned by an individual designer. Moreover, the reviewer is limited by the quality of the reports and the information provided in them. With the literature at hand, the coding procedure revealed the poor quality of the reports and the amounts of missing information.

The main purpose of the extraction of the specific effect sizes was to overcome the absence of specific contextual features in all the codebooks, and provide insights pertaining to type of technology, grade level, and subject matter. Unfortunately, due to missing information it was not as informative as expected. Nevertheless, a quick overview of the specific effect size values reveals that they reflect findings similar to the overall effect size.

The only significant finding from the conducted moderator analyses was the higher effect size with secondary level students ($g^+ = 0.448$, $k = 2$) than with elementary ($g^+ = 0.267$, $k = 2$) or postsecondary ($g^+ = 0.269$, $k = 8$) students. Still, this should be interpreted cautiously due the heterogeneity of the effects and the uneven number of meta-analyses pertaining to each grade level and therefore drawing solid implications does not seem to be appropriate.

Second: Methodological quality

The moderator analysis was very helpful in providing insights pertaining to the methodological quality and its relationship to the effect size. Results revealed that weak and moderate methodological quality meta-analyses tend to significantly overestimate the average effect size as compared with strong methodological quality meta-analyses. The point estimate for low quality meta-analyses was 0.364 ($k = 15$) and 0.273 ($k = 10$) for high quality ones. When moderator analyses were run for the indexes reflecting comprehensiveness and rigour, findings reflected that there was no significant difference between effect sizes from low and high meta-analyses regarding comprehensiveness.

However, there was a significant difference when the rigour index was used with the weak methodological quality resulting in an overestimation of the mean effect size in comparison to strong methodological quality. The point estimate for low quality meta-analyses was 0.334 (k = 20) and 0.275 (k =5) for high quality ones.

A possible implication of this finding is that meta-analysis consumers should be cautious with findings from weak and moderate methodological quality meta-analyses. This requires a higher level of awareness about what a strong meta-analysis should entail. There is a need to provide policy makers, administrators, researchers, and reviewers who use meta-analytic findings with guidelines and standards for evaluating the quality of a given meta-analysis. This is extremely important, since though lately we have observed a sweeping increase in proliferation of meta-analysis, it is evident from their quality that there is still a need for a higher understanding of its procedures. This could be targeted through a variety of dissemination venues including academic courses, training workshops, and conference sessions.

Another possible implication is mainly for meta-analysts who should work harder on the methodological quality of their systematic reviews, particularly regarding rigour. As reflected with the current findings, methodological quality seems to have an impact on the magnitude of the effect size. In light of the other findings in this study in relation to the methodological quality of the conducted meta-analyses, particularly concerning rigour, there seems to be a need to put extra efforts toward establishing adequate and up-to-standard practices. Again, this should be targeted through various venues.

Third: Publication bias

Other relevant findings have to do with publication bias. Moderator analysis for the type of publication of the meta-analyses did not indicate any significant difference between the weighted average effect sizes for journal published meta-analyses versus the non-published dissertations and reports. However, moderator analysis for the type of included literature pointed to a publication bias at the primary literature level. The meta-analyses which included only published empirical research had a significantly higher point estimate ($g^+ = 0.459$, $k = 5$) than those that included both published and unpublished research ($g^+ = 0.306$, $k = 20$). This is in agreement with other researchers' calls for including dissertations and conference proceedings to avoid the file drawer validity threat which refers to effect sizes being overestimated when only published primary studies are included (Sharpe, 1997).

Fifth: Types of effect sizes

Moderator analyses for the type of effect size provided a clear indication that *Cohen's d* and *Hedges' g* offer a smaller point estimate of 0.318 ($k = 13$) than *Glass's* point estimate of 0.384 ($k = 6$). This finding offers support for using *Cohen's d* and *Hedges' g* which take the variance of the experimental group into consideration over *Glass's* Δ which does not.

Summarizing the effect size synthesis, and homogeneity and moderator analyses, findings offered some general perspectives about the average effect size relating to technology's effect on students' achievement. The analysis did not offer specific insights about substantive and contextual features under which technology would be most beneficial however, it was rather informative regarding some meta-analytic methodological and procedural aspects and their relationship with the average effect size.

Strengths and Limitations

Similar to any other research endeavour, this project has its own strengths and limitations which will be addressed in this section.

Strengths

With the increasing number of published meta-analyses in a variety of areas the proposed second-order meta-analysis methodology allows for a more systematic and reliable methodology for synthesizing related results than a narrative review. Moreover, because it enables the synthesis of effect sizes from different meta-analyses while considering the standard errors, it is more adequate than vote counts. With such a methodology, researchers can benefit from published literature while reaching more generalizable findings than individual studies or regular meta-analyses can offer. This is particularly true regarding the effect size because of the larger included sample size.

Similar to the regular meta-analysis approach, one needs to keep in mind that this is only one form of literature review and synthesis and in no way is it capable of meeting each and every expectation of research integration (Bangert-Drowns, 1995). Rather it is one technique that may be helpful in certain contexts and situations. One of the strongest assets for the proposed methodology is its ability to help in answering big questions pertaining to a particular area of research with a considerable number of publicly available meta-analyses without the need to replicate their findings by running a huge new meta-analysis. Thus it will help in the reduction of time constraints, and minimize the costs.

Computer Technology and Student Achievement

Particularly speaking, results from the current study help in settling the controversy around technology's impact on students' achievement in formal educational settings. This second-order meta-analysis synthesized 1055 primary-studies which include approximately 109,700 participants, and after validating it with synthesizing 574 individual effect sizes that include 60,853 participants. Based on its findings we now know that the average effect size is in the order of 0.3. This means that the mean in the experimental condition will be at the 62nd percentile relative to the control group. In other words, this means that the average student in a classroom where technology is used to supplement face-to-face instruction will perform 12 percentile points higher than the average student in the traditional setting that does not use technology to enhance the learning process.

With this we can confidently say that technology does have an impact although we are not able to specify exactly how this impact is achieved, why, and who would benefit the most out of its use. Answering such questions will help us understand further how it works to make sure we gain the best out of what technology has to offer and get highest return on investment. Furthermore, it will help in the design of more adequate learning environments with the tools that are currently available in almost all the classes where the traditional classroom is becoming more and more technology enhanced. Unfortunately a second-order meta-analysis does not offer the answer the question of how and under what conditions technology works the best. This is partially due to the detachment from the primary data as well as the inadequacy of many meta-analytic reports.

Second-Order Meta-Analysis Procedural Aspects

Besides answering the "how much" question related to technology's impact on students' achievement, the major contributions of the current study regarding the proposed and implemented methodology are: a) design and implementation of a set of systematic steps allowing for the synthesis of the effect sizes from various meta-analyses; b) attendance to the methodological quality of the included meta-analyses; c) validation of the point estimate through the calculation with the raw effect sizes; and d) homogeneity testing and the attempt at explaining the variability in the effect sizes with both contextual and methodological quality features. Although researchers have previously experimented with quantitative syntheses of meta-analyses, there is no systematic and standard approach for the implementation of a second-order meta-analyses. Also, methodological quality and the explanation of the variability were not addressed before, and none of the previous attempts had worked on validating the process with other forms of calculations.

First: Systematic procedure

The proposed methodology is a highly systematic one that helps to minimize the level of subjectivity that may be entailed in a narrative review of a set of meta-analyses in a given area of research. By applying the systematic approach and using various checks such as inter-rater agreement and data spot checks a high level of reliability was achieved.

Second: Methodological quality

The approach used to assess the methodological quality for the purpose of this second-order meta-analysis has its own strengths. The methodological quality index was based on the most prominent guidelines and procedural perspectives for conducting a meta-analysis and is inclusive of the different phases of conducting a meta-analysis (e.g. Lipsey & Wilson, 2000). It is simple to implement with the specific descriptions included in the codebook allowing for easy coding. It can be used in the form of a continuous score or scale score and may be easily converted to a categorical score representing weak, moderate or strong methodological quality. Another of its advantage is the ability to check two main aspects, namely the comprehensiveness and rigour of a given meta-analysis. Finally, this approach may serve as the foundation for the development and establishment of a methodological quality tool that may be used by researchers, policy makers, producers and consumers of meta-analyses, as well as academic journal reviewers for the assessment of any given meta-analysis in the social science area.

Although various researchers have called for more adequate implementation of meta-analytic procedures this is the first extensive and in-depth analysis of the methodological quality of such a substantial number of meta-analyses addressing one specific area of study. Moreover, the critical evaluation of the included meta-analyses was very informative about the quality of the meta-analyses in the area.

Hopefully, the implementation of second-order meta-analyses in various research areas may prove to be a drive for the development and improvement of regular meta-analytic reports, similar to the impact that the latter had in relation to improving the quality of primary research reporting.

Third: The validation of the point estimate

The process of calculating the average effect size form the raw effect sizes provided validation for the results of the second-order meta-analysis and its ability to synthesize the effect sizes with the current literature. The heterogeneity results provided some backup for the sample-standard error approach but it did not offer conceptual and theoretical support for either one of the two approaches. The perfect validation would have been conducting moderator analyses with the results from the 574 individual effect sizes, but with the missing information this was impossible.

Fourth: Test of homogeneity and moderator analysis

Although neither the heterogeneity nor moderator analyses could offer conclusive evidence, they did provide some insights about features that may have a substantial influence on the magnitude of the effect sizes. The features that seem to have a moderating effect mainly address methodological quality aspects of the included-meta-analyses. Although the implications from these analyses are not generalizable, they offer support to currently available calls for greater attention to methodological quality and inclusivity in meta-analysis.

Limitations

First: Report quality and limited information

The most important limitation that one may encounter in a systematic review is that pertaining to the boundaries set by the amount of information provided in the documents under review and the quality of the report itself. With a second-order meta-analysis this issue is magnified and becomes the most challenging aspect facing the reviewer. While implementing the proposed methodology with meta-analyses targeting technology integration in formal educational contests and its impact on student achievement, we were highly limited by the information presented in a given meta-analysis, which was only aggravated by the quality of the reports themselves. This also led to restrictions concerning the features that could be extracted and subsequently to the analyses that could be conducted. This is somewhat expected when one is working with a review of reviews; after all, we are trying to synthesize a set of syntheses which should by nature be succinct and condensed regarding certain aspects pertaining to the original primary studies. However, the unexpected constraint had to do with the quality of the reports and the considerable amount and variety of missing information, unfortunately, including such basic facts and figures as number of participants and standard errors. This resulted in the need to depend on certain assumptions particularly in replacing missing values. If we did not have to resort to such measures, we would have been able to put more faith in the results. Poor reporting quality of meta-analyses has already started gaining attention and is being addressed by researchers in the field (e.g. Harwell & Yukiko, 2008).

Furthermore, and as a consequence of the inability to extract and code for a variety of contextual features, it was impossible to explore certain aspects pertinent to the use of technology. Indeed, the quality of the reports denied us an adequate chance to closely address major study features, namely type of technology, grade level, and subject matter. The fluidity and ambiguity of terms in the field did not make the prospect easier, and the difficulty was augmented by the absence of adequate definitions for a variety of the used terms. Although these limitations are not inherent in a second-order meta-analysis, they are substantial and their likelihood is high in this kind of review.

Second: Methodological quality index

To evaluate the methodological quality of the meta-analyses, an overall index was used addressing two main aspects, namely comprehensiveness and rigour. One of the limitations of this index was its inability to create more fine-tuned categories that would reflect more specific aspects of methodological quality. In part, this was the result of the rather large overlap between the different facets such as comprehensiveness, contemporariness, reliability accuracy, and conceptual adequacy. Had that been possible, moderator analyses might have offered more specific information regarding which aspects have a relationship with the effect size and which do not. Once again, as reviewers we are dependent on the quality of the report, and the evaluation of the methodological quality of a meta-analysis is not a pure assessment of the quality of what the meta-analysts actually did, rather it is highly confounded with the report quality.

Third: Vast amount of data

Another challenge faced in the implementation of the proposed methodology has to do with the huge amount of data and the need to be very cautious and organized while working with them to avoid errors. Similar to regular meta-analyses, this is true for the variety of phases, including the review process as well as the study feature and effect size extraction. With the current study, this also applied to the process of investigating the overlap in included primary studies to specify the set of meta-analyses with the minimal overlap. Particularly speaking, the Excel file in which the extracted data from the 37 included meta-analyses were compiled included 63 columns and 143 rows, while the Excel file in which the primary studies included in all 37 meta-analyses were compiled included 40 columns and 1253 rows. Handling the files, organizing the information, and keeping proper records may prove to be overwhelming. In the current situation, the principal investigator's previous experience with systematic reviews and handling similar files was very helpful in enabling the smooth progress in the project.

Fourth: Evolving nature of meta-analysis

Meta-analysis is a constantly evolving research tool and the developments it has witnessed over time (Schmidt, 2008), in addition to the various theoretical underpinnings for each approach and the opposing opinions regarding the more appropriate procedures has understandably resulted in an elevated level of variability in the implementation of meta-analytic procedures.. This leads to certain challenges similar to those faced in synthesizing findings from primary studies in a meta-analysis due to the dissimilar approaches, dependent measures, and variety of study features (Bangert-Drowns, 1986). Such variability makes the process of designing an adequate codebook to address this variability, and target the most relevant aspects, all the more challenging. With this study, the extensive review of various sources allowed for the design and development of a codebook that was capable of meeting the specified objectives and overcoming as many obstacles as possible.

Fifth: Inability to resolve variability issue

For the purpose of this second-order meta-analysis two approaches of testing homogeneity were used, namely the sample-size standard error and the number-of-studies standard error. Each has its strengths and weaknesses, but neither has a definitive conceptual or theoretical justification. The first approach makes use of the strength of meta-analyses and allows for reliable conclusions, while keeping the enormous variability in the study findings intact, thus magnifying the heterogeneity in the findings and enabling its thorough exploration. However, we might be running a higher risk of type I error and finding false positives. On the other hand, the second approach does not overstate the heterogeneity, but it ignores the actual strength offered by the individual point estimates from the different meta-analyses, and increases the chances of committing type II error and finding false negatives. The current inability to resolve this issue forces us to interpret the results from the moderator analyses conducted with the results from the sample-size standard error approach with caution, and consider them as indicators and not reliable moderators of the effect sizes.

Implications and Future Directions

Based on the findings of this research project various implications may be offered to stakeholders interested in meta-analysis as well as technology integration in academic settings. These will be presented in this section in addition to future directions and suggestions for research and practice.

First: Researchers interested in research synthesis

For researchers interested in making use of available literature through reviewing and synthesizing results from various meta-analyses in a given field, the proposed methodology and its implementation demonstrate that it is attainable. Similar to any other research endeavour, it has no safeguards against challenges. Nevertheless, this study has attempted to build on previous trials for conducting such syntheses while moving further in developing and establishing a more systematic approach. The most significant contributions, the present research has to offer, are the development of the methodological quality index and conducting moderator analyses in an attempt to explain the variance in the effect sizes. The methodology has proved to be adequate for answering big questions, but not highly suitable for offering details about specific contextual features. Within the educational sector, this could be of interest to policy makers and administrators who are expected to take informed decisions regarding a variety of issues. The implementation of a second-order meta-analysis offers reliable findings based on a substantive body of literature, in a timely manner, within a specific area of research.

However, if one is interested in answering specific questions pertaining to a given area of study, the second-order meta-analysis does not seem to be the most appropriate venue. Better alternatives may include conducting a regular meta-analysis to address the features of interest, which would allow the reviewers to be more confident in the overall quality of the procedure and moderator analysis findings. Another approach could be contacting authors of meta-analyses and requesting access to their files in order to run the analyses with the raw data and study features, however, the authors' cooperation is absolutely necessary for the success of this approach, still with no guarantees of overcoming the variability in the codebooks and effect size extraction easily. A third approach is to retrieve the available meta-analyses, identify the overall set of primary studies included in them and then extract effect sizes and study features in a consistent fashion allowing for reliable procedures and compatible analyses, thus attempting to replicate all preceding meta-analyses in one mega-review.

Regarding second-order meta-analyses, future developments may prove to be similar to those attained by regular meta-analysis over time. Meanwhile, short term future endeavours should aim at implementing the methodology with different areas of study, while trying to resolve the issue with analysis of variance in the effect sizes, and if possible, to include more contextual study features. However, this should be done with the support and joint efforts of other stakeholders. Finally, it is crucial to keep in mind the relative validity and value of the question being asked. If, for example, an administrator requires information on the effectiveness of laptop programs to inform a major pedagogical/financial decision, a second order meta-analysis such as the one carried out here would be inappropriate. In that case, only meta-analyses examining laptop programs should be used (assuming they exist).

Second: Producers and consumers of meta-analyses

The biggest challenges faced during the implementation of the proposed procedure resulted from the report quality of the included meta-analyses, particularly the missing procedural and statistical information, and the inability to extract more contextual features. Although this influenced what could be done in the second-order meta-analysis, the findings pertaining to the relationship between the methodological quality of a meta-analysis and its effect sizes imply that the repercussions are more far-reaching. It is unfortunate that so many weak and moderate quality meta-analyses are being published. To improve the situation, greater attention should be given by meta-analysis researchers, reviewers, and consumers to methodological quality aspects.

Researchers are calling for higher transparency, specificity and clarity in the reports to allow users to evaluate the quality of a given meta-analysis (Harwell & Yukiko, 2008). Moreover, there seems to be a need to develop researchers' meta-analytic skills in order to better prepare them for conducting their own meta-analyses or assessing others' meta-analyses for their own use or for publication purposes if they are acting as reviewers. This may be achieved through a variety of venues, one of which is designing and developing formal meta-analysis courses for graduate and post-graduate students, which already appears to be more and more common in different academic institutions and departments.

Another approach could be offering professional development workshops and training sessions targeting meta-analytic procedures for both producers and consumers of meta-analyses. In both, formal and informal settings, attention should be given to adequate procedural aspects and the importance of rigour. These approaches may help in improving the quality of published meta-analyses, but what could be even more influential is working on setting standards and guidelines for accepting a meta-analysis for publication. Moreover, meta-analysis users should be aware and familiar with the standards to help them in judging the quality and thus the reliability of the findings of any given meta-analysis. A methodological quality assessment tool would be extremely helpful in achieving such objectives. This was started in this study with the methodological quality index. Future efforts should aim at developing the approach further to design a tool that could be easily utilized by various users. The tool should be validated theoretically with the help of a panel of experts in the field and then it should be validated by practically applying it in different contexts and by various individual users. Potential advantages could include its routine utilization by various associations involved with the implementation of meta-analyses and the dissemination of their findings. One good example is the *Campbell Collaboration* that could make use of a methodological quality tool to support researchers in their meta-analytic work before and during the protocol registration process, and the reviewers while assessing its quality.

Third: Technology integration

Findings of the current synthesis indicate that technology integration has a positive impact on student achievement, although they do not offer insights about why and how. The latter was a direct result of the absence of information pertinent to such questions in the included meta-analyses. Reviewers and meta-analysts who are interested in technology integration and its relationship to student achievement need to shift attention from asking the "technology versus no technology" type of questions to pedagogical issues related to technology integration. Calls asking for such a focus have been very prominent in the literature (Clark, 1994; 2001; Laurillard, 2002), and primary research has started reflecting the response to such calls. However, this is not the case with meta-analytic reviews yet. Including pedagogical aspects of technology integration has to be addressed in meta-analyses not just for the purpose of conducting further second-order meta-analyses, but first and foremost to help in learning more about the active ingredients in technology integration that are benefiting students.

One very good example of using a meta-analysis to validate a theoretical framework is the one recently conducted by Bernard et al. (in press). It compares different distance education instructional conditions using Moore's theoretical framework (1989) for the three types of interaction, namely student-student; student-teacher and student-content to test Anderson's hypothesis (Anderson, 2003) about their relative contribution to learning success in distance education. In this meta-analysis, Bernard et al. go beyond the traditional question of how does distance education compare with traditional face-to-face instruction and focus on answering questions comparing distance education treatments among themselves.

Similarly, within the area of technology integration, the systematic review team at the *Centre for the Study of Learning and Performance* is moving beyond the technology/no technology question by conducting a meta-analysis to answer the following research questions:

> *What is the impact of the educational use of contemporary computer-based technologies on achievement and attitude outcomes of postsecondary students in formal educational settings? How do various pedagogical factors, especially the purpose and the amount of technology use, moderate this effect?*
>
> (Schmid et al., 2009)

Preliminary findings of this meta-analysis will be presented at the next annual *American Educational Research Association.*

In addition, the current review and synthesis, has revealed the absence of a comprehensive meta-analysis with a gold standard methodological quality. This finding in conjunction with the fact that the second order meta-analysis was not capable of answering specific questions related to technology, suggests the need for a high-standard comprehensive regular meta-analysis that will allow for more adequate moderator analyses.

Finally, future meta-analyses should start focusing on more contemporary uses of technology, particularly its online and e-learning applications, such as *Computer Mediated Communication.* According to the Horizon report published as a collaboration between the *New Media Consortium* and the *Educause Learning Initiative* (2008), key emerging technologies that are expected to enter mainstream use in teaching contexts in the near future are grassroots video, collaboration webs, mobile broadband, data mashups, collective intelligence, and social operating systems. Researchers are encouraged to be attentive to the changes taking place in technology usage in learning contexts including keeping an eye on the above-listed technologies.

Final Words

In conclusion, this study has attempted to develop and implement a more systematic and elaborate second-order meta-analysis methodology than previous trials. Results provided general insights about technology integration in educational contexts and offered substantive findings regarding the methodological quality of the included meta-analyses. The major findings from this project include:

- Based on the review and synthesis of a substantive body of literature, it was concluded that technology is helpful for students' achievement in regular formal educational contexts.
- Published meta-analyses addressing technology's impact on students' achievement, vary in methodological quality reflecting various shortcomings that include the different procedural aspects of implementation, with the majority being rather weak regarding rigour.
- Published meta-analyses addressing technology's impact on students' achievement almost completely overlook theoretical and pedagogical frameworks for successful technology use which does not allow for an in-depth understanding of how technology and pedagogy can be coordinated effectively to make the best use of the technology.
- The implementation of the second-order meta-analysis demonstrated the technique's ability to be an adequate and efficient tool for synthesizing effect sizes from various meta-analyses to offer overall answers to focused big questions.

However, currently it falls short on providing answers to specific questions related to study features.

Based on the findings, suggestions for future actions that need to be taken to understand how technology and pedagogy can be coordinated effectively to make the best out of what technology has to offer include the need for:

- Greater attention to definitions and theoretical frameworks in meta-analyses addressing technology integration.
- More focus on comparisons between different pedagogical uses of similar technologies.
- Addressing more contemporary uses of technology particularly the communication and collaboration tools.

Moreover, suggestions for future actions that are needed to improve the quality of meta-analyses in the educational contexts include greater attention to:

- Dissemination of adequate procedures, and proper training of researchers interested in conducting meta-analytic reviews.
- Various methodological aspects of meta-analyses, particularly those related to rigour and reliability throughout the various phases.
- The reporting quality of meta-analyses which should reflect a higher level of transparency to allow for adequate evaluation of the implemented procedures and the generalizability of the subsequent findings.
- Setting standards and guidelines for implementing meta-analyses and reporting their findings.

Finally, it is important to emphasize again that similar to the regular meta-analysis approach, we need to keep in mind that a second-order meta-analysis is only one form of literature review and synthesis and in no way is it capable of meeting all research integration expectations (Bangert-Drowns, 1995). It cannot and should not take the place of a regular meta-analysis the same way that a meta-analysis cannot and should not take the place of primary studies. It is an emerging technique designed to answer big questions. To answer specific questions it is better to depend on primary studies and regular meta-analyses. No one knows what the future of second-order meta-analysis will be, but there seems to be some evolutionary steps comparable to regular meta-analysis. It is starting to gain attention, although gradually, but with the continuing increase in the number of published met-analyses it will gain more attention and may prove to be an effective technique in certain situations. In the future, we may find better statistical approaches to testing homogeneity and explaining variability. Meanwhile we need to work on improving the tools we have and developing them further to make better use of the body of the ever expanding literature.

REFERENCES

References marked with an asterisk indicate studies included in the second-order meta-analysis

Abrami, P. C., Cohen, P. A., & D'Appollonia, S. (1988). Implementation problems in meta-analysis. *Review of Educational Research, 58*(2), 151-179.

Akyol, Z., & Cagiltay, K. (2007, June 25). *The media-method debate revisited.* Paper presented at the World Conference on Educational Multimedia, Hypermedia and Telecommunications Vancouver, Canada.

Alessi, S. M., & Trollip, S. R. (2000). *Multimedia for learning: Methods and development* (3rd ed.). New York: Allyn & Bacon.

American Educational Research Association (2008). *Review of Educational Research.* Retrieved November 15, 2008, from http://ojs.aera.net/journals/index.php/RER/about/editorialPolicies#focusAndScope

Anderson, T. (2003). Getting the mix right again: An updated and theoretical rationale for interaction. *International Review of Research in Open and Distance Learning. 4*(2), 9-14.

APA. (1997). Learner-centered psychology principles: A framework for school redesign and Reform. Retrieved 25 March, 2005, from http://www.apa.org/ed/lcp.html

Ash, R.C. (1985) Analysis of cost-effectiveness of a computer-assisted instruction program. *Dissertation Abstract International,* 46, 2137.

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*(2), 111-127.

Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin, 99*(3), 388-399.

Bangert-Drowns, R. L. (1995). Misunderstanding meta-analysis. *Evaluation and the Health Profession, 18*(3), 304-314.

*Bangert Drowns, R. L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research, 63*(1), 69-93.

Bangert Drowns, R. L., & Rudner, R. L. (1991). Meta-analysis in educational research. *Practical Assessment, research & Evaluation, 2*(8). Retrieved August 20, 2008, from http://pareonline.net/getvn.asp?v=2&n=8

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *Personality and Performance* (1/2), 9-30.

*Bayraktar, S. (2000). *A meta-analysis on the effectiveness of computer-assisted instruction in science education.* Unpublished doctoral dissertation, Ohio University.

Bayraktar, S. (2001). A meta-analysis of the effectiveness of computer-assisted instruction in science education. *Journal of Research on Technology in Education, 34.*

Becker, H. J. (2000). Pedagogical motivation for student computer use that lead to student engagement. *Educational Technology, 40*(5), 5-17.

Bernard, R.M., & Abrami, P.C. (2009). *Statistical applications in meta-analysis, extracting, synthesizing and exploring variability in effect sizes.* Unpublished manuscript, Concordia University, Montreal, Quebec, Canada.

Bernard, R.M., Abrami, P.C., Borokhovski, E., Wade, A.C., Tamim, R.M., Surkes, M.A., et al. (in press). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*

Bernard, R., Abrami, P. C., Lou, Y., & Borokhovski, E. (2004). A methodological morass? How we can improve quantitative research in distance education. *Distance Education, 25*(2), 175-198.

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., et al. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research, 74*(3), 379-439.

Bernard, R., & Naidu, S. (1990). Integrating research into instructional practice: The use and abuse of meta-analysis. *Canadian Journal of Educational Communication, 19*(3), 171-198.

Beuther, D. A., & Sutherland, E. R. (2007). Overweight, obesity, and incident asthma: A meta-analysis of prospective epidemiologic studies. *American Journal of Respiratory and Critical Care Medicine, 175*, 661-666.

Bill Gates quotes (2008). Retrieved November 10, 2008, from http://thinkexist.com/quotation/if_gm_had_kept_up_with_technology_like_the/188702.html

*Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research, 72*(1), 101-130.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (2005). *Comprehensive Metaanalysis Version 2*. Englewood, NJ: Biostat.

Bransford, J. D., Vye, N. J., & Bateman, H. (2002). Creating high-quality learning environments: Guidelines from research on how people learn. In P. A. Graham & N. G. Stacey (Eds.), *The knowledge economy and post secondary education: Report of a workshop*. Washington, DC: National Academy Press.

Butler, A., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review, 26*, 17-31.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. In N. L. Gage (Ed.), *Handbook of research on teaching*. Boston: Houghton Mifflin Company.

*Christmann, E. P., & Badgett, J. L. (2000a). The Comparative effectiveness of CAI on collegiate academic performance. *Journal of Computing in Higher Education, 11*(2), 91-103.

*Christmann, E. P., & Badgett, J. L. (2000b). *Progressive comparison of the effectiveness of computer-assisted instruction on science achievement: A meta-analysis.* Paper presented at the International Conference on Mathematics/Science Education and Technology San Diego, California.

*Christmann, E. P., & Badgett, J. L. (2003). A meta-analytic comparison of the effects of computer-assisted instruction on elementary students' academic achievement. *Information Technology in Childhood Education Annual, 15*(1), 91-104.

*Christmann, E. P., Badgett, J. L., & Lucking, R. (1997). Microcomputer-based computer-assisted instruction within differing subject areas: a statistical deduction. *Journal of Educational Computing Research, 16*(3), 281-296.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445-449.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21-29.

Clark, R. E. (2001). *Learning from media: Arguments, analysis, and evidence*. Greenwich, CT: Information Age Publishers Inc.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, P. A. (1990). Things I have learned (so far). *American Psychologist, 54*(12), 1304-1312.

*Cohen, P. A., & Dacanay, L. S. (1992). Computer-based instruction and health professions education: A meta-analysis of outcomes. *Evaluation and the Health Professions, 15*(3), 259-281.

Computer-assisted instruction (2008). In *CSharpOnline.NET*. Retrieved December 10, 2008, from http://en.csharp-online.net/Glossary:Definition_-_Computer-Assisted_Instruction.

Computer quotes (2008). Retrieved December 22, 2008, from http://www.gdargaud.net/

Humor/QuotesComputer.html

Cook, T. D., & Leviton, L. (1980). Reviewing the literature: A comparison of traditional methods with meta-analysis. *Journal of Personality, 48*, 449-472.

Cooper, H., & Arkin, R. (1981). On quantitative reviewing. *Journal of Personality, 49*, 225-230.

Cuban, L., Kirkpatrick, H., & Peck, C. (2001). High access and low use of technologies in high school classrooms: explaining an apparent paradox. *American Educational Research Journal, 38*, 813-834.

DeBell, M., & Chapman, C. (2006). *Computer and internet use by students in 2003.* Washington DC: U.S. Department of Education. (NCES 2006-065)

Egger, M., Juni, P., Bartlett, C., Holenstein, F., & Sterne, J. (2003). How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technology Assessment, 7*(1), 1-76.

Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist, 33*, 517.

Eysenck, H. J. (1995). Meta-analysis squared - Does it make sense. *American Psychologist, 50*, 110-111.

Ferguson, S. (2005). How computers make our kids stupid. *Maclean's, 118*(23), 24-30.

*Fletcher-Flinn, C. M., & Gravatt, B. (1995). The efficacy of computer assisted instruction (CAI): A meta-analysis. *Journal of Educational Computing Research, 12*(3), 219-242.

Fletcher, J. D. (1989). The effectiveness and cost of interactive videodisc instruction. *Machine-Mediated Learning, 3*(4), 361-385.

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher. 5*(10), 3-8.

Glass, G. V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education, 5.*

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Glass, G. V., & Smith, M. L. (1979). Meta-analysis of research on the relation-ship of class-size and achievement. *Evaluation and Policy Analysis, 1*, 2-16.

*Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992-2002. *The Journal of Technology, Learning, and Assessment, 2*(1), 3-51.

Grabe, M., & Sigler, E. (2002). Studying online: evaluation of an online study environment. *Computers and Education, 38*, 375-383.

Hammill, D. D. (2004). What we know about correlates of reading. *Exceptional Children, 70*(4), 453-469.

Harwell, M., & Yukiko, M. (2008). Deficiencies of reporting in meta-analyses and some remedies. *The Journal of Experimental Education, 76*(4), 403-428.

Health-Evidence (n.d.). *Quality assessment tool.* Retrieved December 22, 2008, from http://health-evidence.ca/downloads/QA%20tool_Doc%204.pdf

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hill, J. R., & Hannafin, M. (2001). Teaching and learning in digital environments: The resurgence of resource-based learning. *Educational Technology Research and Development, 49*(3), 37-52.

*Hsu, Y. C. (2003). *The effectiveness of computer-assisted instruction in statistics education: A meta-analysis.* Unpublished doctoral dissertation, University of Arizona.

Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. New York: Russell Sage Foundation.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis*. Beverly Hills, CA.

Internet World Stats. (2008a). *Internet usage and population in North America.* Retrieved 28 November, 2008, from http://www.internetworldstats.com/stats14.htm

Internet World Stats. (2008b). *World internet users and population stats.* Retrieved 28 November, 2008, from http://www.internetworldstats.com/stats.htm

Jacobson, M. (1998). Adoption patterns of faculty who integrate computer technology for teaching and learning in higher education. Retrieved 15 March, 2005, from http://www.ucalgary.ca/~dmjacobs/phd/phd-results.html

Jonassen, D. (2003). Using cognitive tools to represent problems. *Journal of Research on Technology in Education, 35*(3), 362-381.

Jonassen, D., Campbell, J., & Davidson, M. (1994). Learning with media: Restructuring the debate. *Educational Technology Research and Development, 42*(2), 7-19.

Jonassen, D., & Reeves, T. C. (1996). Learning with technology: Using computers as cognitive tools. In D. Jonassen (Ed.), *Handbook of Research for Educational Communications and Technology*: McMillan.

Joy, E. H., & Garcia, F. E. (2000). Measuring learning effectiveness: A new look at no-significant-difference findings. *Journal of Asynchronous Learning Networks, 4*(1), 33-39.

Kardas, P., Devine, S., Golembesky, A., & Roberts, C. (2005). A systematic review and meta-analysis of misuse of antibiotic therapies in the community. *International Journal of Antimicrobial Agents, 26*, 106-113.

*Khalili, A., & Shashaani, L. (1994). The effectiveness of computer applications: A meta-analysis. *Journal of Research on Computing in Education, 27*, 48-61.

Kimble, C. (1999). *The impact of technology on learning: Making sense of the research.* Aurora, CO: Mid-Continent Regional Educational Laboratory. Retrieved 25 March, 2005, from http://www.mcrel.org/PDF/PolicyBriefs/5983PI_PBImpactTechnology.pdf

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research.* Washington, DC: American Psychological Association.

Kohut, A., Wike, R., & Horowitz, J. M. (2007). *World publics welcome global trade - But not immigration.* Washington, DC: Pew Research Center.

*Koufogiannakis, D., & Wiebe, N. (2006). Effective methods for teaching information literacy skills to undergraduate students: A systematic review and meta-analysis. *Evidence Based Library and Information Practice, 1*(3), 3-43.

Kozma, R. (1991). Learning with media. *Review of Educational Research, 61*(2), 179-221.

Kozma, R. (1994). Will media influence learning: Reframing the debate. *Educational Technology Research and Development, 42*(2), 7-19.

*Kuchler, J. M. (1998). *The effectiveness of using computers to teach secondary school (grades 6-12) mathematics: A meta-analysis.* Unpublished doctoral dissertation, University of Massachusetts Lowell.

Kuh, G. D., & Vesper, N. (2001). Do computers enhance or detract from student learning? *Research in Higher Education, 42*(1), 87-102.

Kulik, C. L. C., & Kulik, J. A. (1986). Effectiveness of computer-based education in colleges. *AEDS Journal, 19*(2-3), 81-108.

*Kulik, C. L. C., & Kulik, J. A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior, 7*(1-2), 75-94.

Kulik, J. A. (1994). Meta-analytic studies of findings on computer-based instruction. In E. L. Baker & H. F. O'Neil Jr. (Eds.), *Technology assessment in education and training*. Hillsdale, NJ: Lawrence Erlbaum.

Kulik, J. A., & Kulik, C. L. C. (1987). Review of recent research literature on computer-based instruction. *Contemporary Educational Psychology, 12*(3), 222-230.

Laurillard, D. (2002). *Rethinking university teaching: A framework for the effective use of educational technology* (2nd ed.): London: Routledge.

*Lee, J. (1999). Effectiveness of computer-based instructional simulation: a meta analysis. *International Journal of Instructional Media, 26*(1), 71-85.

*LeJeune, J. V. (2002). *A meta-analysis of outcomes from the use of computer-simulated experiments in science education.* Texas A & M University

*Liao, Y.-I., & Chen, Y.-W. (2005). *Computer simulation and students' achievement in Taiwan: A meta-analysis.* Paper presented at the World Conference on Educational Multimedia, Hypermedia and Telecommunications, Montreal, Canada.

*Liao, Y.-K., Chang, H. W., & Chen, Y. W. (2008). Effects of computer applications on elementary school students' achievement: A meta-analysis of students in Taiwan. *Computers in the Schools, 24*(3), 43-64.

*Liao, Y. C. (1992). Effects of computer-assisted instruction on cognitive outcomes: a meta-analysis. *Journal of Research on Computing in Education, 24*(3), 367-380.

*Liao, Y. C. (1998). Effects of hypermedia versus traditional instruction on students' achievement: A meta-analysis. *Journal of Research on Computing in Education, 30*(4), 341-360.

Liao, Y. C. (1999a). *Gender differences on attitudes toward computers: A meta-analysis.* Paper presented at the Society for Information Technology and Teacher Education, Chesapeake, VA.

Liao, Y. C. (1999b). *Gender differences on computer anxiety: A meta-analysis.* Paper presented at the Society for Information Technology and Teacher Education, Chesapeake, VA.

Liao, Y. C. (2000). *A meta-analysis of gender differences on attitudes toward computers for studies using Loyd and Gressards' CAS.* Paper presented at the Educational Multimedia, Hypermedia and Telecommunications, Chesapeake, VA.

Liao, Y. K. C. (1999c). Effects of hypermedia on students' achievement: a meta-analysis. *Journal of Educational Multimedia and Hypermedia, 8*(3), 255-277.

*Liao, Y. K. C. (2007). Effects of computer-assisted instruction on students' achievement in Taiwan: A meta-analysis. *Computers and Education, 48*(2), 216-233.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.

Lipsey, M. W., & Wilson, B. G. (1995). Reply to comments on Lipsey and Wilson (1993). *American Psychologist, 50*, 113-115.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological educational, and behavioral treatment. *American Psychologist, 48*(12), 1181-1209.

Lipsey, M. W., & Wilson, D. B. (2000). *Practical meta-analysis* (Vol. 49). London: Sage Publications.

Lou, Y., Abrami, P. C., & D'Appollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71*, 449-521.

Lowerison, G., Sclater, J., Schmid, R. F., & Abrami, P. C. (2005). Are we using technology for learning. *Journal of Educational Technology Systems, 34*(4), 401-425.

Luborsky, L., Rosenthal, R., Diguer, L., Andrusyna, T. P., Berman, J. S., Levitt, J. T., et al. (2002). The dodo bird verdict is alive and well - Mostly. *American Psychological Association, 9*(1), 2-12.

Matt, G. E., & Cook, T. D. (1994). Threats to the validity of research synthesis. In H. Cooper & V. Hedges (Eds.), *The handbook of research synthesis*. New York: Russell Sage.

Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction, 13*, 125-139.

McCombs, B. L. (2000, September 11-12). *Assessing the role of educational technology in the teaching and learning process: A learner-centered perspective.* Paper presented at the Secretary's Conference on Educational Technology, Alexandria, VA. Retrieved September 5, 2005, from http://www.ed.gov/rschstat/eval/tech/techconf00/mccombs_paper.html

McCombs, B. L., & Vakili, D. (2005). A learner-centered framework for E-learning. *Teachers College Record, 107*(8), 1582-1600.

Meehl, P. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 103-115.

*Michko, G. M. (2007). *A meta-analysis of the effects of teaching and learning with technology on student outcomes in undergraduate engineering education.* Unpublished doctoral dissertation, University of Houston, Houston, TX.

Mitra, A., & Steffensmeier, T. (2000). Changes in student attitudes and student computer use in a computer-enriched environment. *Journal of Research on Computing in Education, 32*(3), 417-433.

Møller, A. P., & Jennions, M. D. (2002). How much variance can be explained by ecologists and evolutionary biologists? *Oecologia, 132*, 492-500.

Moore, M. (1989). Three types of interaction. *American Journal of Distance Education, 32*(2), 19-24.

*Niemiec, R., Samson, G. E., Weinstein, T., & Walberg, H. J. (1987). The effects of computer based instruction in elementary schools: a quantitative synthesis. *Journal of Research on Computing in Education, 20*(2), 85-103.

*Onuoha, C. O. (2007). *Meta-analysis of the effectiveness of computer-based laboratory versus traditional hands-on laboratory in college and pre-college science instructions.* Unpublished doctoral dissertation, Capella University, Minnesota.

*Pearson, P. D., Ferdig, R. E., Blomeyer, J. R. L., & Moran, J. (2005). *The effects of technology on reading performance in the middle-school grades: A meta-analysis with recommendations for policy.* Naperville, IL: North Central Regional Educational Laboratory (NCREL).

Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research, 28*(3), 450-461.

Pisapia, J. R., Knutson, K., & Coukos, E. D. (1999). *The impact of computers on student performance and teacher behavior.* Paper presented at the Annual Meeting of The Florida Educational Research Association, Deerfield Beach, Florida.

Plante, J., & Beattie, J. (2004). *Connectivity and ICT integration in Canadian elementary and secondary schools: First results from the information and communications technologies in schools survey.* (Catalogue no. 81-595-MIE-No.017) Ottawa, ON: Statistics Canada and Industry Canada. Retrieved September 19, 2006, from http://www.statcan.gc.ca/pub/81-595-m/81-595-m2004017-eng.pdf

Preiss, R. W., & Allen, M. (1995). Understanding and using meta-analysis. *Evaluation and the Health Profession, 18*(3), 315-335.

Roblyer, M. D. (1988). Assessing the impact of computer-based instruction: A review of recent research. *Computers in the Schools, 5*(3-4).

*Roblyer, M. D., Castine, W. H., & King, F. J. (1988). Assessing the impact of computer-based instruction: A review of recent research. *Computers in the Schools, 5*(3-4).

*Rosen, Y., & Salomon, G. (2007). The differential learning achievements of constructivist technology-intensive learning environments as compared with traditional ones: A meta-analysis. *Journal of Educational Computing Research, 36*(1), 1-14.

Rosenberg, H., Grad, H. A., & Matear, D. W. (2003). The effectiveness of computer-aided, self-instructional programs in dental education: A systematic review of the literature. *Journal of Dental Education, 67*(5), 524-532.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638-641.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage Publications.

*Ryan, A. (1991). Meta-analysis of achievement effects of microcomputer applications in elementary schools. *Educational Administration Quarterly, 27*(2), 161-184.

Salomon, G. (2000, May 3). *Learning today: Not the computer alone...*. Paper presented at the Louvain-la-Neuve, Belgium. Retrieved September 23, 2006, from http://www.ipm.ucl.ac.be/ChaireIBM/Salomon.pdf

*Samson, G. E., Niemiec, R., Weinstein, T., & Walberg, H. J. (1986). Effects of computer-based instruction on secondary school achievement: A quantitative synthesis. *AEDS Journal, Summer*, 312-326.

Schacter, J., & Fagnano, C. (1999). Does computer technology improve student learning and achievement? How, when, and under what conditions? *Journal of Computing Research, 20*(4), 329-343.

*Schenker, J. D. (2007). *The effectiveness of technology use in statistics instruction in higher education: A meta-analysis using hierarchical linear modeling.* Unpublished doctoral dissertation, Kent State University, Ohio.

Schlosser, R. W. (2007). *Appraising the quality of systematic reviews* (Brief No. 17 2007). Austin, TX: National Center for the Dissemination of Disability Research.

Schmid, R.F., Borokhovski, E., Tamim, R.M., Bernard, R.M., Abrami, P.C., Wade, A., Surkes, M.A., Lowerison, G., Lysenko, L. & Galofre, C. (2009, April). *Effect of Computer-based technology use in post-secondary education: preliminary findings of a meta-analysis.* A paper to be presented at the American Educational Research Association, San Diego.

Schmidt, F. (2008). Meta-analysis: A constantly evolving research integration tool. *Organizational Research Methods, 11*(1), 96-113.

Shadish, R. W., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: why validity issues in meta-analysis will not go away. *Clinical Psychology Review, 17*(8), 881-901.

Sheeran, P. (2002). Intention-behavior relations: A conceptual and empirical review. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 12, pp. 1-30). New York: Wiley.

Sipe, T. A., & Curlette, W. L. (1997). A meta-synthesis of factors related to educational achievement: A methodological approach to summarizing and synthesizing meta-analysis. *International Journal of Educational Research, 25*(7), 583-698.

Slavin, R. E. (1984). Meta-analysis in education: How has it been used? *Educational Researcher, 13*, 6-15.

Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy.* Baltimore, MD: Johns Hopkins.

*Soe, K., Koki, S., & Chang, J. M. (2000). *Effect of computer-assisted instruction (cai) on reading achievement: a meta-analysis.* Honolulu, HI: Pacific Resources for Education and Learning.

Spector, J. M., Merrill, M. D., Van Merrienboer, J., & Driscoll, M. P. (Eds.). (2008). *Handbook of Research on Educational Communications and Technology* (3rd ed.). London: Lawrence Erlbaum Associates.

Steiner, D. D., Lane, I. M., Dobbins, G. H., Schnur, A., & McConnell, S. (1991). A review of meta-analyses in organizational behavior and human resources management: An empirical assessment. *Educational and Psychological Measurement, 51*, 609-626.

The New Media Consortium & Educause Learning Initiative (2008). The Horizon report. Retrieved January 15, 2008, from http://www.nmc.org/pdf/2008-Horizon-Report.pdf

Thomas J. Watson. (2008). In *Wikipedia: The free encyclopedia*. Retrieved December 12, 2008, from FL: Wikimedia Foundation, Inc. http://en.wikipedia.org/wiki/Thomas_J._Watson

*Timmerman, C. E., & Kruepke, K. A. (2006). Computer-assisted instruction, media richness, and college student performance. *Communication Education, 55*(1), 73-104.

Torgerson, C.J. (2007). The quality of systematic reviews of effectiveness in literacy learning in English: a 'tertiary' review. *Journal of Research in Reading, 30*(3), 287-315.

*Torgerson, C. J., & Elbourne, D. (2002). A systematic review and meta-analysis of the effectiveness of information and communication technology (ict) on the teaching of spelling. *Journal of Research in Reading, 25*(2), 129-143.

Ungerleider, C., & Burns, T. (2003). *A systematic review of the effectiveness and efficiency of networked ICT in education: A state of the field report to the Council of Ministers of Education, Canada, Industry Canada*. Retrieved October 23, 2005, from http://www.cmec.ca/stats/SystematicReview2003.en.pdf

Van Dusen, G. (2000). Digital dilemma: Issues of access, cost, and quality in media—enhanced and distance education. *ASHE-ERIC Higher Education Report, 27*(5), 1–120.

Van Dusen, G. C. (1998). Technology: Higher education's magic bullet. *Thought and Action, 14*, 59-67.

Wade, A., Abrami, P. C., Bernard, R. M., Turner, H. M., & Peretiatkowicz, A. (2005). *Performing the literature search: Beyond the basics.* Paper presented at the

Workshop presented at the Campbell Collaboration Colloquium, Lisbon, Portugal.

Waight, C. L., Willging, P. A., & Wentling, T. L. (2002). *Recurrent themes in e-learning: A meta-analysis of major e-learning reports.* Paper presented at the Recurrent themes in e-learning: A meta-analysis of major e-learning reports, Honolulu, HI.

Walther, J. B., Anderson, J. F., & Park, D. W. (1994). Interpersonal effects in computer-mediated interaction: A meta-analysis of social and antisocial communication. *Communication Research, 21*(4), 460-487.

Waxman, H. C., Connell, M. L., & Gray, J. (2002). *A quantitative synthesis of recent research on the effects of teaching and learning with technology on student outcomes.* Naperville, IL: Learning Point. Retrieved February 12, 2005, from http://www.ncrel.org/tech/effects/

*Waxman, H. C., Lin, M.-F., & Michko, G. M. (2003). *A meta-analysis of the effectiveness of teaching and learning with technology on student outcomes.* Naperville, IL: Learning Point. Retrieved February 18, 2005, from http://www.ncrel.org/tech/effects2/
waxman.pdf

Whitley, B. E. J. (1997). Gender differences in computer-related attitudes and behavior: a meta-analysis. *Computers in Human Behavior, 13*(1), 1-22.

Wilson, D. B., & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods, 6*(4).

Winn, W. (2002). Current trends in educational technology research: The study of learning environments. *Educational Psychology Review, 14*(3), 331-351.

Wolf, A. W. (2006). *Using technology with learning disabled readers: A meta-analysis.* Unpublished doctoral dissertation, Nova Southeastern University.

Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis.* Beverly Hills, CA: Sage Publications.

*Yaakub, M. N. (1998). *Meta-analysis of the effectiveness of computer-assisted instruction in the technical education and training.* Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University.

*Zhao, Y. (2003). Recent developments in technology and language learning: A literature review and meta-analysis. *CALICO Journal, 21*(10), 7-27.

APPENDIX A

FULL CODEBOOK

Study identification:

- Identification number

- Author

- Title

- Year of publication:

- Type of publication:

  5. Journal
  6. Dissertation
  7. Conference Proceedings
  8. Report / Gray literature

Contextual features:

- Research question

- Technology addressed

- Control group definition/description

- Experimental group treatment definition/description

- Grade level

- Subject matter:

  1- Science/health

  2- Language

  3- Math

  4- Technology

  5- Social Science

  6- Combination

  7- Information literacy

  8- Engineering

  9- Not specified

Methodological features:

Search phase

- Search time frame
- Justification for search time frame:
  3. No
  4. Yes
- Literature covered:
  3. Published studies only
  4. Published and unpublished studies
- Search strategy:
  5. Search strategy not disclosed, no reference to search strategy offered
  6. Minimal description of search strategy with brief reference to resources searched
  7. Listing of resources and databases searched
  8. Listing of resources and databases searched with sample search terms
- Resources used:
  5. Data-base searches
  6. Computerized search of web resources
  7. Hand search of specific journals
  8. Branching
- Databases searched
- Number of data-bases searched

Review phase

- Inclusion/exclusion criteria:
  4. Criteria not disclosed with no description offered
  5. Overview of criteria presented briefly
  6. Criteria specified with enough detail to allow for easy replication

- Included research type:

  4. RCT only
  5. RCT/Quasi
  6. RCT/Quasi/Pre

  999.Not specified

- Article review:

  5. Review process not disclosed
  6. review process by one researcher
  7. Rating by more than one researcher
  8. Rating by more than one researcher with inter-rater agreement reported

Effect size and study feature extraction phase

- ES Extraction:

  5. Extraction process not disclosed, no reference to how it was conducted
  6. Extraction process by one researcher
  7. Extraction process by more than one researcher
  8. Extraction process by more than one researcher with inter-rater agreement reported

- Code Book:

  5. Codebook not described, no reference to features extracted from primary literature
  6. Brief description of main categories in codebook
  7. Listing of specific categories addressed in codebook
  8. Elaborate description of codebook allowing for easy replication

- Study feature Extraction:

  5. Extraction process not disclosed, no reference to how it was conducted
  6. Extraction process by one researcher
  7. Extraction process by more than one researcher
  8. Extraction process by more than one researcher with inter-rater agreement reported

Analysis

- Independence of data:

  3. No
  4. Yes

- Weighting by number of comparisons:

  3. Yes
  4. No

- ES weighted by sample size:

  3. No
  4. Yes

- Homogeneity analysis:

  3. No
  4. Yes

- Moderator analysis:

  3. No
  4. Yes

- Meta-regression conducted:

  3. No
  4. Yes

Further reporting aspects

- Inclusion of list of studies:

  3. No
  4. Yes

- Inclusion of ES table:

  3. No
  4. Yes

- Time between last study and publication date

Effect Size information:

- Effect size type:

5. Glass
6. Cohen
7. Hedges
8. Others: specify

Total ES

- Mean ES
- SE
- SE extraction:

5. Reported
6. Calculated from ES and sample size
7. Calculated from Confidence interval
8. Replaced with weighted average SE

- Time frame included
- Number of studies included
- Number of ES included
- Number of participants
- Number of participants extraction:

3. Calculated
4. Given

Specific ES

- Specific variable
- Mean ES
- SE

- SE extraction:
    1. Reported
    2. Calculated from ES and sample size
    3. Calculated from Confidence interval
    4. Replaced with weighted average SE
- Time frame included
- Number of studies included
- Number of ES included
- Number of participants
- Number of participants extraction:
    1. Calculated
    2. Given