THE NEED FOR META-ANALYTIC THINKING IN EDUCATIONAL TECHNOLOGY RESEARCH

A Dissertation

by

NICOLA LYNN RITTER

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Bruce Thompson
Committee Members,	Victor L. Willson
	Trina J. Davis
	Ronald Zellner
Head of Department,	Victor L. Willson

August 2014

Major Subject: Educational Psychology

ABSTRACT

The present journal article formatted dissertation assessed the extent of meta-analytic thinking currently used educational technology research. In the first study, the author examined the journals, *Computers & Education, International Journal of Computer-Supported Collaborative Learning*, *British Journal of Educational Technology*, *Australasian Journal of Educational Technology*, and *Educational Technology Research and Development*, between 2012 and 2013 to offer empirical evidence of the field's current status with regard to reporting results using meta-analytic thinking. These articles represented a total of 32,131 research methods and statistical techniques recorded from 1,171 articles. Findings point to little change in how educational technology researchers conduct investigations. Quantitative methods continue to dominate the field as a whole. Most authors reported the type of sampling procedure used in their investigations. Few researchers reported score reliability estimates using their own data. Findings also suggest few authors report informationally-adequate statistics. One area of concern is the tendency to report a mean without the SD about the mean. Another area of concern is the lack of reporting correlation matrices with accompanying means and standard deviations or covariance matrices.

In the second study, the author conducted a meta-analysis to offer a glimpse of where the field could go once researchers begin to think meta-analytically. The author cumulated findings from nine studies which used the Technology Acceptance Model (TAM) to explain undergraduate students' acceptance of online learning. The author used meta-analytic structural equation modeling and multiple-group analysis to test four

path models. The meta-analytic findings suggest the TAM is not a valid theoretical model to explain undergraduates' acceptance of college online courses. The multiple group analysis emphasized that the parameter estimates between studies resulted in statistically different findings, suggesting the findings across studies are not replicable.

DEDICATION

This dissertation is dedicated to my mother, Melinda, my inspiration. The challenges you have overcome gave me the strength to complete this. Thank you for your enduring love and support through this process. I am honored to be called your daughter. I love you.

ACKNOWLEDGEMENTS

The most important acknowledgement is to my God, El Sali. The Lord entrusted me with this task and now that it is finished, I commit the dissertation back to Him as an offering. You have demonstrated time and again, in my weakness you are strong (2 Corinthians 12:9), that *nothing* is impossible with You (Luke 1:37), while continuing to do a good work in me (Philippians 1:6). This journey was a step in obedience and stretch of my faith. I hope you find favor in this example of my imperfect progress. May You be glorified in this experience!

Many people have helped me throughout this dissertation process. Undoubtedly, the person that provided the most professional guidance was Dr. Bruce Thompson, my committee chair. Dr. Thompson, I am thankful for your guidance, thoughtfulness, and expertise. You have been an invaluable resource through this process. I particularly want to thank you for giving me the gift to cringe when I see a stepwise regression or the phrase, “reliability of the test,” reported. You have opened my eyes to numerous concepts and expanded my view to see my blind spots, while nurturing my professional growth in ways I often did not realize I needed. I am always eager to learn from you whether in your class, at workshops, in your office, or through your many written works. You are a masterful writer, researcher and teacher. It’s a humbling realization and incredible opportunity to work with you.

Michael Murdock stated, “Each relationship nurtures a strength or weakness within you.” My committee members have done just that. I am thankful for my committee chair, Dr. Bruce Thompson, whose mentorship and support from my first

interaction with him while I considered applying to Learning Sciences program to the concluding level of this document enabled me to develop an in-depth understanding of the subject and the academy. I would also like to thank you for equipping and assisting me with the resources needed during data analysis. I would like to thank my committee members, Dr. Vic Willson, Dr. Trina Davis, and Dr. Ron Zellner for their guidance and support throughout the course of this research. Dr. Willson, thank you for equipping me to apply the theoretical to reality. You are truly an innovative researcher. To Dr. Davis, I appreciate your expertise, kindness, and care. Thank you for providing me with great tools and direction to ensure success through this process. Dr. Zellner, thank you for taking on another doctoral student during your time as professor emeritus. This act alone demonstrates your dedication to students. Your experience, perspective, and expertise was invaluable to me and can be found in no other.

Dr. René Mercer, my allied reader and dear colleague. You were often the only sunshine to MY day and the lead comedian to this dark comedy. I cannot thank you enough for the numerous acts of kindness and encouragement along this journey. If every doctoral student had a René through this process, there would be no ABD suffix. Our talks about faith, family, and life are such a blessing to me. May God continue to bless and pour into your life so that you may continue to share your blessings with others and give Him all the glory.

Amanda Franks, my "agreeable" inter-rater. Thank you for spending 41.5 hours comparing coding schemes with me plus the numerous hours you spent coding articles on your own - all for one sentence in this manuscript. "Cohen's kappa was 0.99" (Ritter,

2014). I am grateful for your patience and attention to detail during data collection to ensure the reliability of the data at hand.

Bilgin Navruz, my incognito superhero. Thank you for your assistance during data analysis phase. You are a talented teacher and I am honored to have learned so much from you.

Dr. Erhan Delen, my gifted co-author and colleague. Thank you for pulling a substantial portion of the articles in Study 1.

I also want to extend my gratitude to the College of Education & Human Development's research grant, which provided the financial support needed to make this study a success.

Lastly, I would like to acknowledge those who have supported me through this laborious, yet rewarding task.

Dr. Kyle Ritter, my loving and brilliant husband. I thank God every day for bringing you into my life. You are more than I ever prayed for. I could never imagine having a husband with all the wonderful attributes you exhibit. Thank you for your unlimited patience through this process. This never would have been done without the sacrifices you have made on my behalf. I love you with all my heart.

Captain and Mrs. Robert Frizzell, my precious grandparents. Thank you for your everlasting love and support. I am at this place in my life because of what you instilled in me.

David Preston Hummadi, my dear brother. Thank you for reminding me to not take life so seriously and what really matters. You have modeled the resilience I hope to demonstrate in my own life. I am honored to be called your sister.

The Ritter & Mathews families, my cherished in-laws. Thank you for your continued prayers, patience, and devoted care for me during my college journey. I am grateful to be a part of your precious family and to experience life alongside you. Thank you for outpouring the love you have into my life. I am a better person because of this family.

Dr. Gisela Lin, my academic coach. Your wisdom and insight through this process is priceless.

Dr. Goodson, my writing coach. Thank you for your guidance in navigating academia and academic writing. You truly are the rockstar to this hit!

A huge thank you to my P.O.W.E.R. colleagues, my Facebook writing group, my Beacon & Mt. Calvary church families, the Association of Baptist Students at Texas A&M University, and the Educational Psychology and Teaching, Learning, & Culture departments faculty and staff for making my 13 years as a student at Texas A&M University a fond and memorable experience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

For over half a century, various scholars have called attention to the problems within educational technology research (Clark, 1983; Hoban, 1958; Mielke, 1968). More recently, researchers have urged the field to reconsider how research is conducted (Bell, Schrum, & Thompson, 2008; Haertel & Means, 2003; Means & Haertel, 2004). Reeves (2000) suggested that the quality of published educational technology research is generally poor. Moreover, Mitchell (1997) described educational technology research as pseudo-scientific, lending itself to unsupported conclusions based on poor measurement practices and statistical blunders. So, where are we now?

One way to assess whether or not the call has been heard is to systematically evaluate the present state of educational technology research practice. As Skidmore (2009) stated, "Indeed, it is only when we hold the mirror up to ourselves and purposely reflect on where we are at the present time that we can begin to consider where we would like to be. To nonsensically consider where we would like to be without considering where we are now, would be like asking for directions without knowing our present location!" (p. 4).

Meta-Analytic Thinking

Educational technology research is plagued with isolated researchers who rarely associate with a greater research agenda (Reeves, 2000). As a result, meta-analysts are challenged to synthesize and interpret the field's findings. Challenges in synthesizing and interpreting findings arise when researchers do not think meta-analytically.

Meta-analytic thinking moves researchers beyond their own study and views their study in the light of the greater body of literature. Thompson (2002) described meta-analytic thinking as "both (a) prospective formulation of study expectations and design by explicitly invoking prior effect sizes and (b) the retrospective interpretation of new results, once they are in hand, via explicit, direct comparison with the prior effect sizes in the related literature" (p. 28). In other words, when researchers conduct studies using meta-analytic thinking, researchers establish study expectations and design before the study takes place and is built upon previous studies. Once the study is conducted, then results are interpreted in the context of the greater body of literature. In a field as scattered as educational technology, meta-analytic thinking is crucial to make better sense of the field's findings and for the field to advance efficiently.

One way to integrate quantitative original research is to conduct meta-analyses. The purpose of meta-analyses is to cumulate findings of studies on a particular topic and estimate effects across an entire body of research. Meta-analyses serve as an example of meta-analytic thinking.

Organization of Document

The current document is divided into four distinct sections. With the exceptions of the first and last sections, the sections are written as individual manuscripts with the intention to submit for publication. A description of each section is provided below:

1) The first section is an introductory section which presents an overview of the topics to be investigated.

2) The second section offers a synthesis of statistical techniques used in the field of educational technology in the past two years. This section presents the introduction, literature review, and methods section of the first journal article.
3) The third section reports a meta-analysis to demonstrate an example of meta-analytic thinking. This section presents the introduction, literature review, and methods section of the second journal article.

The fourth and final section is the concluding section that connects the findings from the two manuscript sections into a coherent, succinct conclusion for the entire thesis.

2. STATISTICAL METHODS USED IN EDUCATIONAL TECHNOLOGY JOURNALS 2012-2013

Research syntheses of statistical techniques are published commonly in various fields, including education. Methodological reviews come in multiple forms, such as investigations of specific techniques published within a single journal or across multiple journals. For example, Willson (1980) reviewed articles published in *American Educational Research Journal* (AERJ) between 1969 and 1978 with respect to statistical methodology. More recently, disciplines within education have also begun to evaluate studies across multiple journals. For example, Warne, Lazo, Ramos, and Ritter (2012) reported the statistical techniques in five gifted education journals. Essentially, the present review models Warne et al. (2012), who reviewed articles across multiple journals

Educational technology researchers have also conducted methodological reviews. Most researchers identified the types of research methods (e.g., quantitative, qualitative, or mixed methods) used in educational technology literature (e.g., Koble & Bunker, 1997; Rourke & Szabo, 2002; Wang & Lockee, 2010). Other researchers extended their investigations beyond identifying research methods to synthesize the types of experimental designs (e.g., Cheung & Hew, 2009; Peterson-Karlan, 2011; Randolph, Julnes, Sutinen, & Lehman, 2008; Shih, Feng, & Tsai, 2008; Şİmşek, Özdamar, Uysal, Kobak, Berk, Kiliçer, & Çİğdem, 2009). Likewise reviews may investigate other methodological issues such as sampling method, reliability, and statistical techniques (e.g., Alper & Gülbahar, 2009; Edyburn, 2000; Karataş, 2008; Lee, Driscoll, & Nelson, 2004, 2007; Randolph, 2008; Randolph et al., 2008).

Literature Review

Previous reviews adequately investigated the empirical nature, research methods, and experimental designs used in educational technology. Educational technology researchers captured information about the literature's empirical and non-empirical nature (e.g., Chen & Hirschheim, 2004; Farhoomand & Drury, 1999; Hrastinski & Keller, 2007). Empirical articles are favored over non-empirical articles and the ratio between the two has remained constant over the past decade. Similarly, many researchers identified the research methods used in educational technology. Historically, quantitative methods dominated the field, yet today the field has seen a more balanced use of quantitative, qualitative, and mixed methodologies. Lastly, researchers identified the experimental designs used, despite the lack of explicit reporting in the original article. Identifying trends in experimental designs are particularly difficult given the variation in classification schemes. Specifically, a synthesis among five reviews resulted in 14 different experimental design categories, with some overlapping others.

Despite the variety of previous reviews in educational technology, few researchers have reviewed sampling methods, score reliability, or statistical techniques. Researchers found convenience samples to plague the field, yet the actual sampling methods are often not directly reported (Alper & Gülbahar, 2009; Edyburn, 2000; Randolph et al., 2008). Likewise, the few researchers who reviewed score reliability information found poor reporting practices (Lee et al., 2004, 2007; Randolph, 2008).

Finally, researchers who evaluated statistical methods simply reported a brief list of statistical techniques (e.g., Karataş, 2008; Lee et al., 2004, 2007; Randolph et al., 2008). Appendix A offers more details on previous findings.

Researchers have a clear picture of the empirical nature, research methods, and experimental designs used in educational technology. Nevertheless, more information is needed about sampling methods, score reliability, and statistical techniques used in the field. In brief, some researchers examined trends over time and across journals, while others did not observe trends. Although these topics are explored to some extent, little is known about the statistical trends in educational technology. Moreover, few reviewed the field of educational technology as a whole. For these reasons, a thorough synthesis of the field regarding the use of statistical techniques is needed.

Purpose

The purpose of the current study was to provide a synthesis of statistical techniques used in the field of educational technology in the past two years. The present article undertook a similar review of published journal articles in educational technology to determine which statistical techniques are used, and whether discernible trends emerged within journals and across the discipline. In essence, this review models Warne et al., (2012), who reviewed articles across multiple journals. Where other methodological reviews of educational technology are limited to specific areas, the current study reviews the field as a whole. Moreover, the current study moved beyond classifying research methods, and examined the statistical techniques used in the educational technology literature to better understand the current state of the field's use of statistics. The present study aimed to answer two research questions: 1) Which statistical techniques are used in educational technology research? and 2) How have the proportion of statistical techniques varied across the top five educational technology journals?

Methods

Inclusion Criteria

The present methodological review examined five influential research journals over a two year period, from 2012 to 2013. The educational technology journals selected for review were based on the journals' impact factors using the *2011 Journal Citation Reports*® (JCR®) *Social Science Edition*. An impact factor is the average number of citations for articles published in social science journals to quantify the relative importance of a journal within its field. Given the 2011 JCR® database did not have an education technology sub-category, I compared the listed JCR® journals with the *Educational Technology of Library Science* directory from the 2012 *Cabell's Directories of Publishing Opportunities* (Cabell's).

Exclusions

The current study reviewed only papers that were identified as empirical articles. Non-empirical papers, such as editorials and book reviews, were excluded from the current study. In addition, the journal, *Educational Research Review*, was excluded, despite being ranked in the top five journals between the JCR® database and Cabell's directory. The journal was excluded because of the lack of emphasis on educational technology.

Sample

The journals with the highest impact rating in rank order were: *Computers & Education, International Journal of Computer-Supported Collaborative Learning*, *British Journal of Educational Technology*, *Australasian Journal of Educational Technology*, and *Educational Technology Research and Development*. Note the top five journals are not representative of the diverse topics discussed within educational technology nor representative of the variety of contexts in which educational technology is applied (e.g., K-12, higher education, or industry).

Elsevier (2012) publishes the top ranked journal, *Computers & Education* (*C&E*). According to Elsevier's (2012) website, the journal's scope is an "interdisciplinary forum for communication in the use of all forms of computing" ("Aims and Scope," para. 1), including open and distance learning environments and spans across primary to tertiary levels in education. The top ranking journal has a 2.62 impact factor.

Springer publishes the second ranked journal, *International Journal of Computer-Supported Collaborative Learning* (*ijCSCL*), quarterly on behalf of the International Society of the Learning Science. The journal promotes a deeper understanding of computer-supported collaborative learning with emphasis on how people learn and how to design in this context. The journal has a 2.24 impact factor.

Wiley-Blackwell publishes the *British Journal of Educational Technology* (*BJET*) bi-monthly on behalf of the British Educational Research Association (BERA). According to the publisher's website (John Wiley & Sons, 2012), the *BJET* (2012) provides the “widest possible coverage of developments in international educational and training technology” (Aims and Scope, para. 1). The third ranked journal has a 1.54 impact factor.

Australasian Society for Computers in Learning in Tertiary Education (ASCILITE; 2012) publishes the fourth ranked journal, *Australasian Journal of Educational Technology* (*AJET*). According to ASCILATE (2012), the journal’s scope entails “research and review articles in educational technology, information and communications technologies for education, online and e-learning, educational design, multimedia, computer assisted learning, and related areas” (*AJET* in brief, para. 1). The journal has a 1.52 impact factor.

Springer (2012) publishes the fifth ranked journal, *Educational Technology Research and Development* (*ETR&D*) bi-monthly on behalf of the Association for Educational Communications & Technology (AECT). The journal’s scope promotes dual emphasis in research and development in educational technology. According to Springer (2012), the research section emphasizes “applications of technology or instructional design in educational settings” (“Aims and Scope,” para. 1), while the development section emphasizes “planning, implementation, evaluation and management of a variety of instructional technologies and learning environments” (“Aims and Scope,” para. 2). The journal has a 1.09 impact factor.

Coding

I developed a coding scheme based on three sources: previous reviews, publication standards, and meta-analytic coding suggestions. First, I consulted previous reviews (Chen & Hirschheim, 2004; Skidmore & Thompson, 2010; Warne et al., 2012), which led to an initial coding scheme. Based on the previous reviews, I observed the original articles under review did not follow publication standards. As a result, next, I appraised the statistical reporting standards from the American Psychological Association (2010), APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008), and Wilkinson and the Task Force on Statistical Inference (1999). Lastly, I reviewed works on meta-analytic coding, which identified the type of information to report in original articles to best synthesize information for future use (Cooper, 2010; Lipsey & Wilson, 2001). The coding sheet and definitions are located in Appendix B.

Procedures

First, each article was classified as empirical or non-empirical. Next, the empirical articles were further classified as applying qualitative, quantitative, or mixed methodology. Lastly, statistical techniques were identified for all articles using quantitative or mixed methods. For example, while all 1,171 articles were screened, only 720 articles using quantitative or mixed research methods were further reviewed for statistical techniques. Qualitative research articles and non-empirical articles were classified accordingly and not reviewed further.

The author and one graduate assistant coded articles using the coding sheet. The author recorded a total of 32,131 items from all 1171 articles. The author trained one graduate student to independently code 35.44% ($n = 415$) of a random sample of all articles using the coding sheet. Each coder used a Microsoft Excel spreadsheet to the indicate presence of the particular methodological technique. The two coders compared items recorded and resolved discrepancies. Inter-rater reliability was computed using Cohen's kappa (κ). Cohen's kappa is a coefficient of agreement for nominal scales, which was the measurement scale used in the coding sheet. Cohen's kappa was 0.99. Cohen's kappa measures the agreement between two raters, taking into account the number of cases, categories, and raters. Values range between +1, indicating complete agreement, and -1, indicating complete disagreement. Cohen's kappa was 0.99.

Results

The current study found empirical articles are favored over non-empirical articles, which mirror the findings of previous methodological reviews in educational technology (e.g., Chen & Hirschheim, 2004; Farhoomand & Drury, 1999; Hrastinski & Keller, 2007). Among the 1,171 articles reviewed, 509 (43.47%) were quantitative, 216 articles (18.45%) were qualitative, and 211 (18.02%) were reports from mixed methods research projects — a proportion similar to the findings from Hrastinski and Keller's (2007) review of educational technology articles from 2000 to 2004. The remaining 235 articles (20.07%) were non-empirical reports, such as editorials, book reviews, and other non-research articles. Figure 1 presents the type of research methods used across the five journals reviewed. Given the scope of the current study, only quantitative and mixed methods research reports were analyzed further.

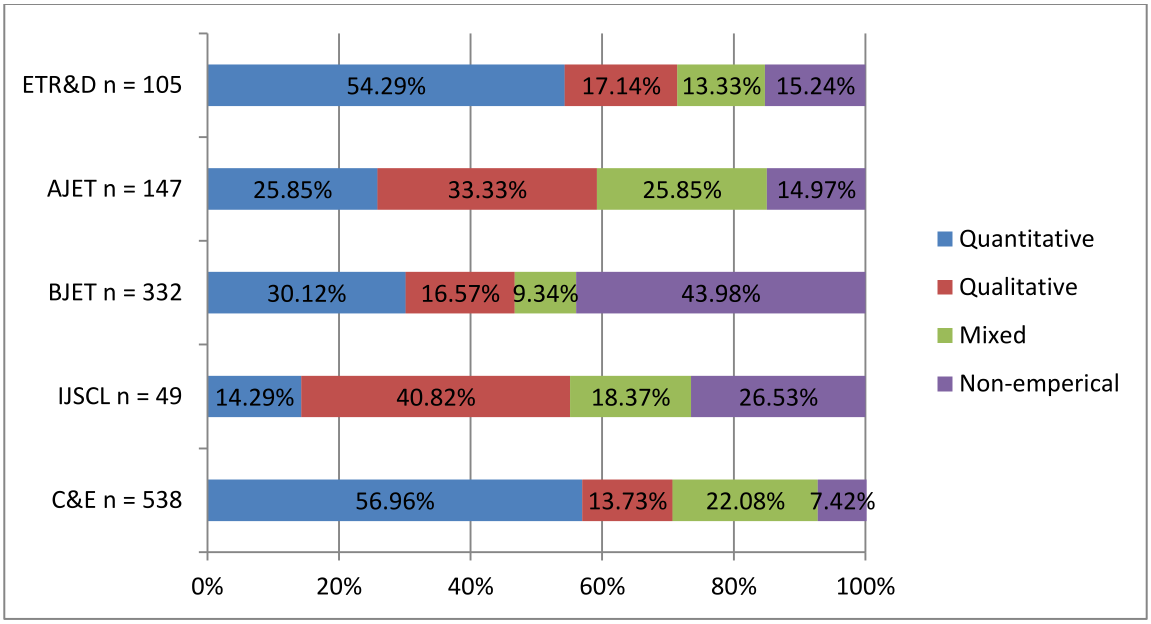


Figure 1 Type of Research Method Used by Journal.

Sampling Procedures

Scholars previously suggested that educational technology research is plagued with convenience samples (Alper & Gülbahar, 2009; Edyburn, 2000; Randolph et al., 2008). After I reviewed the educational technology research between 2012 and 2013, I found most authors did not report the type of sampling procedure used. Specifically, authors of 617 articles (85.69%) did not report the type of sampling procedure used. Among the 100 articles that included the type of sampling procedure, 22 articles (3.06%) indicated they used convenience sampling. The most commonly reported sampling procedure was simple random sampling, appearing in 30 articles (4.17%). Authors of the remaining 70

Table 1.

Sampling Procedure Reported in Education Technology Journals Organized by Journal

Sampling Procedure	Journal					Total
	C&E	*ijSCL*	*BJET*	*AJET*	*ETR&D*	
No sampling procedure reported	374 (87.79%)	15 (93.75%)	108 (82.44%)	62 (81.58%)	58 (81.69%)	617 (85.69%)
Convenience sampling	11 (2.58%)	0	4 (3.05%)	5 (6.58%)	2 (2.82%)	22 (3.06%)
Simple random sampling	14 (3.29%)	0	6 (4.58%)	7 (9.21%)	3 (4.23%)	30 (4.17%)
Database	12 (2.82%)	0	7 (5.34%)	1 (1.32%)	3 (4.23%)	23 (3.19%)
Purposive sampling	6 (1.41%)	1(6.25%)	2 (1.53%)	1 (1.32%)	1 (1.41%)	11 (1.53%)
Stratified random sampling	4 (0.94%)	0	1 (0.76%)	0	1 (1.41%)	6 (0.83%)
Cluster sampling	3 (0.70%)	0	1 (0.76%)	1 (1.32%)	1 (1.41%)	6 (0.83%)
Unspecified non-random sampling	1 (0.23%)	0	1 (0.76%)	0	0	2 (0.28%)

Note. The following sampling procedures were only used once: chain-referral sampling (*C&E*), stratified multi-stage cluster sampling (*C&E*), cluster-randomized sampling (*ETR&D*), probability proportional to size (PPS) in a multistage cluster sampling design (*ETR&D*).

articles reported samples from databases (23 articles; 3.19%), purposive sampling (11 articles; 1.53%), stratified random sampling (6 articles; 0.83%), and cluster sampling (6 articles; 0.83%). Authors of two articles (0.28%) reported they used a non-random sampling method, yet did not specify the type used.

Table 1 presents sampling procedures used across the five journals. Overall, the authors across all five journals rarely reported the type of sampling procedure used. In fact, less than 20% of the quantitative or mixed methods research articles published in each journal reported the type of sampling procedure used. In *ijSCL*, the authors of quantitative and mixed method research reported the type of sampling procedure in only one article.

Based on the current study, random assignment was rarely used in educational technology research, a finding similar to Lee et al. (2007). Authors of 146 articles (20.28%) randomly assigned participants to treatment groups. Random assignment appeared in 93 articles published in *C&E*, 8 articles published in *ijSCL*, 18 articles published in *BJET*, 7 articles published in *AJET*, and 20 articles published in *ETR&D*.

Score Reliability

Scholars argued that researchers should report the reliability statistics from the data at hand because of the possible impact reliability has on the interpretation of research results (American Education Research Association, 2006; Wilkinson et al., 1999). Thompson (1994) forewarned,

> The failure to consider score reliability in substantive research may exact a toll on the interpretations within research studies. For example, we may conduct studies that could not possibly yield noteworthy effect sizes given that score reliability inherently attenuates effects sizes. Or we may not accurately interpret the effect sizes in our studies if we do not consider the reliability of the scores we are actually analyzing. (p. 840)

Despite these warnings, authors of approximately half (420 articles) of the quantitative and mixed method articles reported reliability data from their own sample. The findings of the current study mirror findings of Willson (1980) seen 34 years ago. Authors of 16 articles (2.22%) invoked a reliability coefficient from a previous study or test manual.

In addition, authors of 105 articles (14.58%) referred to reliability as an attribute of the test, instead of data at hand, by using phrases such as "reliability of the test" suggesting there may be misconceptions about reliability (Vacha-Haase, 1998). As Thompson (1992) noted,

> This is not just an issue of sloppy speaking – the problem is that sometimes we unconsciously come to think what we say or what we hear, so that sloppy speaking does sometimes lead to a more pernicious outcome, sloppy thinking and sloppy practice. (p. 436)

This is not good practice because tests are not reliable! Gronlund and Linn (1990) explained,

> Reliability refers to the results obtained with an evaluation instrument and not to the instrument itself… Thus, it is more appropriate to speak of the reliability of the 'test scores' or the 'measurement' than of the 'test' or the 'instrument'. (p. 78).

The following sections and Table 2 presents the type of reliability methods that were found across all five journals.

Internal consistency. By far the most common type of reliability measure reported was internal consistency coefficients. The most common reliability coefficient among all reliability measures was Cronbach's alpha, appearing in 313 articles, almost half of all quantitative and mixed method studies. The widespread use of Cronbach's alpha in educational technology research is also found in the general psychological literature (Hogan, Benjamin, & Brezinski, 2000), and gifted education (Warne et al., 2012). Composite reliability and the Kuder-Richardson formula, and standardized Cronbach's alpha were found in 36 articles (5.00%) and 22 articles (3.06%), and 2 articles (2.82%), respectively. While other internal consistency coefficients, such as McDonald's Omega, and person separation coefficients were reported only once. Authors of four articles (0.56%) computed some type of internal consistency coefficient but did not specify which type of internal consistency coefficient was calculated! This is

Table 2.
Reliability Methods Used in Educational Technology Journals, 2012-2013, Organized by Journal

Method	Journal					Total
	C&E	*ijSCL*	*BJET*	*AJET*	*ETR&D*	
Reliability induction	8 (1.88%)	0	3 (2.29%)	3 (3.95%)	2 (2.82%)	16 (2.22%)
Reported reliability for own data	256 (60.09%)	14 (87.50%)	74 (56.49%)	38 (50.00%)	38 (53.52%)	420 (58.33%)
"Reliability of the test" (not scores)	71 (16.67%)	0	12 (9.16%)	12 (15.79%)	10 (14.08%)	105 (14.58%)
Cronbach's alpha	201 (47.18%)	4 (25.00%)	54 (41.22%)	32 (42.11%)	22 (30.99%)	313 (43.47%)
Composite reliability	25 (5.87%)	0	6 (4.58%)	4 (5.26%)	1 (1.41%)	36 (5.00%)
Kuder-Richardson formula	13 (3.05%)	1 (6.25%)	2 (1.53%)	2 (2.63 %%)	4 (5.63%)	22 (3.06%)
Standardized Cronbach's alpha	2 (2.82%)	0	0	0	0	2 (0.28%)
Unspecified internal consistency	1 (0.23%)	0	3 (2.29%)	0	0	4 (0.56%)
Test-retest	6 (1.41%)	0	1 (0.76%)	0	1 (1.41%)	8 (1.11%)
Cohen's kappa	27 (6.34%)	6 (37.50%)	5 (3.82%)	4 (5.26%)	7 (9.86%)	49 (6.81%)
percent agreement	25 (5.87%)	4 (25.00%)	7 (5.34%)	1 (1.32%)	5 (7.04%)	42 (5.83%)
Krippendorf's alpha	4 (0.94%)	4 (25.00%)	0	0	1 (1.41%)	9 (1.25%)
Correlation between ratings	9 (2.11%)	1 (6.25%)	5 (3.82%)	1 (1.32%)	6 (8.45%)	22 (3.06%)
Free-Marginal Multirater Kappa	0	0	2 (1.53%)	0	0	2 (0.28%)
Intraclass correction coefficient	9 (2.11%)	1 (6.25%)	1 (0.76%)	1 (1.32%)	3 (4.23%)	14 (1.94%)
Unspecified inter-rater method	4 (0.94%)	0	0	1 (1.32%)	1 (1.41%)	6 (0.83%)

Table 2. Continued

Method	Journal					Total
	C&E	*ijSCL*	*BJET*	*AJET*	*ETR&D*	
Alternate/ Parallel forms	6 (1.41%)	0	0	0	0	6 (0.83%)
Spearman-Brown	3 (0.70%)	1 (6.25%)	0	1 (1.32%)	0	5 (0.69%)
Guttmann split-half	4 (0.94%)	0	0	0	0	4 (0.56%)
IRT-based reliability measure	9 (2.11%)	0	0	0	1 (1.41%)	10 (1.39%)

Note. IRT = item response theory; ICC = intraclass correlation coefficient; ROC = receiver operating characteristic curve; Little's MCAR test = Little's Missing Completely at Random test; Q-Q plot = Quantile-quantile plot. The following reliability statistics were used only once: Bland-Altman plot (*C&E*), Bayesian-based reliability measure (*C&E*), fidelity measures (*C&E*), H coefficient (*C&E*), McDonald's Omega (*C&E*), person separation reliability (*C&E*), Raykov's (2001) CFA-based method (*C&E*), squared multiple correlation (*C&E*), an unspecified reliability coefficient (*C&E*), rwg analysis (*BJET*), and Fleiss's kappa (*ETR&D*).

troubling, because different estimates have different properties and may not be directly comparable across types.

Interrater reliability. Interrater reliability measures were the next most common type of reliability measure reported. Cohen's kappa and percent agreement between raters were the most common types of interrater reliability measure – reported in 49 articles (6.81%) and 42 articles (5.83%), respectively. Correlations between ratings were the next most common (22 articles; 3.06%), with intraclass correction coefficients (14 articles; 1.94%), Krippendorf's alpha (9 articles; 1.25%), Free-Marginal Multirater Kappa (2 articles; 0.28%) also appearing in the literature. Other less commonly reported inter-rater reliability coefficients were a Fleiss's kappa, a Bland-Altman plot, and a r_{wg} analysis (e.g., James, Demaree & Wolf, 1993), each appearing in one article. As seen with internal consistency coefficients, authors of 6 articles (0.83%) calculated some type of inter-rater coefficient, yet did not specify which type was computed.

Other types of reliability measures were less commonly reported. Authors of 10 articles (1.39 %%) reported item response theory (IRT) based reliability measures. Split-half correlations were reported for nine articles (1.25%), with five of those correlations being corrected by the Spearman–Brown prophecy formula and four of those using the Guttmann split-half formula. Test–retest reliability coefficients appeared in 8 articles (1.11%), while alternate/parallel forms reliability coefficients appeared in 6 articles (0.83%). Bayesian-based reliability measures, Raykov's (2001) CFA-based method, fidelity measures, squared multiple correlations, and H coefficients were each found in one article. Authors of one article (0.14%) reported an interrater reliability coefficient, yet did not specify which type of reliability coefficient was reported.

Statistical Techniques

Table 3 presents the statistical techniques used across the journals reviewed. The following sections report on multiple methods used in educational technology research over the past two years. Authors of articles often used more than one statistical technique.

Descriptive Statistics

Most authors of quantitative and mixed methods articles reported descriptive statistics. Means were reported for 78.61% of articles and standard deviations for 69.17% of articles. At times, means were reported without accompanying standard deviations or other variance-based statistics. The lack of reporting accompanying means and standard deviations is largely recognized as poor practice (Thompson, 2006) and makes articles harder to include in meta-analyses (Cooper, 2010). Standard deviations quantify how well central tendency statistics do at representing the data set as a whole.

Effect Sizes

Recent publications stressed the importance of effect size reporting. Wilkinson et al. (1999) recommended that researchers should “*always* present effect sizes for primary outcomes” (p. 599, emphasis added). Moreover, the APA Task Force on Statistical Inference (Wilkinson et al., 1999), stated, "reporting and interpreting effect sizes in the context of previously reported effects is *essential* to good research" (p. 599, emphasis added). Likewise, the fifth edition of the APA manual emphasized, “For the reader to

Table 3.

Statistical Methods Used in Educational Technology Journals, 2012-2013, Organized by Journal

Method	Journal					Total
	C&E	*ijSCL*	*BJET*	*AJET*	*ETR&D*	
Descriptive statistics						
Means	343 (80.52%)	14 (87.50%)	86 (65.65%)	61 (80.26%)	62 (87.32%)	566 (78.61%)
SDs	298 (69.95%)	13 (81.25%)	76 (58.02%)	52 (68.42%)	59 (83.10%)	498 (69.17%)
Basic inferential statistics						
t tests	165 (38.73%)	5 (31.25%)	36 (27.48%)	32 (42.11%)	23 (32.39%)	261 (36.25%)
ANOVA	134 (31.46%)	10 (62.50%)	27 (20.61%)	16 (21.05%)	34 (47.89%)	221 (30.69%)
ANCOVA	40 (9.39%)	1 (6.25%)	10 (7.63%)	3 (3.95%)	7 (9.86%)	61 (8.47%)
MANOVA	19 (4.46%)	2 (12.50%)	7 (5.34%)	4 (5.26%)	7 (9.86%)	39 (5.42%)
MANCOVA	9 (2.11%)	1 (6.25%)	1 (0.76%)	0	0	11 (1.53%)
Correlation / regression						
Pearson's r	151 (35.45%)	6 (37.50%)	40 (30.53%)	15 (19.74%)	23 (32.39%)	235 (32.64%)
Spearman's rho	17 (3.99%)	1 (6.25%)	3 (2.29%)	3 (3.95%)	2 (2.82%)	26 (3.61%)
Point-biserial correlation	2 (0. 47%)	0	0	0	1 (1.41%)	3 (0.42%)
Phi coefficient	2 (0. 47%)	1 (6.25%)	0	0	0	3 (0.42%)
Kendall's tau	1 (0.23%)	0	0	0	1 (1.41%)	2 (0.28%)
Canonical correlation	0	0	0	1 (1.32%)	0	1 (0.14%)
Point-measure correlation	1 (0.23%)	0	0	0	0	1 (0.14%)
Multiple regression	48 (11.27%)	3 (18.75%)	12 (9.16%)	4 (5.26%)	7 (9.86%)	74 (10.28%)
Stepwise	12 (2.82%)	1 (6.25%)	5 (3.82%)	1 (1.32%)	0 (0.00%)	19 (2.64%)
Logistic	8 (1.88%)	0	2 (1.53%)	0 (0.00%)	1 (1.41%)	11 (1.53%)
Hierarchical linear	15 (3.52%)	0	4 (3.05%)	1 (1.32%)	3 (4.23%)	23 (3.19%)
HLM	9 (2.11%)	0	0	0	8 (11.27%)	17 (2.36%)

Table 3. Continued

Method	Journal					Total
	C&E	*ijSCL*	*BJET*	*AJET*	*ETR&D*	
Data reduction						
EFA	23 (5.40%)	0	9 (6.87%)	4 (5.26%)	0	36 (5.00%)
PCA	31 (7.28%)	0	6 (4.58%)	6 (7.89%)	1 (1.41%)	44 (6.11%)
Unspecified	2 (0.47%)	0	1 (0.76%)	2 (1.32%)	0	5 (0.69%)
Structural Equation Modeling						
Path analysis	19 (4.46%)	0	6 (4.58%)	8 (10.53%)	0	33 (4.58%)
CFA / measurement model	52 (12.21%)	0	11 (8.40%)	7 (9.21%)	4 (5.63%)	74 (10.28%)
Structural model	32 (7.51%)	0	7 (5.34%)	2 (2.63%)	0	41 (5.69%)
Nonparametric statistics						
Pearson's X^2	46 (10.80%)	6 (37.50%)	12 (9.16%)	9 (11.84%)	7 (9.86%)	80 (11.11%)
Mann-Whitney U	28 (6.57%)	0	7 (5.34%)	1 (1.32%)	0	36 (5.00%)
Cluster analysis	15 (3.52%)	1 (6.25%)	4 (3.05%)	1 (1.32%)	3 (4.23%)	24 (3.33%)
Wilcoxon signed ranks	18 (4.23%)	1 (6.25%)	4 (3.05%)	0	0	23 (3.19%)
Kolmogorov-Smirnoz test	10 (2.35%)	0	1 (0.76%)	1 (1.32%)	1 (1.41%)	13 (1.81%)
Kruskal-Wallis test	8 (1.88%)	0	2 (1.53%)	1 (1.32%)	0	11 (1.53%)
Fisher's exact test	4 (0.94%)	0	0	0	0	4 (0.56%)
Anderson-Darling test	2 (0.47%)	0	2 (1.53%)	0	0	4 (0.56%)
Friendman's X^2	2 (0.47%)	0	0	0	0	2 (0.28%)
Kendall's W	0	0	0	2 (2.63%)	0	2 (0.28%)
McNemar's test	2 (0.47%)	0	0	0	0	2 (0.28%)
Sign test	1 (0.23%)	0	0	0	0	1 (0.14%)
Density Kernel estimation	1 (0.23%)	0	0	0	0	1 (0.14%)

Table 3. Continued

Method	Journal: *C&E*	*ijSCL*	*BJET*	*AJET*	*ETR&D*	Total
Unspecified non-parametric test	1 (0.23%)	0	0	0	0	1 (0.14%)
Miscellaneous procedures						
Frequencies	259 (60.80%)	14 (87.50%)	79 (60.31%)	55 (72.37%)	42 (59.15%)	449 (62.36%)
Average variance extracted (AVE)	30 (7.04%)	0	7 (5.34%)	0	1 (1.41%)	38 (5.28%)
Levene's test	24 (5.63%)	0	2 (1.53%)	2 (2.63%)	5 (7.04%)	33 (4.58%)
Shapiro–Wilk test	7 (1.64%)	0	3 (2.29%)	1 (1.32%)	1 (1.41%)	12 (1.67%)
Power analysis	4 (0.94%)	0	1 (0.76%)	2 (2.63%)	5 (7.04%)	12 (1.67%)
Multicollinearity test	5 (1.17%)	0	2 (1.53%)	0	1 (1.41%)	8 (1.11%)
Social network analysis	4 (0.94%)	0	3 (2.29%)	0	1 (1.41%)	8 (1.11%)
Homogeneity-of-regression test	6 (1.41%)	0	1 (0.76%)	1 (1.32%)	0	8 (1.11%)
Mediation analyses	5 (1.17%)	2 (12.50%)	0	0	0	7 (0.97%)
Unspecified homogeneity of variance	5 (1.17%)	0	0	2 (2.63%)	0	7 (0.97%)
Bayesian statistics	4 (0.94%)	1 (6.25%)	0	0	1 (1.41%)	6 (0.83%)
Sequential analysis	3 (0.70%)	0	2 (1.53%)	0	0	5 (0.69%)
Wald test	3 (0.70%)	0	1 (0.76%)	0	1 (1.41%)	5 (0.69%)
Box's test	2 (0.47%)	0	1 (0.76%)	0	1 (1.41%)	4 (0.56%)
Item difficulty	3 (0.70%)	0	1 (0.76%)	0	0	4 (0.56%)
Item discrimination	2 (0.47%)	0	1 (0.76%)	0	0	3 (0.42%)
Square root of AVE	2 (0.47%)	0	0	1 (1.32%)	0	3 (0.42%)
Pillai's Trace	2 (0.47%)	0	0	0	1 (1.41%)	3 (0.42%)
Little's MCAR test	1 (0.23%)	0	0	1 (1.32%)	0	2 (0.28%)

Table 3. Continued

Method	Journal					Total
	C&E	*ijSCL*	*BJET*	*AJET*	*ETR&D*	
Sobel test	1 (0.23%)	0	0	0	1 (1.41%)	2 (0.28%)
ROC curves	1 (0.23%)	0	1 (0.76%)	0	0	2 (0.28%)
Fisher's Z-transformation analyses	1 (0.23%)	0	1 (0.76%)	0	0	2 (0.28%)
Q-Q plot	1 (0.23%)	0	0	0	1 (1.41%)	2 (0.28%)
Analytic Hierarchy Process	1 (0.23%)	0	0	0	1 (1.41%)	2 (0.28%)
Generalized Estimating Equation	1 (0.23%)	0	0	0	1 (1.41%)	2 (0.28%)
F-test	2 (0.47%)	0	0	0	0	2 (0.28%)
Lawshe's content validity ratio	2 (0.47%)	0	0	0	0	2 (0.28%)
Welch's t-test	2 (0.47%)	0	0	0	0	2 (0.28%)
Mauchly's test	2 (0.47%)	0	0	0	0	2 (0.28%)
Mardia test	2 (0.47%)	0	0	0	0	2 (0.28%)
Bartlett's sphericity X^2 test	2 (0.47%)	0	0	0	0	2 (0.28%)
Unspecified linearity test	2 (0.47%)	0	0	0	0	2 (0.28%)

Note. ANOVA = analysis of variance; ANCOVA = analysis of covariance; MANOVA = multivariate analysis of variance; MANCOVA = multivariate analysis of covariance; HLM = hierarchical linear modeling; EFA = exploratory factor analysis; PCA = principal components analysis; CFA = confirmatory factor analysis. The following miscellaneous statistical methods were used only once: adjusted residuals tables (Z-score) (*C&E*), Box-Cox transformation (*C&E*), Carte & Russell's F-statistic (*C&E*), Cox–Stuart test (*C&E*), Curve estimation (*C&E*), Durbin–Watson-coefficient (*C&E*), Fuzzy Set Analysis (*C&E*), Harman's single factor test (*C&E*), Heterogeneity test: I^2 statistic (*C&E*), Holm-Bonferroni correction (*C&E*), Homogeneity test: Q statistic (*C&E*), Hosmer and Lemenshow Test of Reliability (*C&E*), Hotellings T^2 statistic (*C&E*), IRT-based estimate (*C&E*), Johnson–Neyman method (*C&E*), Kruskal–Wallis test with post hoc, Mann–Whitney U-test (*C&E*), Lagrange multiplier test (*C&E*), Likelihood ratio test (*C&E*), Mahalanobis distance (*C&E*), Mathieu's significance test (*C&E*), Multidimensional scaling (*C&E*), Multivariate normality (*C&E*), Ordinal logistic regression (*C&E*), Prediction interval (C&E), Prediction relevance (*C&E*), Propensity (*C&E*), Sensitivity analysis (*C&E*), Stochastic Frontier Regression (*C&E*), Unspecified normality test (*C&*E), Gini coefficient (*ijSCL*), log-likelihood ratios (G-tests) with post hoc H tests (*BJET*), Parallel analysis (*BJET*), post-hoc General Linear Hypothesis (*BJET*), Somer's d test (*BJET*), Tabachnick and Fidell's ratio validation (*BJET*), Unspecified heteroscedasticity test (*BJET*), Variance-stabilizing transformation (*BJET*), Jonckheere's test (*AJET*), Replication procedures (*AJET*), Bass modeling (*ETR&D*), Constant growth rate modeling (*ETR&D*), Factor score (*ETR&D*), F- test with Bonferroni correction (*ETR&D*), Ordinal Consistency (*ETR&D*), Proportional growth rate modeling (*ETR&D*), Stratified growth rate modeling (*ETR&D*), and Tversky's formula (*ETR&D*).

fully understand the importance of your findings, it is *almost always* necessary to include some index of effect size" (p. 25, emphasis added). More recently, the current APA manual (2010) embraces similar language encouraging researchers to report effect sizes and their accompanying confidence intervals.

Despite scholars stressing the importance of reporting effect sizes, only 49.58% of the articles (357 articles) using quantitative or mixed methods included at least one effect size. The current finding is lower than the findings from Randolph et al.'s, (2008) review, where they found that effect sizes were reported in 97.6% (120 articles) of computer science education research. In the current study, effect sizes appeared in 38.17% of the quantitative or mixed methods articles published in *BJET* (50 articles), 39.47% of the quantitative or mixed methods articles published in *AJET* (30 articles), 52.11% of the quantitative or mixed methods articles published in *C&E* (222 articles), 61.97% of the quantitative or mixed methods articles published in *ETR&D* (44 articles), and 68.75% of the quantitative or mixed methods articles published in *ijSCL* (11 articles),. Table 4 presents the types of effect sizes appearing in educational technology research between 2012 and 2013. The most commonly reported effect size in the literature reviewed was R^2, which was found in 128 articles (17.78%) followed by Cohen's *d* (97 articles; 13.47%). Other popular effect sizes included: η^2 (60 articles; 8.33%), partial η^2 (53 articles; 7.36%), variance explained (48 articles; 6.67%), adjusted R^2 (31 articles; 4.31%), unspecified effect size (24 articles; 3.33%), r^2 (14 articles; 1.94%), Cramer's phi (13 articles; 1.81%), and Cohen's f^2 (8 articles; 1.11%). Other effect sizes were found in less than one percent of the articles and are listed in Table 4.

Occasionally, authors used Cohen's (1992) benchmarks to interpret their effect size. However, this is not good practice. As Cohen (1988) himself noted,

> These proposed conventions were set forth throughout with much diffidence, qualifications, and invitations not to employ them if possible… They were offered as conventions because they were needed in a research climate characterized by a neglect of attention to issues of [effect size] magnitude. (p. 532)

Cohen suggested that a Cohen's d less than .2 are small, a Cohen's d greater than .5 are medium, and a Cohen's d greater than .8 are large. Among the quantitative and mixed method articles reviewed, authors of 72 articles (10.00%) used Cohen's benchmarks to interpret effect sizes. Of these, 39 articles were found in C&E, 3 were found in *ijSCL*, 9 were found in *BJET*, 14 were found in *AJET*, and 7 were found in *ETR&D*.

t Tests and -OVA Methods

The *t* test examines whether the means of two groups are statistically different. T-tests were the most common statistical method found in the educational technology literature, after means and standard deviations. T-tests appeared in 36.25% of all quantitative or mixed methods articles. Moreover, ANOVA, which compares multiple group means to determine if one or more are statistically significantly different from the other(s), was the second most common inferential statistic among educational

Table 4.
Effect Sizes Reported in Education Technology Journals Organized by Journal

	Journal					
Effect Size	*C&E*	*ijSCL*	*BJET*	*AJET*	*ETR&D*	Total
R^2	89	1	20	9	9	128
Cohen's d	54	6	12	9	16	97
η^2	30	3	9	8	10	60
partial η^2	33	2	5	2	11	53
Variance explained	37	0	5	4	2	48
adjusted R^2	19	0	6	2	4	31
Unspecified effect size	16	0	2	3	3	24
r^2	10	0	1	1	2	14
Cramer's phi	8	2	1	1	1	13
Cohen's f^2	5	0	1	0	2	8
Pseudo-R^2	4	0	0	0	0	4
ω^2	2	0	0	0	0	2
Odds ratio	1	0	1	0	0	2
Hedges's g	1	0	0	0	1	2
Generalized Eta-squared (η^2_G)	1	0	0	0	0	1
Squared semi-partial correlations (sR^2)	1	0	0	0	0	1
Canonical R^2/ R	0	0	2	0	0	2
κ^2	0	1	0	0	0	1
partial R^2	0	0	1	0	0	1

technology researchers, with 30.00% of articles using ANOVA. Analysis of covariance (ANCOVA), which is the same as an ANOVA, except with a covariate included in the statistical model, was found in 61 articles (8.47%) and was therefore not commonly used. The multivariate extensions of ANOVA and ANCOVA, respectively, multivariate analysis of variance (MANOVA) and multivariate analysis of covariance (MANCOVA), were less common than their univariate counterparts. MANOVA was used in 5.42% articles, while MANCOVA was found in only 11 articles (1.53%).

Post hoc tests. When an OVA method results in a statistically significant outcome, and there are three or more groups, researchers often follow with post hoc tests to determine exactly where the difference(s) in groups exist. For ANOVA, the most common post hoc tests were the Tukey's Honestly Significant Difference test (HSD) (28 articles), Scheffé post hoc (19 articles), post hoc t tests (13 articles), Bonferroni post hoc analysis (11 articles), and Fisher's least significant differences (LSD; 11 articles). Paired-sample t-tests with Bonferroni correction, Dunnett's t, Tamhane's multiple comparison tests, Games-Howell post hoc tests, and Welch's t-test were each found in a single article. Authors of 16 articles indicated they conducted post hoc tests after their ANOVA, yet did not specify which post hoc test(s) were performed. The most common post hoc tests for ANCOVA were the Bonferroni post hoc analysis, Fisher's LSD, and unspecified post hoc tests, with each of these found in 4 articles. Sidaks post-hoc test, Tukey's HSD test, and Dunn's procedure with Bonferroni adjustment were each found in one article. A series of conventional post hoc t-tests were found in 2 articles.

Considerably the most common post hoc statistical test for MANOVA was a series of ANOVAs, which occurred in 13 articles – one-third of all articles where at least one MANOVA was performed. This is unacceptable practice, because MANOVA and ANOVA test completely different hypotheses. As Tatsuoka (1971) noted, "one would usually be well advised to follow up a significant multivariate test with a [descriptive] discriminant function analysis in order to study the nature of the group differences more closely" (p. 50).

The other specified post hoc tests after MANOVA were Tukey's HSD test (3 articles), Fisher's LSD (2 articles), t-test (1 articles), Post-hoc Bonferroni corrected comparisons, (1 articles), ANOVA with post hoc Scheffé (1 article), and discriminant analysis (1 article). Authors of three articles who conducted multivariate null hypothesis statistical significance testing did not specify which post hoc test(s) they conducted. The most common post hoc test for MANCOVA included a series of ANCOVAs (2 articles) or a series of ANOVAs (2 articles). Authors of one article conducted a post hoc test after the MANCOVA, yet did not specify which post hoc test was performed.

Correlational and Regression Procedures

The most commonly used correlational procedure is Pearson's *r*, which examines the statistical relationship between two interval or ratio scaled variables – appearing in 235 articles (32.64%). Other correlational techniques were rare: Spearman's ρ was found in 26 articles (3.61%), point–biserial and phi correlations were each found in three articles (0.42%), Kendall's tau was found in two articles (0.28%), and canonical and point-measure correlations were each found in a single article (0.14%). Multiple regression was the most commonly found regression procedure and appeared in 74 articles (10.28%). As with the correlational procedures, other regression procedures were much less common: hierarchical linear regression was found in 23 articles (3.19%), stepwise regression was used in 19 articles (2.64%), and logistic regression was found in 11 articles (1.53%).

Complex Statistical Methods

Complex statistical methods are more accessible to researchers than ever before. Closed- and open-source statistical programs offer researchers statistical tools to test more complex models. This section reports on the impact complex statistical methods have had on educational technology research.

Hierarchical linear modeling. HLM allows researchers to analyze nested data such as data represented by three nesting levels (e.g., student, college, and university). HLM is rarely found in educational technology literature. The method was used in 17 articles (2.36%), which were all published in *C&E* or *ETR&D*.

Data reduction methods. Data reduction methods create latent variables, which summarize a set of observed variables in a more parsimonious manner (Thompson, 2004). There are two types of data reduction, exploratory factor analysis (EFA) and principal components analysis (PCA). I found that educational technology researchers favor both methods similarly well: EFA was used in 36 articles (5.00%), and PCA was used in 44 articles (6.11%). Authors of five articles (0.69%) used some type of data reduction method but did not specify which type of data reduction method was used.

A variety of subjective decisions come into play when conducting a data reduction method (Thompson, 2004). One of the first decisions is selecting an extraction method to obtain a factor structure based on assumptions made about the data. Principal axis factoring was the most common method used in educational technology research, with 11 articles found. Other extraction methods found were maximum likelihood (ML; 4 articles), and unweighted least square (1 article).

Another decision when conducting an EFA or PCA is to determine the number of factors to retain. Often authors use more than one factor retention method when making their decision and I found educational technology researchers often used more than one factor retention method when they reported the factor retention method used. The most commonly specified method for determining the number of factors was the Guttman rule (also called the K1 rule, or eigenvalues-greater-than-one rule), appearing in 35 articles. The other factor retention methods used were the Kaiser–Meyer–Olkin test of sampling adequacy (29 articles), visual inspection of the pattern coefficients (22 articles), Bartlett's test of sphericity (22 articles), the variance accounted by the factors (11 articles), scree test (8 articles), parallel analysis (5 articles), and communality coefficients (3 articles). Moreover, researchers used reliability coefficients, component matrices, homogeneity of proportions test, critical ratios, and their own ability to interpret different factor solutions to determine the number of factors to retain, with each found once in the educational technology literature.

A third decision is to choose a rotation method to produce more interpretable results. Varimax rotation was the most common method used – appearing in 31 articles. The next most common methods specified were varimax with Kaiser normalization (7 articles), oblimin (4 articles), promax (5 articles), direct oblimin (2 articles), quartimin (1 article). In six articles, the author(s) did not specify the rotation method performed.

After an EFA or PCA is performed the results are outputted in the form of a factor pattern matrix if an orthogonal rotation is used or a structure matrix if a non-orthogonal rotation is used. Among the 85 articles that used a data reduction method, authors of 39 articles reported a factor pattern matrix, while only authors of 5 articles reported structure matrices. Of course, factor pattern coefficients equal the factor structure coefficients if the factors are rotated orthogonally (Thompson, 2004).

Structural equation modeling. Structural equation modeling (SEM) is a family of statistical methods which extend beyond the constraints of the statistical methods underlying the general linear model (Kline, 2011). Note that researchers could use more than one SEM method: path analysis, CFA/measurement model, or structural model.

Among the quantitative and mixed method articles reviewed, 101 articles (14.03%) utilized at least one SEM method. Path analysis compares measured variables to other measured variables. Path analysis was seldom used in the recent educational technology literature. Only 33 articles (4.58%) from the last two years used path analysis. While path analysis compares measured variables, structural models compare latent variables. Structural models were more common than path analysis, appearing in 41 articles (5.69%) from the past two years. Confirmatory factor analysis (CFA) and measurement models compare measured variables to latent variables. CFA is the most common advanced statistical method used in educational technology with CFA/measurement models appearing in 74 articles (10.28%). Tests of invariance or result replicability that are often used SEM and CFA were performed in only 1 article (0.14%).

As with data reduction methods, SEM requires researchers to report a substantial amount of details so that their procedures can be properly evaluated. One reporting requirement is for researchers to report a covariance matrix or a correlation matrix with accompanying means and standard deviations of all observed variables in a model (Kline, 2011). Among the 101 articles utilizing SEM, authors of merely 27 articles reported a correlation matrix with accompanying means and standard deviations, while authors of only two articles reported a covariance matrix. Furthermore, authors of 28 articles reported a correlation matrix without accompanying means and standard deviations. By only reporting a correlation matrix, researchers cannot properly evaluate SEM procedures.

Kline (2011) also recommended that researchers report the estimation method used by their statistical software package when conducting SEM. The most common estimation method was maximum likelihood (21 articles), followed by partial least squares (15 articles), weighted least squares with missing variables (WLSMV; 3 articles) and robust ML (2 articles). Standardized robust ML, full information maximum likelihood (FIML), maximum likelihood with robust standard errors (MLR), ordinary least squares, and weighted least squares were each found in one article. Furthermore, multivariate normality is assumed when using ML estimation in SEM (Kline, 2011). Therefore, researchers who use ML estimation should either check their data for violations of the normality assumption or use methods that compensate for a lack of normality. Authors of only 15 out of 101 articles using SEM reported that they examined the normality of their data.

Another reporting requirement of SEM, is for researchers to report parameter estimates. Of the 101 articles that used SEM, parameter estimates were reported for 80 articles. Of these, 72 were standardized estimates and 8 included unstandardized estimates.

Lastly, researchers who use SEM have a wide variety of fit statistics with which they can evaluate how closely their observed data fit their SEM. The most common fit statistic reported was the root mean square error of approximation (RMSEA; 66 articles) followed by the comparative fit index (CFI; 61 articles), χ^2/df (44 articles), χ^2 (36 articles), standardized root mean square residual (SRMR; 30 articles), goodness-of-fit index (GFI; 29 articles), normed fit index (NFI; 28 articles), adjusted goodness-of-fit index (AGFI; 23 articles), Tucker–Lewis index (TLI; 21 articles), nonnormed fit index (NNFI; 16 articles), incremental fit index (IFI; 10 articles), root mean residual (RMR; 8 articles). The Akaike's information criterion (AIC), expected cross-validation index (ECVI), global goodness-of-fit index (GoF), parsimony goodness of fit index (PGFI), and parsimony normed fit index (PNFI) were reported in few articles, each appearing in 2 articles. Other reported fit statistics, which were reported only once, were the Bayesian information criterion (BIC), normal theory weighted least squares χ^2, parsimonious comparative fit index (PCFI), relative fit index (RFI), and the Satorra–Bentler χ^2.

Nonparametric Statistics

Nonparametric statistics do not require assumptions of normality and associated probability distributions, which the aforementioned parametric methods do require. Nonparametric methods are more appropriate than parametric when assumptions are not met. The most common nonparametric null hypothesis statistical significance testing method was the Pearson's X^2, which was found in 80 articles (11.11%), followed by the

Mann–Whitney *U* test (36 articles; 5.00%), cluster analysis (24 articles; 3.33%), Wilcoxon signed ranks (23 articles; 3.19%), Kolmogorov-Smirnoz test (13 articles; 1.81%), and Kruskal-Wallis test (11 articles; 1.53%). Moreover, Fisher's exact test and Anderson-Darling test were each found 4 articles (0.56%). Friendman's X^2, Kendall's W, and McNemar's test were each found 2 articles (0.28%). Density Kernel estimation, sign test, and an unspecified non-parametric test were each found in 1 article (0.14%) and all were published in *C&E*.

Miscellaneous Statistical Procedures

There were a variety of other statistical procedures, which did not easily fit into other sections of the current article. Note that some of these miscellaneous statistics are used in conjunction with other statistical methods described elsewhere in the present article. For example, the most common miscellaneous statistics reported were frequencies, which were found in 449 articles (62.36%). The second most common miscellaneous statistics reported were average variance extracted (AVE), which were found in 38 articles (5.28%) and mostly in the context of validity. Moreover, the Levene's test (33 articles, 4.58%) was the second most common and was often used apriori to parametric statistical methods to examine the homogeneity of variances assumption in an ANOVA or a *t* test. Shapiro–Wilk tests and power analyses each occurred in 12 articles (1.67%), while multicollinearity tests, social network analyses, and homogeneity-of-regression tests each appeared in 8 articles (1.11%). Other miscellaneous statistics reported were mediation analyses (7 articles; 0.97%), unspecified types of homogeneity of variance tests (7 articles; 0.97%), Bayesian statistics (6 articles; 0.83%), sequential analysis (5 articles; 0.69%), Wald test (5 articles; 0.69%), Box's test (4 articles; 0.56%), Item difficulty (4 articles; 0.56%), Pillai's Trace (3 articles; 0.42%), square root of AVE (3 articles; 0.42%), and item discrimination (3 articles; 0.42%). The following miscellaneous statistics were each found in two articles (0.28%): Little's MCAR test, Sobel test, ROC curves, Fisher's Z-transformation analyses, Q-Q plot, analytic hierarchy process, and generalized estimating equation (GEE). Welch's t-test, Lawshe's content validity ratio, Mauchly's test, Mardia test, Bartlett's sphericity X^2 test, F-test, and an unspecified linearity test also appeared in two articles (0.28%); however these statistics were only reported in *C&E*. Miscellaneous statistical methods that were used in only one article are listed in the footnotes of Table 3.

P values. Another miscellaneous reporting standard examined was how often researchers reported at least one exact p value and/or inexact p value. According to APA guidelines (2010), "exact p values" are defined as p values that are equal to a specific number and not a range (e.g., $p < .01$ is defined as an inexact p). The exceptions to this guideline were p values permitted in the APA manual (e.g., if p is less than .001, or if p is in a table). Overall 72.22% of articles reported at least one exact p value, whereas 46.94% of articles reported at least one inexact p value. There were many authors who reported a both exact and inexact p values. At times, authors noted that p was equal or less than zero. Reported values of $p = .000$ or $p < .000$ are impossible. However, 160 articles (22.2%) reported a p value equal to or less than zero.

Confidence intervals. Researchers also recommended that authors report confidence intervals (CI). Few authors reported CIs, which appeared in only 69 articles (9.58%). All five journals published articles with CIs with 46 articles (10.80%) published in *C&E* 3 articles (18.75%) published in *ijSCL*, 8 articles (6.11%) published in *BJET*, 7 articles (9.21%) published in *AJET*, and 5 articles (7.04%) published in *ETR&D*.

Discussion

The current study found that quantitative methods continue to dominate the field as a whole, yet journals appear to favor certain research methods over others. As demonstrated in Figure 1, *C&E* and *ETR&D* favor quantitative methods more than other journals, while *ijSCL* favors qualitative methods. *AJET* has the most balanced representation of research methodologies than any other journal, by only slightly favoring qualitative methods. *BJET* has more non-empirical articles than any other journal – mostly due to numerous book reviews in each issue.

The choice of research method should be based on the researcher's questions. For example, the *International Journal of Computer-Supported Collaborative Learning (ijSCL)* focuses on in-depth collaboration within educational technology. Because of this, one would expect *ijSCL* to utilize more qualitative methods compared to general educational journals such as *BJET* or *AJET*. Additionally, researchers recently discussed the commonalities between quantitative and qualitative research, and advocated for mixed method research (e.g., Johnson & Onwuegbuzie, 2004). As reported in Figure 1, the current study suggests educational technology researchers are utilizing mixed methods approaches.

Previous publications suggested that convenience samples plague the field, yet note that sampling methods are often not directly reported (Alper & Gülbahar, 2009; Edyburn, 2000; Randolph et al., 2008). As with previous reviews of educational technology literature and as reported in Table 1, most authors did not report the type of sampling procedure used in their investigations. However, in most cases the descriptions of their participants suggested the sample was collected out of convenience. Nevertheless, convenience samples should not be considered poor research practice. Convenience samples are as helpful as other sampling methods. However, when convenience samples are used, authors should adequately describe the sample for the reader to contextualize findings appropriately. In brief, authors should provide enough information about participants so that readers can determine generalization parameters.

As reported in Table 2, the current study found poor reporting practices just as Lee et al., (2004), Lee et al., (2007) and Randolph (2008) found. The lack of reporting reliability estimates continues to degrade educational research. Over 30 years ago, Willson (1980) observed,

> Only 37% of the AERJ studies explicitly reported reliability coefficients for the data analyzed. Another 18% reported only indirectly through reference to earlier research… [and] unreported [reliability coefficients]

> in almost half the published research is… inexcusable at this late date. (p. 8-9)

In the current study, approximately half of the authors reported reliability of the scores produced by their instrument, while fewer authors invoked a reliability coefficient from a previous study or test manual. Surely by now, the lack of reporting reliability coefficients for the data at hand is inexcusable. Accordingly, Wilkinson et al. (1999) stated, "if a questionnaire is used to collect data, [a researcher should] summarize the psychometric properties of its scores with specific regard to the way the instrument is used in a population" (p. 596). Authors should report evidence that the *scores* they are analyzing are reliable. I emphasize scores in the previous sentence because authors sometimes referred to reliability as an attribute of the test, instead of data at hand, by using phrases such as "reliability of the test." Verbiage relating reliability to a test instead of scores, suggests there may be misconceptions about reliability (Ritter, 2010). For a more thorough explanation about the importance of score reliability see Thompson (1992, 2003). Nevertheless, the lack of reliability information is a clear weakness in the body of educational technology research.

The results of the current review suggest few authors report informationally-adequate statistics. Certain statistics should be reported to effectively evaluate findings and to allow meta-analysts to conduct secondary research studies. Although the current study identifies all statistical techniques used in educational technology between 2012 and 2013, two specific areas of concern are emphasized.

One area of concern is the tendency to report a mean without the SD about the mean. As reported in Table 3, among the authors of 566 articles reporting means, SDs were not reported for 68 of the articles. In a similar vein, authors tended to report means of the instrument's items instead of variables used when performing SEM. The lack of reporting means and SDs of the variables used in the SEM hinders the ability of researchers in secondary studies to synthesize results from SEM analysis appropriately. Authors should always report means with their accompanying SDs and report means and SDs of the variables used when performing SEM.

Another area of concern is the lack of reporting correlation matrices with accompanying means and standard deviations or covariance matrices. Few authors reported matrices and often only provided a correlation matrix. Correlation matrices with accompanying means and SDs or covariance matrices should always be reported when conducting data reduction methods or SEM. These matrices of associations are used to analyze data and can be used to replicate or expand analyses (Zientek & Thompson, 2009). For an example on how to report a correlation matrix with accompanying means and SDs in a research study using SEM, see Merchant, Goetz, Keeney-Kennicutt, Kwok, Cifuentes and Davis (2012).

Limitations

As with any study, there are some limitations to the present study. First, while previous methodological reviews of educational technology are limited to specific areas, the current study aimed to review the field as a whole by examining the top five educational journals between 2012 and 2013. Given there are at least 423 educational technology journals (cf., *2012 Cabell's Directories of Publishing Opportunities*) and only five, high-impact journals were selected, the current study is limited in representing the entire field. By selecting the top five educational technology journals, the most impactful literature to the field was reviewed. Moreover, the top five journals ensure the level of quality research under review in a field with numerous publication outlets. Nevertheless, given that educational technology is a global issue, future studies should examine other publications as well as publications in languages other than English.

Despite efforts to ensure the quality of the coding procedures, at times subjective judgment in classifying articles and interpreting authors' descriptions of their methodological and statistical procedures was required. Thus, if other researchers conducted a similar review, they would likely produce somewhat different results. In fact, Skidmore and Thompson (2010) found that occasionally in other reviews when the same journal volumes of journals were analyzed by different authors different results were produced. Nevertheless, by having two coders and demonstrating sufficient inter-rater agreement (kappa = 0.99), suggests that such problems were minimal in the present study.

Implications

Although the nature of the current study is descriptive, the results can provide authors, editors, and reviewers with some insightful ideas about the publishing trends in high impact educational technology journals. The results presented here can also help identify some of the strengths and weaknesses of the current methods used in educational technology research. Additionally, decision-makers in educational technology doctoral programs can use this information to determine the types of statistical techniques their graduate students need to interpret and conduct research in the field.

The discussion section reported on what I consider to be the most important evidence-based recommendations for improving the current state of the educational technology literature. Educational technology researchers should report the type of sampling procedure they used and provide adequate description of their samples so that readers can determine generalizability. Moreover, educational technology researchers should report evidence that the *scores* they are analyzing are reliable because of the possible impact reliability has on the interpretation of research results. Lastly, researchers should report informationally-adequate statistics so that readers can evaluate findings appropriately.

3. TECHNOLOGY ACCEPTANCE MODEL OF ONLINE LEARNING MANAGEMENT SYSTEMS IN HIGHER EDUCATION: A META-ANALYTIC STRUCTURAL EQUATION MODEL

Technology adoption practices have received considerable attention in the last five years. With more funding offered for technology integration and implementation in a time when other funding is cut, universities are looking to online learning as a cost-effective option to deliver instruction. However, little is known about whether undergraduate learners will readily accept an online learning environment. In this technology age, many practitioners consider students proficient in technology. This assumption often stems from students' fluency with social media and entertainment media. However, researchers previously demonstrated that proficiently using technologies in personal and social settings does not necessarily transfer to the technology skills needed in an academic setting (Goode, 2010; Lloyd, Dean, & Cooper 2007; Presley & Presley, 2009; Teclehaimanot & Hickman, 2011).

Practitioners need to gain an understanding of students' acceptance of the online learning environment, in addition to the instructor's preferences in delivering instruction. Instructors' preferences are often taken into consideration before implementing online learning. However, students' preferences are often explored only after adoption or when issues emerge during the course. Both parties' acceptance of the online learning environment is crucial to the success of online learning programs and for funds to be wisely invested. Before investing in online learning technologies, practitioners should determine whether the online learning environment will be accepted by all the parties involved.

The Technology Acceptance Model

The Technology Acceptance Model (TAM; Davis, Bagozzi, & Warshaw, 1989) is one of many underlying theories used in technology adoption. The TAM is one of the most commonly used models to explain user's technology acceptance behavior. The TAM is rooted in Social Psychology Theory and the Theory of Reasoned Action. The core constructs in the original TAM include perceived ease of use (EU), perceived usefulness (PU), and behavioral intention to use (BI). Over time, the model has been modified by adding constructs such as attitude toward using (A) and actual system use (U), as noted in Figure 2. Note, the TAM also specifies relationships between numerous endogenous variables (i.e., predictor variables) and other variables within the model.

The TAM posits perceived ease of use and perceived usefulness of the technology will individually predict user's behavioral intention to use the technology. In other words, the easier the technology is to use or the more useful a particular technology is found to be, the more likely the user intends to use the technology again. The TAM also proposes that perceived usefulness mediates the relationship between perceived ease of use and behavioral intention to use the technology. This mediation effect may be observed when a technology is easy to use, but the technology is not useful to a person. If the technology is not perceived as useful, then it does not matter how easy the technology is to use; the end user will not continue to use the technology.

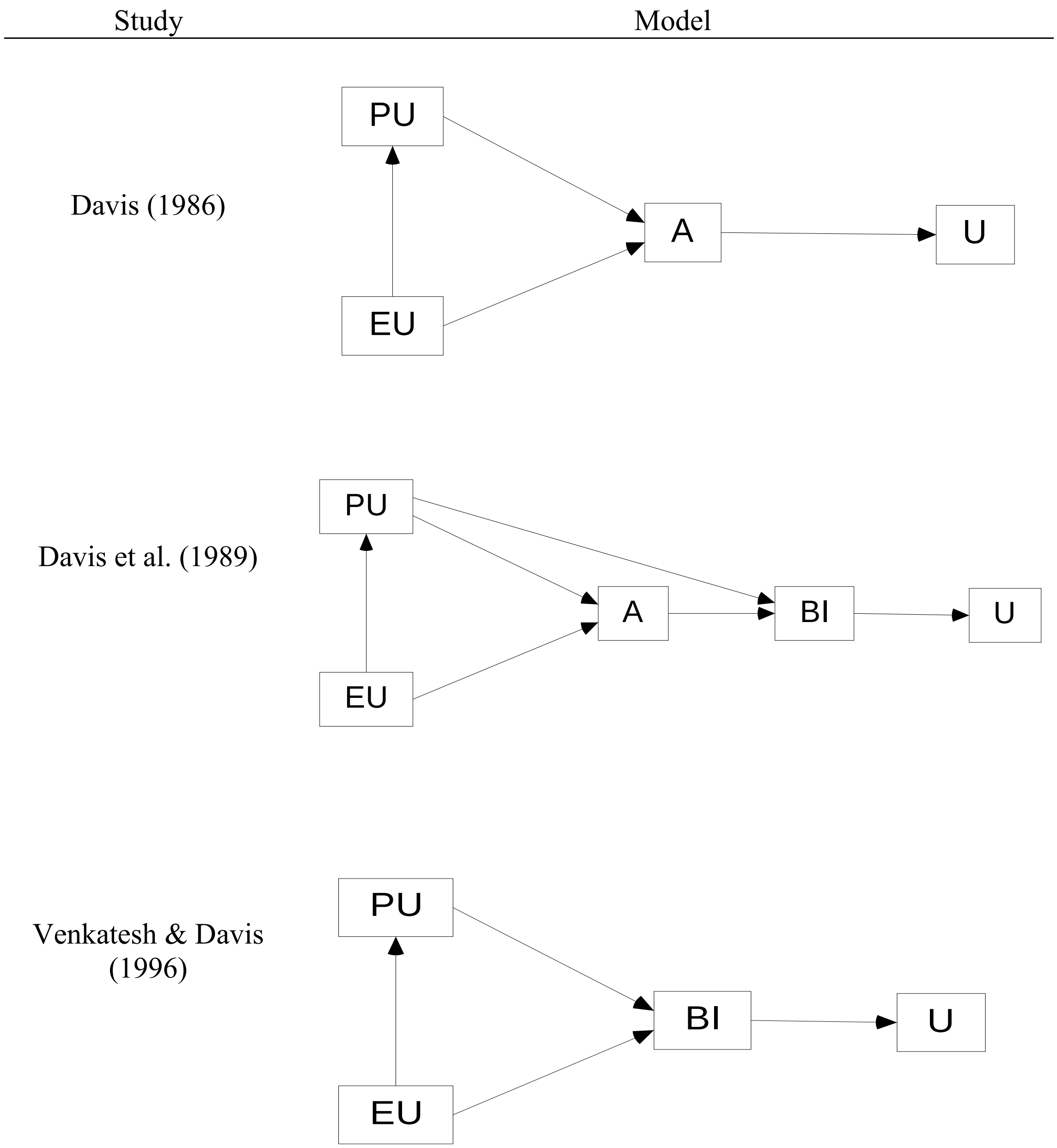


Figure 2 Evolution of Core Constructs in the Technology Acceptance Model

Literature Review

Meta-analysts have faced numerous challenges in validating the TAM. One challenge meta-analysts face is the inability to conduct moderator analyses relating to a specific type of technology used. For example, Schepers and Wetzels (2007) reviewed all empirical studies assessing the TAM. The authors had sufficient information to code for four types of technologies. However, there were not enough studies to a conduct moderator analysis based on a single technology. Instead, the meta-analysts converted the previous four categories into two categories to conduct a less informative moderator analysis.

Another challenge meta-analysts face is the inability to validate the TAM with a specific population. For example, King and He (2006) conducted a meta-analysis on the TAM using different users (e.g., students, professionals, and general users) and found differences between types of users. More specifically, King and He (2006) concluded that, although students were similar to professionals, students were "not exactly like either of the other two groups" (p. 751) (i.e., professionals and general users).

The failure to explore technology adoption with a specific type of technology in a single population limits the meta-analyst's ability to explore subjective norms, such as culture. Previous meta-analysts (Schepers & Wetzels 2007; Straub, Keil, & Brenner, 1997) have demonstrated the impact of subjective norms, and that the type of technology influences adoption practices using the TAM. Schepers and Wetzels (2007) findings confirmed Straub et al., (1997) who found that the TAM did not fit equally well across cultures.

Previous studies have used meta-analytic techniques to validate the TAM; however, researchers failed to explore the TAM with a single type of technology or among a specific population using the TAM (King & He, 2006; Ma & Liu, 2004). The results from prior meta-analyses suggest a lack of understanding of a specific population's ability to accept a specific type of technology.

Purpose

Before investing in online learning technologies, decision-makers should determine whether online learning environment will be accepted by all the parties involved. Most studies have focused on multiple parties and their relationship to technology. The present study, however, and in contrast to previous meta-analyses, isolated one population and one technology. For this reason, the purpose of the current study was to determine whether the core variables of the TAM explains undergraduates' acceptance of online learning.

Methods

Eligibility Criteria

Articles included in the current meta-analysis met a set of certain criteria. The article must have been published after Davis (1986) proposed the TAM. Any type of article was open for inclusion to avoid publication bias, including book chapters and papers presented at conferences. Articles included in the research synthesis had to meet eight criteria; the study had to: 1) be written in English, 2) report quantitative results, 3) test the TAM, 4) include samples which came from an undergraduate student population, 5) use the technology in an online learning context, 6) report adequate statistics to calculate covariances, 7) measure variables in the TAM, and 8) measure variables at one time point (e.g., longitudinal studies were excluded). Note, mobile learning devices were excluded from the current study. Online learning included fully online courses or blended courses (e.g., partially face-to-face and partially online).

Search Procedures

First, I found articles using three academic databases: 1) ERIC database via EBSCO Host, 2) Educational Full Text via Wilson Web, and 3) Proquest Dissertations & Theses database via ProQuest. I used similar thesaurus terms and keywords across all three databases to minimize search error. The database searched for the following words: “Technology Acceptance Model,” AND "e-learning" OR "distance education" OR "online learning" AND "undergraduate" OR "college" The database search retrieved a total of 38 articles. After I removed external duplicates, 34 articles remained for screening. Table 5 presents the total number of articles retrieved from each database.

Second, I found articles while hand-searching articles from Manuscript #1. The hand-search retrieved 4 articles. Lastly, I searched the reference section of the articles found in the database search and hand-search. The reference list search retrieved 39 articles.

Screening

The screening process occurred in two phases: a primary screening and secondary screening (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009), using an online reference management system, RefWorks. During the primary screening, I reviewed each article’s title and abstract to determine if the article was written in

Table 5

Articles Retrieved

Search	Database	Vendor	Number retrieved	External dups	New articles added
1	ERIC	EBSCO	16	0	16
	Education Full Text	EBSCO	19	4	15
	ProQuest Dissertations & Theses	ProQuest	3	0	3
2	Hand-searching	-	4	0	4
3	Reference lists	-	39	0	39
Total			81	4	77

Note. External dups = External duplicates between databases.

English, was quantitative in nature, and tested the TAM. The articles which met the first screening's criteria progressed to the second screening phase. During the secondary screening, I reviewed the entire article to determine if the article included samples which came from an undergraduate student population, used the technology in an online learning context, reported adequate statistics to calculate covariances, measured variables in the TAM, and measured variables at one time point. Interested readers may access the screened articles using the following permalink: http://goo.gl/5NYDKV.

Among the 77 studies in the primary screening, I removed one article, which was qualitative in nature, and two articles, which did not test the TAM. Among the remaining 74 articles, I excluded 61 articles during the secondary screening. Among the 61 excluded articles, six articles did not include an undergraduate student sample, two articles were not in an online learning context, 49 articles did not include statistics to

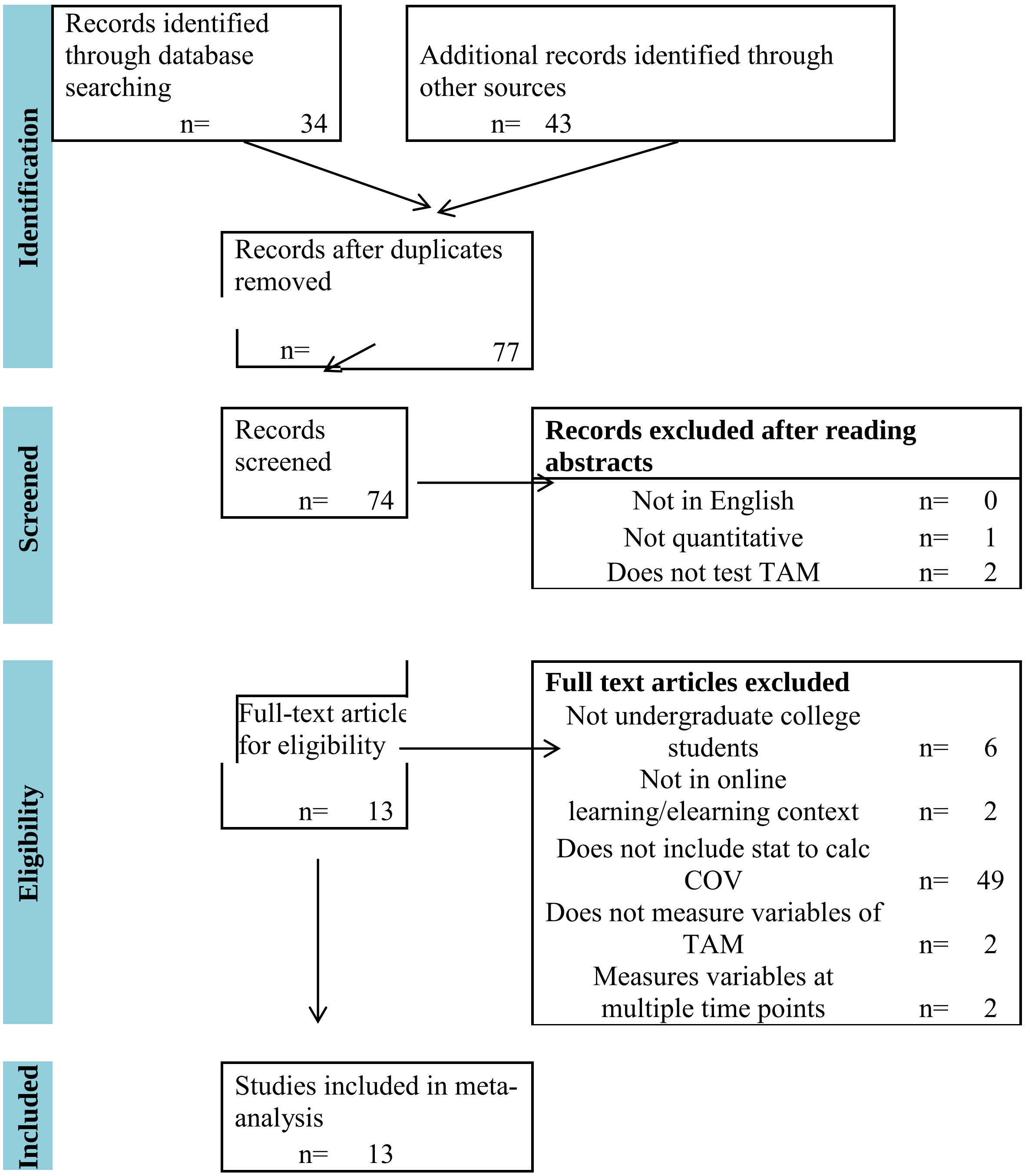


Figure 3 Article Screening Process.

calculate a covariance matrix, two articles did not measure variables in the TAM, and two articles measured variables at more than one time point. Figure 3 presents the screening process, which includes the number of articles excluded and the reason for exclusion.

Most of the articles did not report sufficient statistics to synthesize results. Given that 49 articles would be excluded due to lack of statistics reported, I sent two emails to the authors whose articles did not contain statistics to compute a covariance. The first email requested missing information (e.g., means and SDs and/or correlations), and a second email followed two weeks later with a reminder of the initial request. Appendix C presents a sample of the initial email and the second email.

Few authors responded and even fewer shared the requested missing information. Of the 49 articles with missing information, 13 authors responded, with one author providing me with the information requested. The most common response was to refer me to a co-author or suggest the data was lost. For example, one author noted both situations,

> "I do not have the data set or the cov[arinace] matrix any longer – I lost quite a number of files due to hard drive crash a couple of years ago. You might contact [co-author's name], my co-author. She ran the analyses so maybe she would still have those files."

Appendix D presents other responses from authors. I had little success in retaining articles with missing information, given authors' responses.

All 13 articles meeting inclusion criteria were published in seven peer-reviewed journals. Among the 13 articles, most came from a single journal, *C&E*. Table 6 presents the distribution of the articles meeting the inclusion criteria.

Table 6

Journals Represented in Meta-Analysis

Rank	Journal	Count	%
1	*Computers & Education*	6	46.15
2	*Educational Technology & Society*	2	15.38
3	*International Review of Research in Open and Distance Learning*	1	7.69
4	*Turkish Online Journal of Educational Technology - TOJET*	1	7.69
5	*British Journal of Educational Technology*	1	7.69
6	*Behaviour & Information Technology*	1	7.69
7	*Journal of Educational Computing Research*	1	7.69

Note. n = 13.

Coding Procedures

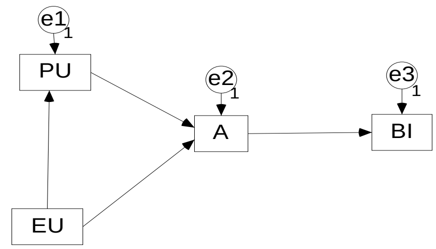
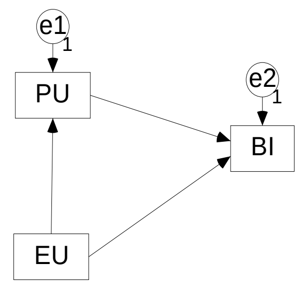
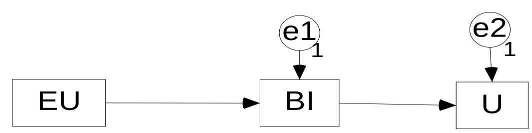
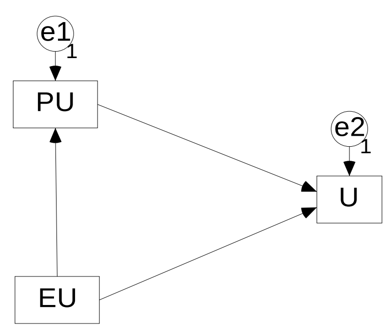
I created a coding scheme for the attributes of interest for the current study. Appendix E presents the different attributes coded in each category. I coded all 13 articles using an Excel spreadsheet.

Appendix G presents the 13 studies measured different combinations of the variables in the TAM. Thus, I grouped the studies based on the common variables measured. Table 7 presents the four groups tested. Each group included a different number of studies. Once the studies were grouped, I conducted a meta-analysis within each group and then compared the studies within each group using multiple-group analysis. Given the number of parameters estimated in Groups 2 and 4, the SEM could not be identified. Hence, only the multiple-group analysis is reported for Groups 2 and 4. Note, by grouping the studies based on the common variables measured, some of the 13 articles were not included in a given analysis (e.g., Davis & Wong, 2007; Lee & Lee, 2008; Saadé, 2007; and Pituch & Lee, 2006). Furthermore, some studies were analyzed in more than one group (e.g., Martins & Kellermanns, 2004; Ramayah, 2006; Saadé et al., 2007; and Yi & Hwang, 2003).

Meta-analysis. The current meta-analysis used MASEM. MASEM combines meta-analysis and structural equation modeling by pooling covariance matrices and testing structural equation models using the pooled covariance matrix. The current study used Cheung and Chan's (2005) proposed two-stage structural equation modeling (TSSEM) approach to fit MASEM using covariance matrices. In stage one, the homogeneity of the covariance matrices was tested and covariance matrices were pooled

Table 7

Groups of Studies Tested

Groups	Studies	Model
Group 1	Saadé & Galloway (2005) Saadé et al. (2007)	
Group 2	Almrashdah et al. (2010) Martins & Kellermanns (2004) Saadé & Bahli (2005) Saadé et al. (2007) Yi & Hwang (2003)	
Group 3	Liao & Lu (2008) Martins & Kellermanns (2004) Yi & Hwang (2003)	
Group 4	Brown (2002) Martins & Kellermanns (2004) Ramayah (2006) Yi & Hwang (2003)	

together. In stage two, the pooled covariance matrix was used to fit the structural equation model specified using the reticular action model (RAM) formulation (McArdle & McDonald 1984) and estimated using weighted least squares (WLS).

The current study utilized both fixed- and random-effects models. By definition, a fixed-effects MASEM was selected because the fixed-effects model assumes all population covariance matrices are the same. However, when researchers expect a large variance across the effect sizes of the studies due to differences in situational factors (e.g., setting), the studies are considered heterogeneous. When studies are heterogeneous, a random-effects MASEM is more appropriate. In a similar vein, a random-effect model assumes the population covariance matrices may vary across studies because the selected studies are random samples of the population. The random-effects model uses the weighted average of the effect sizes to reduce the possible bias introduced by a large variance of the effect size across studies. For a more exhaustive explanation of the TSSEM approach to a MASEM, see Appendix H.

The MASEM using the TSSEM approach was conducted using the metaSEM package version 0.8-4 (Cheung, 2013), the OpenMx package version 1.3.1-2301, and R version 2.15.3. The metaSEM package is a particularly useful package because metaSEM automatically takes the stage one model into account when estimating parameters, standard errors, and goodness-of-fit indices. Appendix I presents the syntax ran in R to conduct the meta-analysis for each group.

Multiple group analysis. By definition, multiple group analysis tests whether there are differences in the structural parameters across studies. For the present study, a path analysis using maximum likelihood estimation was used to estimate the structural parameters of the variables measured in each of the studies. To test the invariance across studies, I conducted a multi-group analysis of structural invariance for each group of studies. The first step established a baseline model, labeled as Model 1 in each group. Secondly, a constrained model was established and labeled as Model 2 in each group. In the constrained model, each parameter was forced to be equal across all studies in the group. Thirdly, a chi-square difference test between Model 2 and Model 1 was conducted. If the chi-square difference test resulted in a non-statistically significant difference across the studies, I concluded that the studies found similar results. If the chi-square difference test resulted in a statistically significant difference across the studies, I determined where the differences were by reviewing the critical ratios (e.g., z- statistics) of the parameter estimates in each study. The AMOS software was utilized to conduct the multiple-group analysis.

Multiple fit indices were reported and used to interpret model fit. While the chi-square test measures the model's ability to reproduce the sample covariance matrix; the chi-square test is sensitive to sample size and non-normality. Thus, several fit indices were considered to assess model fit, including root mean square error of approximation (RMSEA), root mean square residual (RMR), normed fit index, (NFI), goodness of fit index (GFI), and comparative fit index (CFI). RMSEA below .06 indicate a reasonable fit. An RMR of zero indicates a perfect fit; thus, the closer RMR is to zero, the better model fit. NFI, GFI and CFI values greater than 0.95 suggest reasonable model fit (Thompson, 2004).

Results

Meta-Analyses

Group 1. A fixed-effects MASEM combines two studies from Group 1. Figure 4 presents the model tested in the two studies.

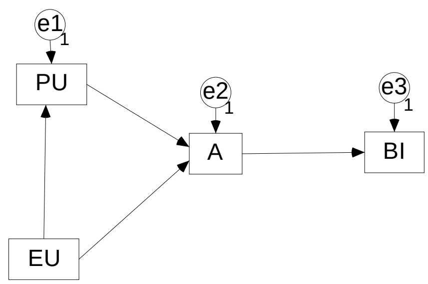


Figure 4 Group 1 Model

In Stage 1, homogeneity of the covariance matrices was met based on the goodness-of-fit indices: X^2 (df = 6, N = 490) = 12.78; p = .05, CFI = 0.99, TLI = 0.98, SRMR = 0.06, and RMSEA = 0.07. Given that the covariance matrices were homogeneous, the analysis continues to Stage 2 to fit structural model using RAM specification. In Stage 2, the fit indices of the structural model indicate good fit, X^2 (df = 2, N = 490) = 12.77; p = .0017, CFI = 0.99, TLI = 0.96, SRMR = 0.04, and RMSEA = 0.10. These indicators were consistent in indicating a generally acceptable fit of the hypothesized model to the data. Table 8 presents the standardized parameter estimates of the model.

Table 8
Group 1 Synthesis

		95% CI	
Parameter	Stand.	Lower	Upper
PU -> EU	0.51	0.44	0.57
PU -> A	0.52	0.45	0.60
EU -> A	0.16	0.08	0.24
A -> BI	0.61	0.55	0.67

Note. CI = confidence interval; Stand. = standardized estimate.

Group 3. A random-effects MASEM combines three studies from Group 3. Figure 5 presents the model tested in the three studies.

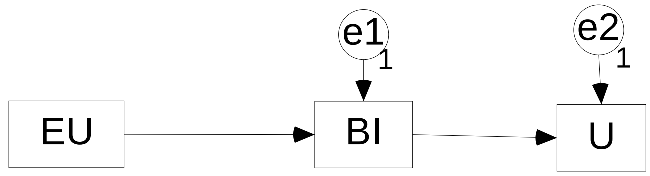


Figure 5 Group 3 Model

In Stage 1, homogeneity of the covariance matrices was not met based on the goodness-of-fit indices: X^2 (df = 6, N = 489) = 34.03; p = 0.98, CFI = 0.81, TLI = 0.71, SRMR = 0.16, and RMSEA = 0.17. Given that the covariance matrices were heterogeneous, a random-effects model is appropriate. In Stage 1, heterogeneity was confirmed Q (6) = 27.93, $p < .001$. The heterogeneity of EU, BI, and U were 97.30%, 96.64%, and 96.52%, respectively. In Stage 2, the fit indices on structural model indicates a perfect fit, X^2 (df = 1, N = 489) = 0.00, $p < .001$, CFI = 1.00, TLI = 1.00, SRMR = 0.00 and RMSEA = 0.00. Table 9 presents the standardized parameter estimates of the model.

Multi-group Analyses

Group 1. Group 1 compared two studies. Figure 4 presents the model tested in the two studies. Table 10 presents the model fit statistics and the invariance test between the constrained and unconstrained model. Recall, the constrained model assumes the parameters from each study are equal to each other. The chi-square difference test was not statistically significant; thus, the parameter estimates across the two studies were statistically the similar or invariant. Table 11 presents the standardized and

Table 9

Group 3 Synthesis

		95% CI	
Parameter	Stand.	Lower	Upper
EU -> BI	0.55	0.38	0.68
BI -> U	0.36	0.23	0.50

Note. CI = confidence interval; Stand. = standardized estimate.

unstandardized parameter estimates.

Group 2. Group 2 compared five studies. Figure 6 presents the model tested in the five studies. Table 12 presents the model fit statistics and the invariance test between the constrained and unconstrained model. The chi-square difference test was statistically significant; thus, there is a lack of model invariance across the five studies in this group. In other words, the parameter estimates across the five studies were statistically different. Table 13 presents the standardized and unstandardized parameter estimates.

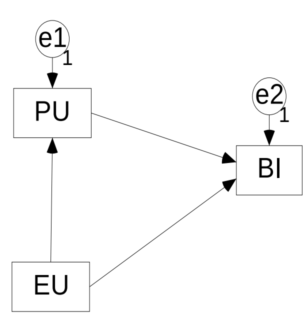


Figure 6 Group 2 Model

Table 10

Model Fit Statistics and Invariance Analysis of Group 1

No.	Model	X^2	df	p-value	RMSEA	RMR	NFI	GFI	CFI	ΔX^2	Δ df	p-value
1	unconstrained model	16.5	4	.002	.080	.060	.972	.984	.979			
2	constrained model	23.2	8	.003	.062	.078	.961	.976	.974	6.7	4	.153

Note. n = 490.

Table 11

Parameter Estimates of Group 1

Study	EU -> PU		PU -> A		EU -> A		A -> BI	
	Stan.	Unst.	Stan.	Unst.	Stan.	Unst.	Stan.	Unst.
Saadé & Galloway (2005)[a]	0.47	0.41	0.47	0.46	0.03	0.02	0.55	0.63
Saadé et al. (2007)[b]	0.51	0.47	0.51	0.51	0.21	0.19	0.60	0.61

Note. [a]n_1 = 128. [b]n_2 = 36. Stan = Standardized estimate, Unst. = Unstandardized estimate.

Table 12

Model Fit Statistics and Invariance Analysis of Group 2

No.	Model	X^2	df	p-value	RMSEA	RMR	NFI	GFI	CFI	ΔX^2	Δ df	p-value
1	unconstrained model	0.0	0	-	.256	0.000	1.000	1.000	1.000			
2	constrained model	143.4	12	$p < .001$	.094	.453	.883	.931	.892	143.4	12	$p < 0.0001$

Note. n = 1267.

Table 13

Parameter Estimates of Group 2

	EU -> PU		PU -> BI		EU -> BI	
Study	Stan.	Unst.	Stan.	Unst.	Stan.	Unst.
Saadé & Bahli (2005)[a]	0.26	0.23	0.36	0.47	0.06	0.07
Saadé et al. (2007)[b]	0.51	0.47	0.42	0.42	0.05	0.05
Almrashdah et al. (2010)[c]	0.79	0.83	0.62	0.69	0.19	0.23
Martins & Kellermanns (2004)[d]	0.49	0.72	0.37	0.45	0.25	0.44
Yi & Hwang (2003)[e]	0.29	0.29	0.46	0.50	0.22	0.24

Note. [a]n_1 = 128. [b]n_2 = 362. [c]n_3 = 425. [d]n_4 = 243. [e]n_5 = 109. Stan. = Standardized estimate, Unst. = Unstandardized estimate.

Given that the five studies had different parameter estimates, a post-hoc analysis was conducted to determine which studies had similar or different parameter estimates. To identify differences, I compared the critical ratios of the parameters estimates between each study. Table 14 presents the critical ratios. A statistically significant critical ratio suggests that the parameter estimate in one study is statistically different than the parameter estimate in another study. For example, the critical ratio of EU -> PU in Almrashdah et al. (2010) and Yi and Hwang (2003) is $z = -5.54$ and is statistically significant at $z = 1.96$ ($p = .05$). Thus, the parameter estimates of the EU -> PU in Almrashdah et al. (2010) and Yi and Hwang (2003) were statistically different from one another. Conversely, the critical ratio of EU -> PU in Almrashdah et al. (2010) and Martins and Kellermanns (2004) is $z = -1.328$ and is not statistically significant at $z = 1.96$ ($p = .05$). Thus, the parameter estimates of EU -> PU in Almrashdah et al. (2010) and Martins and Kellermanns (2004) were similar to each other. The results of the post hoc analysis across the five studies suggest that their parameter estimates may be different for EU -> PU and PU -> BI, but similar for EU -> BI. While the post hoc analysis provides potential insight to the nature of the differences in parameter estimates among the five studies, these results should be interpreted with caution. The post hoc analysis is exploratory in nature.

Group 3. Group 3 compared three studies. Figure 5 presents the model tested in the three studies. Table 15 presents the model fit statistics and the invariance test between the constrained and unconstrained model. The chi-square difference test was

Table 14

Critical Ratios of Parameter Estimates of Group 2

Study	Almrashdah et al. (2010)			Martins & Kellermann (2004)			Saadé & Bahli (2005)			Saadé et al. (2007)		
	EU - PU	PU - BI	EU - BI	EU - PU	PU - BI	EU - BI	EU - PU	PU - BI	EU - BI	EU - PU	PU - BI	EU - BI
Almrashdah et al. (2010)[a]												
Martins & Kellermann (2004)[b]	-1.328	-2.564*	1.748									
Saadé & Bahli (2005)[c]	-6.501*	-1.629	-1.222	-4.048*	0.153	-2.370*						
Saadé et al. (2007)[d]	-7.012*	-3.414*	-2.3*	-2.705*	-0.264	-3.244*	2.438*	-0.349	-0.197			
Yi & Hwang (2003)[e]	-5.54*	-1.761	0.123	-3.435*	.446	-1.427	0.461	0.203	1.154	-1.736	0.730	1.814

Note. [a]$n_1 = 425$. [b]$n_2 = 243$. [c]$n_3 = 102$. [d]$n_4 = 362$. [e]$n_5 = 109$. * |z-value| statistically significant at $z \geq 1.96$.

Table 15

Model Fit Statistics and Invariance Analysis of Group 3

No.	Model	X²	df	p-value	RMSEA	RMR	NFI	GFI	CFI	ΔX²	Δ df	p-value
1	unconstrained model	23.5	3	p < .001	.118	7.526	.848	.971	.859			
2	constrained model	38.7	7	p < .001	.097	21.762	.750	.952	.782	15.2	12	p = 0.0043

Note. n = 489.

Table 16

Parameter Estimates of Group 3

	EU -> BI		BI -> U	
Study	Stand.	Unstand.	Stand.	Unstand.
Martins & Kellermanns (2004)[a]	0.43	0.76	0.30	0.31
Yi & Hwang (2003)[b]	0.35	0.38	0.26	18.74
Liao & Lu (2008)[c]	0.47	0.47	0.17	0.33

Note. [a]n_1 = 243. [b]n_2 = 109. [c]n_3 = 137.

statistically significant; thus, there was a lack of model invariance across the three studies in this group. Table 16 presents the standardized and unstandardized parameter estimates.

Given that the three studies had different parameter estimates, a post-hoc analysis was conducted to determine which studies had similar or different parameter estimates. To identify differences, the critical ratios of the parameters estimates between each study were compared. Table 17 presents the critical ratios between the three studies. The results of the post hoc analysis across the three studies suggest that the parameter estimates were statistically different for both EU -> BI and BI -> U.

Group 4. Group 4 compared four studies. Figure 7 presents the model tested in the four studies. Table 18 presents the model fit statistics and the invariance test between the constrained and unconstrained model. The chi-square difference test was statistically significant; thus, there was a lack of model invariance across the four studies in this group. Table 19 presents the standardized and unstandardized parameter estimates.

Given that the four studies had different parameter estimates, a post-hoc analysis was conducted to determine which studies had different parameter estimates. To identify differences, the critical ratios of the parameters estimates between each study were compared. Table 20 presents the critical ratios. The results of the post hoc analysis across the four studies suggested that the parameter estimates were invariant for both EU -> PU, PU -> U, but non-invariant for the path, EU -> U.

Table 17

Critical Ratios of Parameter Estimates of Group 3

Study	Liao & Lu (2008)		Martins & Kellermanns (2004)	
	EU -> BI	BI -> U	EU -> BI	BI -> U
Liao & Lu (2008)[a]				
Martins & Kellermanns (2004)[b]	2.330*	-0.114		
Yi & Hwang (2003)[c]	-0.643	2.752*	-2.649*	2.756*

Note. [a]$n_1 = 137$. [b]$n_2 = 243$. [c]$n_3 = 109$. * |z-value| statistically significant at $z \geq 1.96$.

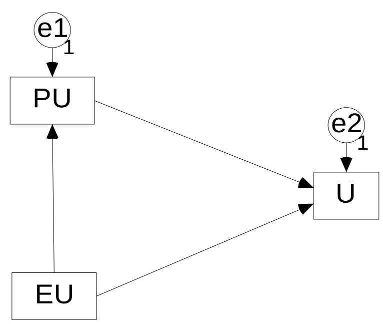


Figure 7 Group 4 Model

Table 18

Model Fit Statistics and Invariance Analysis of Group 4

No.	Model	X^2	df	p-value	RMSEA	RMR	NFI	GFI	CFI	ΔX^2	Δ df	p-value
1	unconstrained model	0	0	-	.213	0.000	1.000	1.000	1.000			
2	constrained model	37.9	9	$p < .001$	.068	14.509	.903	.965	.924	37.9	9	$p < 0.0001$

Note. n = 700.

Table 19

Parameter Estimates of Group 4

	EU -> PU		PU -> U		EU -> U	
Study	Stand.	Unstand.	Stand.	Unstand.	Stand.	Unstand.
Brown (2002)[a]	0.39	0.40	0.04	0.05	0.32	0.37
Martins & Kellermanns (2004)[b]	0.49	0.72	0.23	0.28	0.07	0.13
Ramayah (2006)[c]	0.55	0.46	0.32	0.41	0.45	0.48
Yi & Hwang (2003)[d]	0.29	0.29	-0.04	-3.16	0.24	19.15

Note. [a]$n_1 = 73$. [b]$n_2 = 243$. [c]$n_3 = 275$. [d]$n_4 = 109$.

Discussion

The purpose of the current study was to determine whether the TAM explains undergraduates' acceptance of online learning. In contrast to previous meta-analyses, which focused on a variety of populations and an array of technologies, the current study isolated both only one population and only one technology. Specifically, the present study investigated undergraduate students and online learning management systems. Furthermore, the present study utilized multiple group analysis to identify similarities and differences between studies. First, I tested the relative fit of four groups of studies using multiple group analysis. In Group 1, the studies were replicable, as assessed by the ΔX^2 test and fit indices seen in Table 10. In the remaining three groups, the studies had statistically different results (e.g., Tables 12, 15 and 18). Second, I examined the critical ratios for each path in the proposed model. The results suggest that the parameter estimates were different across certain paths and similar across other paths. These differences may be due to cultural differences (Schepers & Wetzels, 2007; Straub et al., 1997) or gender differences (Gefen & Straub, 1997) across studies. The following section expounds on the results from each group of studies.

Group 1 Studies

The results from the meta-analysis of Group 1 suggested that the fixed-effects model was an acceptable fit. The present study confirmed that perceived ease of use has

Table 20

Critical Ratios of Parameter Estimates of Group 4

Study	Brown (2002)			Martins & Kellermann (2004)			Ramayah (2006)		
	EU -> PU	PU -> U	EU -> U	EU -> PU	PU -> U	EU -> U	EU -> PU	PU -> U	EU -> U
Brown (2002)[a]									
Martins & Kellermanns (2004)[b]	-0.196	1.472	-1.283						
Ramayah (2006)[c]	0.454	2.410*	0.741	-2.795*	1.098	2.486*			
Yi & Hwang (2003)[d]	-0.774	-0.417	2.427*	-3.433*	-0.447	2.459*	-1.640	-0.463	2.414*

Note. [a]$n_1 = 73$. [b]$n_2 = 243$. [c]$n_3 = 275$. [d]$n_4 = 109$. * |z-value| statistically significant at $z \geq 1.96$.

a strong effect on perceived usefulness as demonstrated in previous studies. Both perceived usefulness and perceived ease of use influence individual attitudes. However, the relationship between perceived usefulness and attitude is stronger, $r = .52$, 95% CI [0.45, 0.60], than the relationship between perceived ease of use and attitude, $r = .16$, 95% CI [0.08, 0.24]. The influence of attitude to behavioral intention is also profound. Likewise, the multiple group analysis echoed the findings of Saadé et al. (2007), suggesting that the parameter estimates were the similar across studies. In fact, Group 1 was the only group of studies that were statistically similar. The results were reasonable given the studies' sample, learning management system, and instrumentation. The two studies were similar in that the two studies both draw on a sample from the same population. For example, Saadé and Galloway (2005) described their sample as students taking a "core management information systems course at Concordia University in Montreal, Canada" (p. 291). Moreover, both studies used the same in-house-developed, learning management system, in which Saadé et al. (2007) referred to as a "multimedia learning system (MMLS)" (p. 175). Additionally, Saadé et al. (2007) noted, that both studies used the same "methodology" and instruments (p. 178). Given the similarities between the two studies, one would expect the results to be replicable, which the studies were.

The current study offers further insight into the primary findings of Saadé et al. (2007). While Saadé et al. (2007) used visual inspection of the parameter estimates across both studies, the present multiple group analysis utilized statistical-based invariance testing. The multiple group analysis strengthens the previous findings of

Saadé et al. (2007) and offers a clearer conclusion regarding the equality between each parameter estimate in the path model. Moreover, the relationship between attitude and behavioral intention echoes the findings of Ursavaş (2013).

Group 2 Studies

The multiple group analysis results suggested that the parameter estimates were different across the five studies. First, the current study found the EU -> PU path was statistically different between 7 of the 10 pairs of studies. For example, Almrashdah et al. (2010) and Saadé et al. (2007) were statistically different from each other. Unlike Tai, Zhang, Chang, Chen, and Chen (2012), the present results suggest mixed findings across the five studies regarding the relationship between ease of use and perceived ease of use. Second, the current study found the relationship between perceived usefulness and behavioral intention was relatively consistent across studies, with only 2 of the 10 pairs of studies diverging from each other. The results mirror Saadé et al.'s (2007) meta-analysis, which suggested a consistent and slight relationship between perceived usefulness and behavioral intention. Lastly, the relationship between ease of use and behavioral intention was statistically different between only 3 of the 10 pairs of studies, a result emulating King and He's (2006) findings. Given the differences across studies, researchers should be cautious when forming conclusions regarding the relationships between the three variables: perceived usefulness, ease of use, and behavioral intention.

Group 3 Studies

The results from the meta-analysis suggest that the random-effects model represented in Group 3 was an acceptable fit. Despite the adequate model fit, researchers should be cautious when forming conclusions regarding the relationships between the three variables: ease of use, behavioral intention, and actual use, because the model tested only two relationships within the TAM. For example, the relationship between ease of use and behavioral intention was relatively strong, while the relationship between behavioral intention and use was moderate.

Furthermore, the multiple group analysis suggests that the parameter estimates were different across the four studies. First, the current study found the EU -> BI path was statistically different between 2 of the 3 pairs of studies. Second, the relationship between behavioral intention and actual use was statistically different across 2 of the 3 pairs of studies. The results suggest that the parameters estimates were different across studies, which also maintains the idea that findings were not replicable across studies.

Group 4 Studies

As seen in the multiple group analysis, results suggest that the parameter estimates were different across the four studies. First, the current study found the EU -> PU path was statistically different between only 2 of the 6 pairs of studies, suggesting Group 4 studies relatively reproduce a similar relationship between ease of use and perceived ease of use. Second, the relationship between perceived usefulness and actual use was relatively consistent across studies, because only 1 of the 6 pairs of studies were different from each other. Lastly, the relationship between ease of use and actual use was statistically different between 4 of the 6 pairs of studies. The results suggest mixed findings across the four studies regarding the relationship between ease of use and actual use, a finding which resonates with Ma and Liu (2004).

Although primary studies have validated the TAM with undergraduate students in an online learning context, I caution practitioners in the field when making decisions about undergraduate online learning based on the TAM. Moreover, most of the prior meta-analyses have only looked at the bivariate relationships represented in the TAM, instead of the model as a whole, with one exception: Tai et al. (2012) meta-analytically tested the model as a whole using correlation matrices. Tai et al.'s (2012) attempt was a progressive step and should be commended. However, the study was limited by only using a pooled correlation matrix. The current study attempted to use a pooled covariance matrix, which provides more information for the path analysis. However, I faced many challenges in attempting to meta-analyze studies. Perhaps, Tai et al. (2012) encountered similar challenges, and therefore chose to utilize a more accessible correlation matrix to synthesize findings.

Limitations

The present study faced many challenges in attempt to meta-analyze studies. First, the current study was limited by the range of variables included in past research, a limitation that Fried, Shirom, Gilboa, and Cooper (2008) also found in their meta-analysis using structural equation modeling. Second, the present meta-analysis was limited to the statistics provided by the authors. The following two paragraphs expound on these limitations.

More specifically, the current study was limited by the range of variables included in past research. For example, studies that used the TAM tested different combinations of the variables within the multiple iterations of the TAM. By testing different combinations of variables, all 13 studies could not be synthesized together due to missing variables. Although the TSSM approach handles missing covariances, there is currently no MASEM approach to handle missing variables. Hence, in the current MASEM, the studies were grouped based on the common variables measured. Given this restriction, some of the 13 articles were not included in the analysis, (e.g., Davis & Wong, 2007; Lee & Lee, 2008; Saadé, 2007; and Pituch & Lee, 2006), and some studies were analyzed in more than one group (e.g., Martins & Kellermanns, 2004; Ramayah, 2006; Saadé, Nebebe, & Tan, 2007; and Yi & Hwang, 2003). Meta-analysts are limited by the extent to which articles can be meta-analyzed, when authors report modified versions of a theoretical model.

Another glaring challenge was the inadequate reporting of statistics to conduct the meta-analysis. Among the 77 articles identified, authors of 49 articles did not report the appropriate statistics to compute a covariance matrix. This denotes authors did not report either the means and/or standard deviations and/or correlations of the variables. Despite multiple attempts to request missing statistics from the authors and co-authors, only one author responded with the missing information. When authors do not report adequate statistics in primary studies, meta-analysts cannot include these studies in a meta-analysis; thus, the information from those primary studies was essentially lost.

Future Meta-Analysis Research

For future research, researchers may test the aforementioned models without using some of the studies in the current study. For example, results from Yi and Hwang (2003) study appeared dissimilar to the results from the other studies in the present study. The dissimilarity noted may be the result of different measurement scales for the variable, actual use. Consequently, without the results from Yi and Hwang (2003) study, researchers may reveal a stronger parameter estimate for the model described in the current study.

Additionally, future studies could compare meta-analysis results to longitudinal results. The current study found three articles that measured variables within the TAM at multiple time points. These studies measured students' perspectives before taking an online course and after taking the online course. Future meta-analyses could investigate variables based on novice and proficient users and then compare whether the meta-analysis results correspond to the longitudinal results.

Summary

The advancement of online learning technologies has provided unmatched accessibility for colleges to meet the educational needs of students than ever before. As Bennett and Green (2001) noted, "There is little doubt that more and more college classes will be placed online in the future, and we are fast approaching the point when it will be the norm to have several courses online at the universities throughout the nation" (p. 495). Although prophetic in its time, today this statement seems commonplace. While college administrators advocate for online courses, the current study suggests practitioners are making decisions based on non-replicable results.

The TAM is a popular model for explaining and predicting undergraduates' learning management system use. To date, researchers have conducted numerous studies on the TAM and obtained numerous confirmatory results through primary studies.

Researchers have selected a variety of ways to validate or extend the TAM. For example, some researchers conducted replication studies, such as Adams, Nelson, and Todd (1992), while other researchers rely on meta-analyses (e.g., Šumak, Heričko, & Pušnik, 2011). Moreover, some researchers look to longitudinal studies (e.g., Venkatesh & Davis, 2000), while other researchers relied on a series of single primary studies to validate or extend the TAM. The current study attempted to use meta-analytic structural equation modeling to validate or extend the TAM. Unfortunately, there were too many obstacles to definitively confirm any version of the TAM meta-analytically.

However, researchers should heed the concerns expressed here regarding the application and accuracy of the model in an undergraduate online learning context. As demonstrated in the current study, some researchers may have formed erroneous conclusions regarding the relationships between the variables in the TAM. Moreover, the multiple group analysis suggests that the studies included here resulted in statistically different findings. Hence, the findings across studies were not replicable.

Consequently, researchers have spent over a decade modifying a theoretical model based on primary studies that has demonstrated little explanatory or predictive power. Hence, future research should be careful not to develop new models which would exploit the strengths of the TAM while ignoring the model's weaknesses. In sum, decision-makers should carefully consider students' preferences before investing in online learning technologies; however, decision-makers should base their decisions on the findings from theoretical models validated in an online learning context.

4. SUMMARY AND CONCLUSIONS

Today's educational technology researchers tend to conduct research in isolation or with a narrow view of the field. Further, educational technology research is limited by inadequate reporting practices. For example, Study 1 found researchers rarely report the type of sample, score reliability, or informationally-adequate statistics. Likewise, as observed in the screening phase of Study 2, educational technology researchers' lack of response to requests for missing information suggests that researchers lack of awareness, or regard for, meta-analytic research practices. With over 400 educational technology journals currently published (cf. *2012 Cabell's Directories of Publishing Opportunities*), researchers are encouraged to stop conducting research in isolation.

Furthermore, meta-analysts are challenged to synthesize and interpret the field's findings given researchers' current reporting practices. Challenges in synthesizing and interpreting findings arise when researchers do not think meta-analytically. Meta-analytic techniques, such as meta-analyses, offer stronger evidence for researchers to form conclusions and make critical decisions in both current practice and future research efforts in educational technology.

Study One

The first study offered empirical evidence of the field's current status with regard to reporting results using meta-analytic thinking. The current study found that quantitative methods continue to dominate the field as a whole, yet journals appear to favor certain research methods over others. Additionally, the present study found overall poor reporting practices - approximately only half of the authors reported reliability of the scores produced by their instrument. Moreover, findings suggest few authors report informationally-adequate statistics. For example, sometimes means were reported without the accompanying standard deviations. The lack of reporting accompanying means and standard deviations is largely recognized as poor practice (Thompson, 2006), while this type of practice limits the number of articles included in meta-analyses (Cooper, 2010).

Although the nature of the current study was descriptive, the results identified some of the strengths and weaknesses of the current research methods used in educational technology research. Moreover, meta-analysts rely on the quality of primary studies to conduct secondary research studies. Accordingly, educational technology researchers should report evidence that the *scores* they are analyzing are reliable because of the possible impact reliability has on the interpretation of research results. Moreover, researchers should report informationally-adequate statistics so that readers can evaluate findings appropriately.

Study Two

The preliminary aim of the second study was to offer a glimpse of where the field could go once researchers begin to think meta-analytically. However, the current meta-analysis revealed numerous challenges that impeded the synthesis of primary studies. Authors rarely reported adequate statistics to synthesize studies. In fact, among the possible 77 articles which met inclusion criteria, 49 studies were removed from the synthesis due to insufficient statistics. Additionally, despite multiple attempts to contact authors for missing information, few responded with the requested information and were unconcerned with having their article excluded from the meta-analysis. Once the number of articles was finalized, I found that researchers reported different combinations of variables within the TAM. As a result, the 13 studies included were forced into smaller groups to analyze the data without missing variables.

The second study attempted to combine meta-analysis and structural equation modeling to extend and refine theoretical models. Unfortunately, there were too many obstacles to confirm any version of the TAM meta-analytically. Instead, the second study investigated different combinations of variables and formed conclusions about the relationships between the variables that were available.

Although the TAM is a popular model for explaining and predicting undergraduates' learning management system use, the results of the current meta-analysis suggested that educational technology researchers should be cautious when forming conclusions about undergraduate online learning based on the TAM model. As demonstrated through meta-analysis, researchers may have formed erroneous conclusions regarding the relationships between the variables in the TAM. More specifically, the studies included in the meta-analysis resulted in statistically different findings across multiple parameter estimates. Therefore, the findings across studies are not replicable. Perhaps researchers have spent over a decade modifying a theoretical model based on primary studies, which has demonstrated little explanation or predictability.

In closing, the current thesis does not intend to vilify educational technology research. Instead, the current work attempts to reflect on the field of educational technology research by offering empirical evidence of issues expressed over a decade ago and pleading for reform in the field. The current thesis encourages authors to think meta-analytically when conducting primary studies. Instead, our research must be thoughtful and extend the knowledge of the field systematically.

We must acknowledge and connect our own research to the greater body of literature. We should no longer only form conclusions from a single study. Instead, we should view the findings as one more piece of the greater picture. But behind the winsome primary studies lies an uncompromising conviction: secondary studies propel the field forward in a unified direction. Therefore, as researchers investigate topics further, someone must come through to synthesize findings across studies. Educational technology researchers should engage in more secondary studies so that the field can move at more reasonable pace in relation to the quickly-changing pace of technology in education. Once meta-analyses are used for the purpose they serve in the social sciences, the educational technology field can move forward empirically, rather than jumping to the next novel technology.

REFERENCES

Adams, D. A., Nelson, R. R., & Todd, P. A. (1992). Perceived usefulness, ease of use, and usage of information technology: A replication, *MIS Quarterly, 16,* 227-247.

Alper, A. & Gülbahar, Y. (2009). Trends and issues in educational technologies: A review of recent research in TOJET. *The Turkish Online Journal of Educational Technology, 8*(2), 124-135.

*Almrashdah, I. A., Sahari, N., Mat Zin, N. A. H., Alsmadi, M. (2010, November). Distance learners acceptance of learning management system. Paper presented at the meeting of Advanced Information Management and Service (IMS), International Conference Seoul, 304 – 309.

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher, 35*, 33-40. doi:10.3102/0013189x035006033

American Psychological Association (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

APA Publications & Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology Why do we need them? what might they be? *American Psychologist, 63*, 839-851. doi: http://dx.doi.org.lib-ezproxy.tamu.edu:2048/10.1037/0003-066X.63.9.839

Australasian Society for Computers in Learning in Tertiary Education (ASCILITE). (2012). *Australasian Journal of Educational Technology*. Retrieved from http://www.ascilite.org.au/ajet/about/about.html

Bell, L., Schrum, L., & Thompson, A. D. (2008). *Framing research on technology and student learning in the content areas: Implications for educators*. Charlotte, NC: Information Age Publishing.

Bennett, G. & Green, F. P. (2001). Promoting service learning via online instruction. *College Student Journal, 35,* 491-497.

*Brown, I. T. J. (2002). Individual and technological factors affecting perceived ease of use of web-based learning technologies in a developing country. *The Electronic Journal on Information Systems in Developing Countries, 9*(5), 1-15.

Chen, W. & Hirschheim, R. (2004). A paradigmatic and methodological examination of information systems research from 1991 to 2001. *Information Systems Journal, 14*, 197-235. doi: 10.1111/j.1365-2575.2004.00173.

Cheung, M. W.-L. (2013). Meta-analysis using structural equation modeling (Version 0.8-4) [Computer Software]. Retrieved from http://courses.nus.edu.sg/course/psycwlm/Internet/metaSEM/

Cheung, M.W.-L. & Chan, W. (2005). Meta-analytic Structural Equation Modeling: A two-stage approach. *Psychological Methods, 10*, 40-64. doi: 10.1037/1082-989X.10.1.40

Cheung, W. S. & Hew, K. F. (2009). A review of research methodologies used in studies on mobile handheld devices in K-12 and higher education settings. *Australasian Journal of Educational Technology, 25*, 153-183.

Clark, R.E. (1983). Reconsidering research on learning with media. *Review of Educational Research, 53*(4), 445-459.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Thousand Oaks, CA: Sage.

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Davis, F. D. (1986). *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. (Doctoral dissertation). Retrieved from DSpace@MIT. (http://hdl.handle.net/1721.1/15192)

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*, 982–1003.

*Davis, R. & Wong, D. (2007). Conceptualizing and measuring the optimal experience of the elearning environment. *Decision Sciences Journal of Innovative Education, 5,* 97-126. doi: 10.1111/j.1540-4609.2007.00129.x

Edyburn, D. L. (2000). 1999 in review: A synthesis of the special education technology literature. *Journal of Special Education Technology, 15*, 7-30.

Elsevier B. V. (2012). *Computers & Education*. Retrieved from http://www.journals.elsevier.com/computers-and-education/

Farhoomand, A. & Drury, D.H. (1999). A historiographical examination of information systems. *Communications of the Association for Information Systems*, *1*(5es), 1-27.

Fried, Y., Shirom, A., Gilboa, S., & Cooper, C. L. (2008). The mediating effects of job satisfaction and propensity to leave on role stress-job performance relationships: Combining meta-analysis and structural equation modeling. *International Journal of Stress Management, 15,* 305-328. doi: 10.1037/a0013932

Gall, M. D., Gall, J. P., & Brog, W. R. (2007). *Educational research an introduction* (8th ed.). Boston, MA: Pearson Education.

Gefen, D. & Straub, D. W. (1997). Gender Differences in the Perception and Use of E-Mail: An Extension to the Technology Acceptance Model. *MIS Quarterly*, 21, 389-400.

Goode, J. (2010). Mind the gap: The digital dimension of college access. *Journal of Higher Education, 81*, 583-618.

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan.

Haertel, G. D. & Means, B. (2003). *Evaluating educational technology: Effective research designs for improving learning*. New York, NY: Teachers College Press.

Higgins, J. P. & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539–1558. doi: 10.1002/sim.1186

Higgins, J. P. , Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557-560. doi: 10.1136/bmj.327.7414.557

Hoban, C. F. (1958). Research on media. *AV Communication Review, 6*, 169-178.

Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement, 60*, 523-531. doi:10.1177/00131640021970691

Hrastinski, S. & Keller, C. (2007). An examination of research approaches that underlie research on educational technology: A review from 2000 to 2004. *Journal of Educational Computing Research, 36*, 175-190.

James, L. R., Demaree, R. G. & Wolf, G. (1993). r_{wg}: An assessment of within group interrater agreement. *Journal of Applied Psychology*, 78, 306-309.

Johnson, R. B., & Onwuegbuzie. A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*, 14-26.

John Wiley & Sons. (2012). *British Journal of Educational Technology*. Retrieved from http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1467-8535/homepage/ProductInformation.html

Karataş, S. (2008). Interaction in the internet-based distance learning researches: Results of a trend analysis. *The Turkish Online Journal of Educational Technology*, 7(2), 11-19.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

Koble, M. A & Bunker, E. L. (1997). Trends in research and practice: An examination of The American Journal of Distance Education 1987 to 1995. *American Journal of Distance Education, 11*, 19-38. doi: http://dx.doi.org/10.1080/08923649709526959

King, W. R. & He, J. (2006). A meta-analysis of the TAM. *Information & Management 43,* 740-755.

Lee, Y., Driscoll, M. P., & Nelson, D. W. (2004). The past, present, and future of research in distance education: Results of a content analysis. *American Journal of Distance Education*, 18, 225-241. doi: http://dx.doi.org/10.1207/s15389286ajde1804_4

Lee, Y., Driscoll, M. P., & Nelson, D. W. (2007). Trends in research: A content analysis of major journals. In M. G. Moore (Ed.), *Handbook of Distance Education* (2nd ed., pp. 31-41). Mahwah, NJ: Lawrence Erlbaum Associates.

*Lee, J. K. & Lee, W. K. (2008). The relationship of e-Learner's self-regulatory efficacy and perception of e-Learning environmental quality. *Computers in Human Behavior, 24,* 32-47 . doi: 10.1016/j.chb.2006.12.001

*Liao, H. & Lu, H. (2008). The role of experience and innovation characteristics in the adoption and continued use of e-learning websites. *Computers & Education, 51,* 1405-1416. doi:10.1016/j.compedu.2007.11.006

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Lloyd, J. M., Dean, L. A., & Cooper, D. L. (2007). Students' technology use and its effects on peer relationships, academic involvement, and healthy lifestyles. *NASPA Journal, 44*, 481-495.

Ma, Q., & Liu, L. (2004). The TAM: A meta-analysis of empirical findings. *Journal of Organization and End User Computing, 16*(1), 59-72.

*Martins, L. L. & Kellermanns, F. W. (2004). A model of business school students' acceptance of a web-based course management system. *Academy of Management Learning and Education, 3,* 7-26.

McArdle J.J. & McDonald R.P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology*, 37, 234-251. doi:10.1111/j.2044-8317.1984.tb00802.x.

Means, B. & Haertel, G. D. (2004). *Using technology evaluation to enhance student learning.* New York, NY: Teachers College Press.

Merchant, Z., Goetz, E. T., Keeney-Kennicutt, W., Kwok, O., Cifuentes, L., and Davis, T. J. (2012). The learning characteristics, features of desktop 3D virtual reality environments, and college chemistry instruction: A structural equation modeling analysis. *Computers & Education*, *59,* 551-568. doi: http://dx.doi.org/10.1016/j.compedu.2012.02.004

Mielke, K.W. (1968). Questioning the question of ETV research. *Educational Broadcasting, 2*, 6-15.

Mitchell, P. D. (1997). The impact of educational technology: A radical reappraisal of research methods. *Association for Learning Technology Journal, 5*(1), 48-54. doi:10.1080/0968776970050108

Moher, D. Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement PLoS Med 6(6): e1000097. doi: 10.1371/journal.prned1000097

Peterson-Karlan, G. P. (2011). Technology to support writing by students with learning and academic disabilities: Recent research trends and findings. *Assistive Technology Outcomes and Benefits, 7*, 39-62.

*Pituch, K. A., & Lee,Y. (2006). The influence of system characteristics on e-learning use. *Computers & Education, 47*, 222-244. doi:10.1016/j.compedu.2004.10.007

Presley, A., & Presley, T. (2009). Factors influencing student acceptance and use of academic portals. *Journal of Computing in Higher Education, 21*, 167-182. doi: 10.1007/s12528-009-9022-7

*Ramayah, T. (2006). Course website usage: Does prior experience matter?. *WSEAS TRANS. on Information Science & Applications, 2,* 299-306.

Randolph, J. J. (2008). A methodological review of the program evaluations in K-12 computer science education. *Informatics in Education, 7*, 237-258.

Randolph, J., Julnes, G., Sutinen, E., & Lehman, S. (2008). A methodological review of computer science education research. *Journal of Information Technology Education, 7*, 135-162.

Raykov, T. (2001). Estimation of congeneric scale reliability via covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology*, 54, 315–323.

Reeves, T.C. (2000, April). Enhancing the worth of instructional technology research through "design experiments" and other development research strategies. In Session 41.29, *International perspectives on instructional technology research for the 21st century*. Symposium conducted at the annual meeting of the American Educational Research Association, New Orleans, LA.

Ritter, N. L. (2010, February). *Understanding a widely misunderstood statistic: Cronbach's alpha*. Paper presented at the meeting of Southwestern Educational Research Association Conference 2010, New Orleans, LA (ED526237).

Rourke, L., & Szabo, M. A. (2002). A content analysis of the Journal of Distance Education 1986–2001. *Journal of Distance Education, 17*(1), 63-74.

*Saadé, R. G. (2007). Exploring dimensions to perceived usefulness: Towards an enhanced assessment. *Decision Sciences Journal of Innovative Education, 5,* 289-310.

*Saadé, R., & Bahli, B. (2005). The impact of cognitive absorption on perceived usefulness and perceived ease of use in on-line learning: an extension of the technology acceptance model. *WSEAS TRANS. on Information Science & Applications, 42,* 317-327. doi:10.1016/j.im.2003.12.013

*Saadé, R. G. & Galloway, I. (2005). Understanding intention to use multimedia information systems for learning. *Issues in Informing Science and Information Technology, 2,* 287-296.

*Saadé, R., Nebebe, F., & Tan, W. (2007). Viability of the "Technology Acceptance Model" in multimedia learning environments: A comparative study. *Interdisciplinary Journal of Knowledge and Learning Objects, 3,* 175-184.

Schepers, J. & Wetzels, M. (2007). Meta-analysis of the technology acceptance model: Investigating subjective norm and moderation effects. *Information & Management*, 44, 90-103.

Shih, M., Feng, J. & Tsai, C. C. (2008). Research and trends in the field of e-learning from 2001 to 2005: A content analysis of cognitive studies in selected journals. *Computers & Education, 51*, 955-967.

Şİmşek, A., Özdamar, N., Uysal, O., Kobak, K. Berk, C., Kiliçer, T., & Çİğdem, H. (2009). Current trends in educational technology research in Turkey in the new millennium. *Educational Sciences: Theory and Practice, 9*, 961-966.

Skidmore, S. T. (2009). *Effect size matters: Empirical investigations to help researchers make informed decisions on commonly used statistical techniques*. Retrieved from ProQuest Digital Dissertations. (UMI 3399894)

Skidmore, S. T. & Thompson, B. (2010). Statistical techniques used in published articles: A historical review of reviews. *Educational and Psychological Measurement, 70*, 777-795. doi: 10.1177/0013164410379320

Springer. (2012). *Educational Technology Research and Development*. Retrieved from http://www.springer.com/education+%26+language/learning+%26+instruction/journal/11423

Straub, D., Keil, M., & Brenner, W. (1997). Testing the technology acceptance model across cultures: A three country study. *Information & Management*, *33*, 1–11.

Šumak, B., Heričko, M., & Pušnik, M. (2011). A meta-analysis of e-learning technology acceptance: The role of user types and e-learning technology types. *Computers in Human Behavior, 27,* 2067, 2077. doi:10.1016/j.chb.2011.08.005

Tai, D. W. S., Zhang, R., Chang, S., Chen, C., & Chen, J. (2012). A Meta-Analytic Path Analysis of e-Learning Acceptance Model. *World Academy of Science, Engineering and Technology*, *65*, 104-107.

Tatsuoka, M. M. (1971). *Significance tests: Univariate and multivariate*. Champaign, IL: Institute for Personality and Ability Testing.

Teclehaimanot, B., & Hickman, T. (2011). Student-teacher interaction on facebook: What students find appropriate. *TechTrends: Linking Research and Practice to Improve Learning, 55*(3), 19-30.

Thompson, B. (1992). Two and one-half decades of leadership in measurement and evaluation. *Journal of Counseling and Development*, *70*, 434-438.

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement, 54*, 837-847.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, *31*(3), 25-32. doi: 10.3102/0013189X031003025

Thompson, B. (Ed.) (2003). *Score reliability: Contemporary thinking on reliability issues.* Newbury Park, CA: Sage.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington, DC: American Psychological Association.

Thompson, B. (2006). *Foundations of behavioral statistics*. New York, NY: Guilford Press.

Ursavaş, Ö. F. (2013). Reconsidering the role of attitude in the TAM: An answer to Teo (2009) and Nistor and Heymann (2010), and Lopez-Bonilla and Lopez-Bonilla (2011). *British Journal of Educational Technology, 44,* E22–E25. doi:10.1111/j.1467-8535.2012.01327.x

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58*, 6-20.

Venkatesh, V. & Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision Sciences, 27,* 451-481. doi: 10.1111/j.1540-5915.1996.tb00860.x

Venkatesh, V. & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46*, 186-204. doi: http://www.jstor.org/stable/2634758

Wang, F. & Lockee, B. B. (2010). Virtual worlds in distance education a content analysis study. *Quarterly Review of Distance Education, 11*, 183-186.

Warne, R.T. Lazo, M., Ramos, T., & Ritter, N. (2012). Statistical methods used in gifted education journals, 2006-2010. *Gifted Child Quarterly, 56*, 134-149. doi: 10.1177/0016986212444122

Willson, V. L. (1980). Research techniques in *AERJ* articles: 1969 to 1978. *Educational Researcher, 9*(5), 5-10. doi: 10.3102/0013189X009006005

Wilkinson & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604. doi: 10.1037/0003-066X.54.8.594

*Yi, M. Y. & Hwang, Y. (2003). Predicting the use of web-based information systems: self-efficacy, enjoyment, learning goal orientation, and the technology acceptance model. *International Journal of Human-Computer Studies, 59*, 431-448. doi:10.1016/S1071-5819(03)00114-9

Zientek, L. R. & Thompson, B. (2009). Matrix summaries improve research reports: Secondary analyses using published literature. *Educational Researcher, 38*(5), 343-353. doi: 10.3102/0013189X09339056

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

APPENDICES

Appendix A. Expanded Literature Review

Previous reviews have appropriately investigated the empirical nature, research methods, and experimental designs used in educational technology. Educational technology researchers captured information about the literature's empirical and non-empirical nature. Chen and Hirschheim (2004) reviewed eight major information systems journals between 1991 and 2001. Among the 1893 articles, 60% (n = 1131) were empirical and 40% (n = 762) were non-empirical studies. Chen and Hirschheim (2004) echo findings from Farhoomand and Drury (1999) who examined 2098 articles published between 1985 and 1996 with 61% empirical and 39% non-empirical studies. Likewise, more recent findings from Hrastinski and Keller (2007) supported this ratio. Among the 660 articles examined, 68% were empirical studies, while the remaining 32% were non-empirical studies. Overall, the balance between empirical and non-empirical articles remains steady over the past decade.

Many researchers have identified the research methods used in educational technology. Koble and Bunker (1997) reviewed the research methods used in *The American Journal of Distance Education* between 1987 and 1995. Among the 129 articles, 28.7% were classified as quantitative studies, 4.7% were qualitative studies, 4.7%, were literature reviews, and 1.5% used mixed methods. Rourke and Szabo (2002) reviewed the *Journal of Distance Education* from 1986 to 2000. Among the 49 empirical studies, 25% used quantitative methods, 31% used qualitative methods, and 31% use mixed methods. Hrastinski and Keller (2007) reviewed 660 educational technology articles between 2000 and 2004 and found 51% used quantitative methods, 25% used qualitative methods and 24% used mixed methods. Randolph et al. (2008) reviewed 352 articles from eight major peer-reviewed computer science education publications published between 2000 and 2005. Among the 144 reviewed articles, 74.3%, 95% CI [68.1, 80.2] were quantitative, 15.3%, CI [10.4, 20.8] were qualitative, and 10.4%, CI [6.3, 14.6] were mixed methods. Alper and Gülbahar (2009) found between 2004 and 2006 in the *Turkish Online Journal of Educational Technology* (TOJET) a shift to more qualitative methods, with only 40% using quantitative methods. Although, most studies in TOJET during 2007 were quantitative. Şİmşek et al. (2009) reviewed master's theses in educational technology completed in Turkey between 2000 and 2007. Among the 259 theses, 204 (79%) used quantitative methods, 21 (8%) used qualitative methods and 34 (13%) used mixed methods. More recently, Wang and Lockee (2010) conducted a content analysis on four studies on virtual worlds in distance education. All studies used qualitative research methods. As demonstrated here, there is no lack of empirical information on the research methods used in educational technology. Research methods appear to fluctuate over time and varies based on review's literature specificity.

Researchers have also identified the experimental designs used, despite the lack of explicit reporting in the original article (Cheung & Hew, 2009; Koble & Bunker, 1997; Peterson-Karlan, 2011; Randolph et al., 2008; Shih et al., 2008; Şİmşek et al., 2009). Identifying trends of experimental designs is particularly difficult given the variation in researcher's classification schemes. Specifically, a synthesis among five reviews resulted in 14 different experimental designs, with some categories overlapping one another based on the original review's classification scheme. Additionally, reviewers blurred the lines between types of experimental designs and types of research studies by grouping the two together. For example, Cheung and Hew (2009) calculated the frequency of both from 44 articles on mobile handheld devices used in educational settings. Among the experimental designs classified, 11.4% used one group pretest-posttest design, 6.8% used more than one design, 4.5% used quasi-experimental design, 4.5% used experimental design, and 2.3% single subject design. Among the types of studies categorized, 65.9% were descriptive, 2.3% ex-post facto design, and 2.3% design-based research. Moreover, researchers often encounter articles using more than one experimental design. For example, Randolph and colleagues (2008) classified 183 designs from 144 computer science articles. Among the 144 articles, 93 (64.6%), 95% CI [58.3, 70.8] articles used experimental or quasi-experimental designs, 38 (26.4%), 95% CI [20.8, 31.3] used qualitative designs, 26 (18.1%), 95% CI [13.2, 22.9] used causal comparative designs, 15 (10.4%) 95% CI [7.0, 14.6] used correlational design, and 11 (7.6%), 95% CI [4.2, 11.1] used surveys. ŞİmŞek et al. (2009) reviewed 204 quantitative articles and found 55% (n = 160) used surveys while only 27% (n = 79) used experimental designs. More recently, Peterson-Karlan (2011) reviewed 85 studies, 21% (n = 18) were experimental, 20% (n = 17) were case studies, 14% (n = 12) were quasi-experimental, and 15% (n = 13) single-subject designs, and 29% (n = 25) did not report the experimental design used. Conversely, Koble and Bunker (1997) listed the types of studies and experimental designs. Koble and Bunker (1997) found in 37 quantitative articles, researchers used surveys, quasi-experimental designs, and experimental designs. Table A1 presents experimental designs classified in previous reviews.

Few researchers have reviewed sampling methods, score reliability, or statistical techniques. Researchers have estimated the sampling methods used despite the lack of reporting. Edyburn (2000) examined 114 articles from 26 special education technology in 1999. Edyburn concluded most articles used convenience samples given the "over-abundance" in post-secondary samples (p.13). Randolph et al. (2008) reviewed 144 reviewed articles and found 124 (86.1%) used convenience sample, 14 (9.7%) articles used a purposive sample and only 6 (4.2%) used a random sample. Alper and Gülbahar (2009) reviewed *The Turkish Online Journal of Educational Technology* between 2003 and 2007. Among the 98 articles Alper and Gülbahar (2009) reviewed which excluded literature reviews and discussions, most used a convenience sample (65.3%, n = 64), 22 (22.4%) used a cluster sample, and 12 (12.2 %) used a random sample.

Few researchers have captured the extent to which reliability is reported. Overall, the few studies which reviewed reliability information found poor reporting practice. For example, Randolph (2008) reported on K-12 computer science education programs. Among the 29 reports, only one report conveyed reliability or validity estimates. Randolph (2008) found among the 107 articles reviewed for reliability, only 13 articles (12.1%) reported reliability or validity information about the scores produced by instruments. Likewise, although Lee et al. (2004) and (2007) did not report specifically on reliability, the authors concluded reliability and validity issues were not reported in the sample under review.

Table A1

Frequency of Experimental Designs Across Studies

	Experimental Designs						
	Quasi-experimental	Experimental	Experimental or Quasi-experimental	One group pretest-posttest	Single subject	More than one design	Unknown
Koble & Bunker (1997) [a]	x	x					
Randolph, et al. (2008) [b]			64.6				
Cheung & Hew (2009) [c]	4.5	4.5					
Simsek, et al. (2009) [d]	27.0			11.4	2.3	6.8	
Peterson-Karlan (2011) [e]	14.0	21.0			15.0		29.0

Notes. [a] n = 129; [b] n = 352; [c] n = 44; [d] n = 259; [e] n = 249.

Finally, there is a dearth of reviews on statistical methods used in educational technology literature. Moreover, the few researchers who evaluated the statistical methods, reviewed small samples. In addition, some of these reviews have examined trends over time and across journals, while other reviewers did not observe trends.

Despite the limitations, these results offer a glimpse of the field's statistical techniques. For example, Lee et al. (2004) examined 383 articles in four distance education journals between 1997 and 2002. Among the 47 experimental studies reviewed for statistical techniques, 8 (17.0%) used ANOVA or ANCOVA, 8 used regression analysis, 8 used chi-square test, 5 (11%) reported correlations, 8 (17.0%) used factor analysis, 4 (8.5%) used t-test, 2 (4.3%) used path analysis, 1 (2.1%) used MANOVA or MANCOVA and 1 used a cluster analysis. Three years later, similar results were found when Lee et al. (2007) updated the review to examine 553 articles from 1997 to 2005. Among the 86 quantitative studies examined, 14 (16%) articles used a t-test; 14 (16%) used ANOVA or ANCOVA; 14 (16%) conducted a factor analysis; 13 (15%) used a regression analysis; 11 (14%) used a chi-square test; 7 (9%) reported correlations; 4 used other methods such as discriminant analysis, Mann-Whitney U-test, and structural equation modeling; 3 used MANOVA or MANCOVA; 2 (2%) used path analysis; and 1 (1%) used a cluster analysis.

Karataş (2008) also reviewed statistics used in 25 articles from three distance education journals between 2003 and 2005. Karataş found 5 (20%) articles reported percentages, 2 (8%) articles reported frequencies, and 3 (12%) reported correlations. Among the statistical analyses conducted, 3 (12%) studies used an ANOVA, 3 (12%)

used a t-test, 1 (4%) used a z-test, 4 (16%) used factor analysis, 2 (8%) used a MANOVA, 2 (8%) used multiple regression, 1 (4%) used structural equation modeling, and 2 (8%) conducted a cross-tabulation. Likewise, Shih et al. (2008) examined 444 articles intersecting cognition and e-learning found in five educational technology journals between 2001 and 2005. Among the 16 articles reviewed for statistical analyses, 11 (68.75%) included descriptive statistics, 6 (37.5%) used ANOVA with only 1 reporting post hoc tests, 4 (25%) used t-tests, 2 (12.5%) included frequencies, 2 (12.5%) included ANCOVA, 1 (6.3%) conducted a factor analysis, 1 (6.3%) used a MANCOVA, and 1 (6.3%) used a chi-square test.

Randolph et al. (2008) conducted the most comprehensive review of statistical techniques used in computer science education. Among the 123 articles reviewed, 44 (35.8%) used inferential statistics. Randolph and colleagues (2008) further classified the inferential statistics used. Among the 44 articles using inferential statistics, 25 (56.8%), 95% CI [47.7, 65.9] used parametric analyses, 13 (29.5%), 95% CI [23.3, 37.2] used correlational analyses, 11 (25.0%), 95% CI [13.2, 31.8] used nonparametric analyses, 2 (4.5%), 95% CI [0.0, 9.1] used small sample analysis, and 1 (2.3%), 95% CI [0.0, 2.3] used a multivariate analysis. In addition, Randolph and colleagues found among the 25 articles using parametric statistics, 15 (60.0%), 95% CI [48.0, 72.0] reported central tendency and dispersion. Moreover, among the 13 articles using correlational analyses, 10 (76.9%), 95% CI [53.9, 92.3] articles reported sample size, while only 5 (38.5%), 95% CI [15.4, 61.5] reported correlation or covariance matrices. Additionally, the one article using multivariate analysis did not report adequate information to interpret the analysis such as cell means, cell sample size, or a matrix of associations. Furthermore, among the 123 quantitative articles reviewed, 97.6% (n = 120) reported an effect size. The remaining the three articles reported only probability values or stated whether or not the resulting effect was statistically significant or not. More specifically, among the 120 articles which reported effect sizes, 97.5% (n = 117), 95% CI [95.0, 100.0] were raw differences, 6.7% (n = 8), 95% CI [3.3, 6.7] were correlational effect sizes, and 5.0% were standardized mean differences (n = 6), 95% CI [1.7, 8.3]. Table A2 and Table A3 presents a list univariate and multivariate statistical techniques reported here, respectively.

Table A2

Frequency of Univariate Statistical Techniques Used

	%	Frequency	Cross tab	Descriptive	Central tendency & dispersion	Correlation	χ^2 test	Z-test	T-test	Effect size	ANOVA or ANCOVA	ANCOVA	Regression
Lee et al. (2004)						11	17		8.5		17		17
Lee et al. (2007)						9	14		16		16		15
Karatas (2008)	20	8	8			12		4	12		12		8
Shih et al. (2008)		12.5		68.75			6.3		25		37.5	12.5	
Randolph et. al (2008)					60					97.6			

Table A3

Frequency of Multivariate Statistical Techniques Used

	MANOVA or MANCOVA	MANCOVA	Path analysis	Factor analysis	Cluster analysis	SEM	Other methods
Lee et al. (2004) [a]	2.1		4.3	17	2.1		
Lee et al. (2007) [b]	2		2	16	1		4.7
Karatas (2008) [c]	8			16		4	
Shih et al. (2008) [d]		6.3		6.3			

Notes. [a] n = 383; [b] n = 553; [c] n = 25; [d] n = 16. The category "other methods" included an unspecified frequency of the following analyses: discriminant analysis, Mann-Whitney U-test, and structural equation modeling.

Appendix B. Coding Sheet for Study 1

Coding Template (adapted from Warne et al. (2012).

Elements	Categories	Definition
Basic Information	Journal	Pulled from RefWorks database.
	Year	Pulled from RefWorks database.
	Volume	Pulled from RefWorks database.
	Issue	Pulled from RefWorks database.
	Author(s)	Pulled from RefWorks database.
	Title	Pulled from RefWorks database.
Research Method	Quantitative	Quantitative Research "uses numerical analysis to illustrate the relationship among factors in the phenomenon studied” (Chen & Hirschheim, 2004, p. 205)
	Qualitative	Qualitative research is "an inquiry process of understanding based on a distinct methodological tradition of inquiry that explores a social or human problem. The researcher builds a complex, holistic picture, analyzes words, reports detailed views of informants, and conducts the study in a natural setting." (Creswell, 2007, p. 249)
	Mixed Methods	Mixed methods research is “where the researcher mixes or combines quantitative and qualitative research techniques, methods, approaches, concepts or language into a single study” (Johnson & Onwuegbuzie, 2004, p. 17).
	Non-empirical	Non-empirical articles are ancillary materials, such as editorials, book reviews, methodological papers, discussions, and acknowledgements, which support empirical articles included in the journal issue.
Participant Characteristics	Size	Total intended sample size. At times, the sample size will differ from the analysis. In such cases, note the intended sample size.
	Demo. (list race/ethnicity %s)	Demographic characteristics reported (e.g., ethnicity, gender, age, level of education, etc.)
Sampling Procedures	No sampling procedure listed	
	Convenience sampling	A convenience sample is a sample selected which "suits the purposes of the study and that is convenient" (Gall, Gall, & Borg, 2007, p. 175). When a convenience sample is used, the reader "must infer a population to which the results might generalize" (Gall et al., 2007, p. 175). When authors provide a detailed description of the sample, the reader can more easily infer the population.
	Simple random sampling	"A simple random sample is a group of individuals drawn by a procedure in which all the individuals in the defined population have an equal and independent chance of being select as a member of the sample" (Gall et al., 2007, p. 170).
	Systematic random sampling	Systematic random sampling involves selecting a sample from a list and selecting every nth person in the list (Gall et al., 2007).
	Stratified random sampling	"A stratified random sample involves a sample selected so that certain subgroups in the population are adequately represented in the sample" (Gall et al., 2007, p. 173). Other types of stratified random samples include: proportional stratified random sampling and nonproportional stratified random sampling. If the author(s) list either of these, type "proportional" in the stratified random sampling column.
	Cluster sampling	"In cluster sampling, the unit of sampling is a naturally occurring group of individuals. Cluster sampling is used when it is more feasible to select groups of individuals rather than individuals from a defined population" (Gall et al., 2007, p. 173). Another type of cluster sampling is multistage cluster sampling. If the author(s) specify multistage sampling, type multistage sampling in the cluster sampling column.
	Database (list database)	Select if the data comes from archival data set and type in the name of the database or other information provided by the

Coding Template (adapted from Warne et al. (2012).

Elements	Categories	Definition
		author(s).
	Other sampling method (list)	Select if any other type of sampling method is used that is not listed in the sampling procedures category. List the name of the sampling method used under the "Other sampling method" column.
	Random Assignment	Select if the sample was randomly assigned to groups. If not randomly assigned or the author does not state whether the sample was randomly assigned, leave the cell blank.
Descriptive Statistics	Descriptive Statistics	Descriptive statistics include: 1) location or central tendency (e.g., mean, median, mode); 2) dispersion (e.g., range, sum of squares, standard deviation (SD), variance); and 3) shape (e.g., skewness, kurtosis). Note. Relationship statistics such as Pearson r are listed separately in the coding sheet. These statistics should be noted in correlational statistics section instead.
	Effect size reported?	List all types of effect sizes reported.
	Effect Size Value	List all effect size values
Inferential statistics	t-test	
	paired t-test	
	ANOVA	
	ANCOVA	
	ANOVA/ ANCOVA post hoc tests (specific)	Record the name of the post hoc test used followed by page number.
	MANOVA	
	MANCOVA	
	MANOVA/MANCOVA post hoc tests (specific)	Record the name of the post hoc test used followed by page number.
	Descriptive discriminant analysis (DDA)	
	Exact p-values	If exact p-values (e.g. p = .01 or p < .001) are reported in the text, select the cell. If non-exact p-values (e.g., p < .05) are reported in the note section of the table, consider this an exact value. APA (2010) allows reporting of non-exact p-values in the notes section of the table. If the only exact p-value reported is p =.000, do not select this cell. Instead, select the column, p = .000.
	Non-exact p-values	If non-exact p-values (e.g., p < .05) are reported in the text, select the cell. If non-exact p-values are reported in the table, do not select the cell. APA (2010) allows reporting of non-exact p-values in the notes section of the table.
	p = .000?	If the p-value reported is p =.000, select this cell.
	Confidence Intervals	
Correlational statistics	Pearson's r	Also known as zero-order correlation coefficient. When Pearson r is only used for inter-rater reliability. Do not select this cell. Instead, select the statistic as inter-rater reliability only.
	Spearman's rho	
	Tetrachoric correlation	
	Biserial correlation	
	Point-biserial correlation	
	Phi coefficient	
Regression	Multiple regression	
	Stepwise regression	
	Logistic regression	
	Hierarchical linear regression	

Coding Template (adapted from Warne et al. (2012).

Elements	Categories	Definition
Data Reduction	HLM	
	Commonality analysis	
	EFA or PCA?	
	Extraction method	e.g., maximum likelihood analysis (or canonical factor analysis), alpha factor analysis, image factor analysis, principle axes factor analysis, principal axes factor analysis with iterated communalities (or least squares).
	Rotation method	e.g. Orthogonal rotations include varimax, equamax, and quartimax; and oblique rotations - promax, Procrustean, oblimin, and direct oblimin
	Factor retention method (list all used)	e.g. scree plot, Guttman rule (or the K1 rule, or eigenvalue-greater-than-one rule), variance accounted by the number of factors, Kaiser–Meyer–Olkin test of sampling adequacy, a priori theory, parallel analysis, Bartlett's test of sphericity, visual inspection of the item loadings, and inspection of the residual correlation matrix
	Loading matrix reported?	When an orthogonal rotation method is used, we expect to see a loading matrix reported. If reported, place an x and page number. If not reported, leave the cell blank.
	Structure matrix reported? (Oblique rotation)	When a nonorthogonal rotation method (oblique rotation method) is used, we expect to see a structure matrix reported. If reported, place an x and page number. If not reported, leave the cell blank.
SEM	Path analysis	Models specifying relationships between observed variables only. The model does not include latent variables. Authors should use boxes for observed variables and circles for latent variables; however, not all authors follow this rule.
	CFA/ Measurement model	The measurement model is the part which relates measured variables to latent variables. Authors should use boxes for observed variables and circles for latent variables; however, not all authors follow this rule.
	Structural model	The structural model is the part that relates latent variables to one another. Authors should use boxes for observed variables and circles for latent variables; however, not all authors follow this rule.
	Test of invariance	
	Covariance matrix?	
	Estimation method (specific)	e.g., maximum likelihood, weighted least squares mean/variance adjusted, robust ML, etc.
	Examined normality of data?	
	Identification method?	
	Standardized or unstandardized results?	
	Fit statistics used	List fit statistics used and separate each by a comma. Record the page number each statistic is found. [e.g. (X2 , CFI, RMSEA, SRMR, goodness-of-fit index (GFI), nonnormed fit index (NNFI), expected cross-validation index (ECVI), Tucker–Lewis index (TLI), adjusted goodness-of-fit index (AGFI), normed fit index (NFI), root mean residual (RMR), the Satorra–Benter adjusted X2, weighted root mean residual (WRMR), relative fit index (RFI), robust CFI, and the robust RMSEA, p. 121)]
Nonparametric statistics	Frequencies	
	Cross-tabulations	
	Non-parametric NHSST (specific)	List which specific non-parametric inference tests were conducted (e.g., Pearson's $\chi2$, Likelihood ratio, linear-by-linear association, Cochran-Mantel-Haenszel test, and Kruskal-Wallis test).
Reliability	No mention of reliability	

Coding Template (adapted from Warne et al. (2012).

Elements	Categories	Definition
	Reliability induction	Reliability induction is when the "researcher takes a reliability statistic from one sample and applies it to another" (Warne et al., 2012). For example, from a test manual or previous study.
	Reported reliability for own data	
	"Reliability of the test" (not scores)	
	Internal Consistency	(e.g., Cronbach's alpha, KR20, KR21, or Kuder Richardson formula)
	Test-retest	
	Interrater	e.g., Cohen's kappa, percent agreement
	Alternate/ Parallel forms	
	Split half (with correction or not?)	Indicate if the split half coefficient has a correction or not.
	Other reliability measure (specific)	List any reliability measure not specified in the coding sheet (e.g., IRT-based reliability).
	Replication procedure	
Miscellaneous	Power Analysis	
	Other unlisted statistical method	List any miscellaneous statistic reported in the article that is not already listed in the coding sheet.
Notes		Notes to coder or both coders about the article.

Appendix C. Emails to Authors

Initial Email:

Greetings Dr. ____,

I hope this email reaches you well. I am a PhD candidate in the Educational Technology program at Texas A&M University, USA. I am currently collecting data for my dissertation.

I recently read your article, ____. I am very interested in the conclusions of the study and would like to include your study in my dissertation. I am writing today to ask, would you be willing to share with me [the correlation matrix (with means & standard deviations) or the covariance matrix] you and your colleagues used to conduct the structural equation model?

If you have any questions regarding my request, please don't hesitate to contact me. I appreciate any time and attention you might devote to the above request.

Thank you for your time and consideration,

Follow-up Email:

Greetings Dr. ___

I hope this email reaches you well. I previously contacted you regarding your article, ____. I am very interested in the conclusions of the study and would like to include this study in my dissertation.

Would you be willing to share your [the correlation matrix (with means & standard deviations) or the covariance matrix] with me? If not, I understand and will remove your article from my study.

Thank you again for your time and consideration.

Best regards,

Appendix D. Extended discussion on email responses.

Authors had an array of responses including: referrals to co-authors, traveling pretexts, perceptions of old research, or lost data. Authors often referred me to a co-author. Another author stated, "…My co-author, [name of co-author], completed the statistical analysis, so I will forward your email to her for a response." Unfortunately, none of the referrals responded to my request in the forwarded message. Another common response from authors was that they were traveling. One author stated, "I am travelling abroad now and I am afraid I can not find the data for the student." In another instance, an author suggested the information I requested from an article published in 2005 was too long ago. The author responded, "…I will try and check my hard disk for the data which I am not sure if I have them as it has been quite a while."

One surprising response was about an article where the data was collected in Chile and Spain during 2009. The author stated prior agreements did not allow them to share the information I requested (e.g., means and standard deviations and a covariance matrix). The author responded, "I am writing to respond negatively to your email. Unfortunately we have a prior agreement on the survey data indicating that we can not give other researchers in any form."

Lastly, authors stated limitations of the statistical software prevented them from providing this information. One author stated,

> "I am sorry very much, but I cannot send you the correlation matrix you ask me for. When we did that research, we used a software (PLS-Graph)

w[h]ich developed internally this matrix and, unfortunately, we don't have the licence [license] of this software by the time. It's impossible for us to recover those data. We have the initial data (questionnaires) …. We also have the final document of the project with similar tables from the paper you have read. … If you are interested on them, please, ask me. I will not have any problem in sending them to you. If you feel you have to remove our article from your study, there is no problem."

Although I responded to request the information mentioned by the author, I never received the information.

Appendix E. Coding Sheet for Study 2

Article Attributes
Reference Type
Author
Publication Year
Article Title
Name of Journal
Volume
Issue
Start Page
Other Pages
Keywords
Abstract
Notes
Publisher
ISSN/ISBN
Availability
Accession Number
Links
Passed Primary 1 Screening
Passed Secondary2 Screening
Study ID #
Sample Attributes
Type of Sample
Description
Location of Study
Total Sample Size
of males
of females
Age
Research Design
Major
Technology Proficiency Index
Internet experience
Type of LMS
Learning Management System
Instruments

Description
Type of Model
TAM Model
Reliability of Variables (Cronbach's alpha)
Perceived Usefulness (PU)
Ease of Use (EU)
Attitude (A)
Behavioral Intention (BI)
Actual Use (U)
Means & SDs
Perceived Usefulness (PU)
SD
Ease of Use (EU)
SD
Attitude (A)
SD
Behavioral Intention (BI)
SD
Actual Usage (U)
SD
Correlations
r (PU, BI)
r (EU, BI)
r (PU, EU)
r (BI, U)
r (EU, A)
r (PU, A)
r (EU, U)
r (PU, U)
r (A, BI)
r (A, U)
Covariances
COV (PU, BI)
COV (EU, BI)
COV (PU, EU)
COV (BI, U)
COV (EU, A)
COV (PU, A)

COV (A, BI)
COV (A, U)
COV (EU, U)
COV (PU, U)

Appendix F. Covariance Matrices for Studies Included in the Meta-Analysis

Group 1

Saadé & Galloway (2005)

	PU	EU	A	BI
PU	0.894916	0.487	0.487	0.487
EU	0.487	1.177225	0.251	0.366
A	0.487	0.251	0.857476	0.536
BI	0.487	0.366	0.536	1.089936

Saadé, Nebebe, & Tan (2007)

	PU	EU	A	BI
PU	0.7744	0.430848	0.474672	0.34848
EU	0.430848	0.9216	0.392544	0.228096
A	0.474672	0.392544	0.7569	0.45936
BI	0.34848	0.228096	0.45936	0.7744

Group 2

Almrashdah et al. (2010)

	PU	EU	BI
PU	0.816601	0.61406	0.69907
EU	0.61406	0.738001	0.588056
BI	0.69907	0.588056	1.001541

Martins & Kellermann (2004)

	PU	EU	BI
PU	1.69	0.56693	1.00646
EU	0.56693	0.7921	0.60467
BI	1.00646	0.60467	2.4964

Saadé & Bahli (2005)

	PU	EU	BI
PU	0.49	0.14	0.24
EU	0.14	0.6	0.11
BI	0.24	0.11	0.81

Saadé et al. (2007)

	PU	EU	BI
PU	0.7744	0.430848	0.34848
EU	0.430848	0.9216	0.228096
BI	0.34848	0.228096	0.7744

Yi & Hwang (2003)

	PU	EU	BI
PU	2.9584	0.852948	1.681472
EU	0.852948	2.9241	1.12518
BI	1.681472	1.12518	3.5344

Group 3

Liao & Lu (2008)

	EU	BI	U
EU	1.377806	0.64068	-0.656
BI	0.64068	1.377806	0.45417
U	-0.656	0.45417	5.24181

Martins & Kellermann (2004)

	EU	BI	U
EU	0.7921	0.60467	0.2611
BI	0.60467	2.4964	0.77262
U	0.2611	0.77262	2.6569

Yi & Hwang (2003)

	EU	BI	U
EU	2.9241	1.12518	53.30395
BI	1.12518	3.5344	66.24706
U	53.30395	66.24706	18368.38

Group 4

Brown (2002)

	PU	EU	U
PU	1.3225	0.497835	0.248285
EU	0.497835	1.2321	0.479298
U	0.248285	0.479298	1.6129

Martins & Kellermann (2004)

	PU	EU	U
PU	1.69	0.56693	0.55094
EU	0.56693	0.7921	0.2611
U	0.55094	0.2611	2.6569

Ramayah (2006)

	PU	EU	U
PU	0.5625	0.37125	0.406125
EU	0.37125	0.81	0.53865
U	0.406125	0.53865	0.9025

Yi & Hwang (2003)

	PU	EU	U
PU	2.9584	0.852948	6.993348
EU	0.852948	2.9241	53.30395
U	6.993348	53.30395	18368.38

Appendix G. Summary of Studies Using different combinations of the variables from the TAM

Table G1

Summary of Studies Using different combinations of the variables from the TAM

Study	PU - EU	PU - A	EU - A	A - BI	PU - BI	EU - BI	BI - U	PU - U	EU - U
Almrashdah et al. (2010)	x				x	x			
Liao & Lu (2008)						x	x		x
Martins & Kellermanns (2004)	x				x	x	x	x	x
Saadé & Bahli (2005)	x				x	x			
Saadé & Galloway (2005)	x	x	x	x	x	x			
Ramayah (2006)	x							x	x
Yi & Hwang (2003)	x				x	x	x	x	x
Brown (2002)	x							x	x
Saadé et al. (2007)	x	x	x	x	x	x			
Davis & Wong (2007)	x								
Saadé (2007)			x	x		x			
Lee & Lee (2008)	x								
Pituch & Lee (2006)	x								

Note. PU = perceived usefulness; EU = perceived ease of use; A = attitude; BI = behavioral intention to use; U = actual use.

Appendix H. Extended Explanation of the TSSEM Approach

The current study uses a two-stage structural equation modeling (TSSEM) approach to fit fixed-effects MASEM using covariance matrices proposed by Cheung and Chan (2009). In stage one, the covariance matrices are pooled together. In stage two, the pooled covariance matrix is used to fit the structural equation model using with weighted least squares (WLS) estimation method.

Stage 1

The purpose of the stage 1 is to obtain a pooled covariance matrix. Under the fixed-effects model, all population covariance matrices are the same. Under the assumption of homogeneity of covariance matrices, a common covariance matrix may be obtained by allowing the variance to vary across studies. When there are missing covariances, metaSEM filters out the missing data. Since the studies were grouped before the meta-analysis, there are no missing data within each group.

Stage 2

After the stage 1, a pooled covariance matrix and corresponding asymptotic covariance matrix are estimated. A structural model is fitted with weight least squares (WLS) estimation method which assumes unequal error variance and as a result gives less weight to observations with larger error variance. The likelihood-ratio statistics and multiple goodness-of-fit indices are used to judge whether the proposed structural model is appropriate. The standard errors may be used to test the significance of individual parameter estimates.

Appendix I. Syntax from R to Conduct the Meta-Analysis

Group 1

```
##INTALL PACKAGES TO RUN metaSEM
install.packages('OpenMx', repos='http://openmx.psyc.virginia.edu/packages/')
install.packages(c('ellipse','MASS'))
install.packages(pkgs="e:/metaSEM_0.8-4.zip", repos=NULL)
## LOAD THE metaSEM LIBRARY
library(metaSEM)
## LOAD DATA FILE INTO CONSOLE BY CLICKING AND DRAGGING FILE
INTO CONSOLE
## READ DATA FILE AT THIS LOCATION
PUEUABIdata <- readFullMat(file = "C:\\PUEUABIdata.dat")
## SEE DATA NAMED "PUEUABIdata"
PUEUABIdata
##WRITE MATRIX FOR SAMPLE SIZE OF EACH STUDY- Code modified from:
http://www.r-tutor.com/r-introduction/matrix
PUEUABIn = matrix (c(128, 362), nrow=1, ncol=2, byrow = TRUE)

## SEE MATRIX FOR SAMPLE SIZE OF EACH STUDY
PUEUABIn
##FIXED EFFECTS MASEM USING TSSEM - UNDER THE FIXED EFFECTS
MODEL, IT IS ASSUMED THAT ALL POPULATION COVARIANCE MATRICES
ARE THE SAME WHILE THERE ARE STUDY SPECIFIC COVARIANCE
MATRICES UNDER THE RANDOM-EFFECTS MODEL.
## STAGE 1 - OBTAINS A POOLED COVARIANCE MATRIX
## FOR DATA SET PUEUABIdata AND PUEUABIn
head(PUEUABIdata)
head (PUEUABIn)
## TSSEM: STAGE 1 - The tssem1() function is used to pool the correlation matrices
with a fixed-effects model in the first stage analysis by specifying method="FEM" in the
argument:
fixed1 <- tssem1(PUEUABIdata, PUEUABIn, method = "FEM")
## AFTER TSSEM1 RUNS ANALYSIS OF CORRELATION MATRIX (IN THIS
CASE THE COVARIANCE MATRIX), OUTPUT RESULTS
summary(fixed1)
## OUTPUTS POOLED COVARIANCE MATRIX (THE PARAMETER
ESTIMATES) EXTRACTED
coef(fixed1)
## PREPARE FOR TSSEM: STAGE 2 (RAM)
## CREATES "A MATRIX" WHICH SPECIFIES THE ASYMMETRIC PATHS
(ASYMTOTIC MATRIX)
```

```
A2 <-
matrix(c(0,"0*x2tox1",0,0,0,0,0,0,"0*x1tox3","0*x2tox3",0,0,0,0,"0*x3tox4",0),nrow=4
, ncol=4, byrow=TRUE)
dimnames(A2) <- list(c("x1", "x2", "x3", "x4"), c("x1", "x2", "x3", "x4"))
A2
A2 <- as.mxMatrix(A2)
## CREATES "S MATRIX" WHICH SPECIFIES THE SYMMETRIC
VARIANCE/COVARIANCE MATRIX
S2 <- Diag(c("0.2*Varx1",1,"0.2*Varx3", "0.2*Varx4"))
dimnames(S2) <- list(c("x1", "x2", "x3", "x4"), c("x1", "x2", "x3", "x4"))
S2
S2 <- as.mxMatrix(S2)
## TSSEM: STAGE 2
fixed2 <- tssem2(fixed1, Amatrix = A2, Smatrix = S2, diag.constraint=TRUE,
intervals.type="LB")
## AFTER TSSEM2 RUNS ANALYSIS, OUTPUT RESULTS
summary(fixed2)
```

Group 3

```
##INSTALL PACKAGES TO RUN metaSEM
install.packages('OpenMx', repos='http://openmx.psyc.virginia.edu/packages/')
install.packages(c('ellipse','MASS'))
install.packages(pkgs="e:/metaSEM_0.8-4.zip", repos=NULL)
## LOAD THE metaSEM LIBRARY
library(metaSEM)
## LOAD DATA FILE INTO CONSOLE BY CLICKING AND DRAGGING FILE
INTO CONSOLE
## READ DATA FILE AT THIS LOCATION
PUEUUdata <- readFullMat(file = "C:\\PUEUUdata.dat")
## SEE DATA NAMED "PUEUAUdata"
PUEUUdata
##WRITE MATRIX FOR SAMPLE SIZE OF EACH STUDY- Code modified from:
http://www.r-tutor.com/r-introduction/matrix
PUEUUn = matrix (c(73, 243, 275, 109), nrow=1, ncol=4, byrow = TRUE)
## SEE MATRIX FOR SAMPLE SIZE OF EACH STUDY
PUEUUn
##FIXED EFFECTS MASEM USING TSSEM - UNDER THE FIXED EFFECTS
MODEL, IT IS ASSUMED THAT ALL POPULATION COVARIANCE MATRICES
ARE THE SAME WHILE THERE ARE STUDY SPECIFIC COVARIANCE
MATRICES UNDER THE RANDOM-EFFECTS MODEL.
## STAGE 1 - OBTAINS A POOLED COVARIANCE MATRIX
## FOR DATA SET PUEUABIdata AND PUEUABIn
head(PUEUUdata)
```

```
head (PUEUAUn)
## TSSEM: STAGE 1 -  The tssem1() function is used to pool the correlation matrices
with a fixed-effects model in the first stage analysis by specifying method="FEM" in the
argument:
fixed1 <- tssem1(PUEUUdata, PUEUUn, method = "FEM")
## AFTER TSSEM1 RUNS ANALYSIS OF CORRELATION MATRIX (IN THIS
CASE THE COVARIANCE MATRIX), OUTPUT RESULTS
summary(fixed1)
## OUTPUTS POOLED COVARIANCE MATRIX (THE PARAMETER
ESTIMATES) EXTRACTED
coef(fixed1)
##RANDOM EFFECTS MASEM USING TSSEM
random1 <- tssem1(PUEUUdata, PUEUUn, method = "REM")
summary(random1)
##TSSEM - STAGE 2
## CREATES "A MATRIX" WHICH SPECIFIES THE ASYMMETRIC PATHS
(ASYMTOTIC MATRIX)
A2 <- matrix(c(0,"0*x2tox1",0,0,0,0,"0*x1tox3","0*x2tox3",0),nrow=3, ncol=3,
byrow=TRUE)
dimnames(A2) <- list(c("x1", "x2", "x3"), c("x1", "x2", "x3"))
A2
A2 <- as.mxMatrix(A2)
## CREATES "S MATRIX" WHICH SPECIFIES THE SYMMETRIC VARIANCE/
COVARIANCE MATRIX
S2 <- Diag(c("0.2*Varx1",1,"0.2*Varx3"))
dimnames(S2) <- list(c("x1", "x2", "x3"), c("x1", "x2", "x3"))
S2
S2 <- as.mxMatrix(S2)
## TSSEM: STAGE 2 – RANDOM EFFECTS
random2 <- tssem2(random1, Amatrix = A2, Smatrix = S2, diag.constraint=TRUE,
intervals.type="LB")
## AFTER TSSEM2 RUNS ANALYSIS, OUTPUT RESULTS
summary(random2)
```