A Survey of Mathematics Education Technology Dissertation Scope and Quality: 1968–2009

Citation	Ronau, R. N., Rakes, C. R., Bush, S. B., Driskell, S. O., Niess, M. L., & Pugalee, D. K. (2014). A Survey of Mathematics Education Technology Dissertation Scope and Quality 1968–2009. American Educational Research Journal, 51(5), 974-1006. doi:10.3102/0002831214531813
DOI	10.3102/0002831214531813
Publisher	SAGE Publications
Version	Accepted Manuscript
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

A Survey of Mathematics Education Technology Dissertation Scope and Quality: 1968-2009

Robert N. Ronau
University of Cincinnati

Christopher R. Rakes
University of Maryland, Baltimore County

Sarah B. Bush[1]
Bellarmine University

Shannon O. Driskell[1]
University of Dayton

Margaret L. Niess[1]
Oregon State University

David Pugalee[1]
University of North Carolina-Charlotte

Correspondence concerning this article should be addressed to:

Robert N. Ronau
PO Box 210002
School of Education, CECH, Room 610d
University of Cincinnati
Cincinnati, OH 45221

Phone: 502-693-1114
Email: bob.ronau@uc.edu

[1] These authors contributed to the paper equally.

Robert N. Ronau is a Visiting Professor in the School of Education at the University of Cincinnati. His research focuses on and teacher preparation and assessment, teacher knowledge, and the use of technology to improve mathematics instruction.

Christopher R. Rakes is an Assistant Professor in the Department of Education at the University of Maryland, Baltimore County. His research focuses on methods for improving student learning in mathematics (especially addressing misconceptions), teacher knowledge and the role of teacher knowledge in classroom practice, and the use of technology in mathematics education.

Sarah B. Bush is an Assistant Professor in the School of Education at Bellarmine University. Her research focuses on students' misconceptions in the learning of algebra and role of technology in the teaching and learning of mathematics.

Shannon O. Driskell is an Associate Professor in the Department of Mathematics at the University of Dayton, Dayton, OH. Her research focuses on enhancing mathematical pedagogical knowledge for teachers of grades 9-12.

Margaret L. Niess is a Professor Emeritus of Science and Mathematics Education in the Department of Science and Mathematics Education at Oregon State University. Her research focuses on integrating technology in education, specifically in the preparation of mathematics and science teachers for teaching with technology.

David Pugalee is a Professor of Education at the University of North Carolina Charlotte. His research focuses on the relationship between language and mathematics teaching and learning.

Abstract

We examined 480 dissertations on the use of technology in mathematics education and developed a Quality Framework (QF) that provided structure to consistently define and measure quality. Dissertation studies earned an average of 64.4% of the possible quality points across all methodology types compared to studies in journals that averaged 47.2%. Doctoral students as well as their mentors can play a pivotal role in increasing the quality of research in this area by attending to the QF categories as they plan, design, implement, and complete their dissertation studies. These results imply that the mathematics education research community should demand greater clarity in its published papers through the preparation of their own manuscripts and how they review the works of others.

Key Words: Mathematics Education, Technology, Dissertations, Research Quality, Quality Framework

Includes 3 Tables and 4 Figures

Dissertation Studies in Mathematics Education Technology: 1968-2009

The introduction of technology to education has historically been met with resistance. Over the last four decades, research has led to consistent findings that digital technologies such as calculators and computer software improve student understanding and do no harm to student computational skills (e.g., Hembree & Dessart, 1986; Ellington, 2003, 2006).

Yet traditional mathematics teaching fails to fully take advantage of such research findings (Battista, 1999; Stigler & Hiebert, 1997). One possible explanation for this resistance may be a general perception of low quality research in the education field. A key leverage point for improving the quality of education research is in the doctoral program, in which doctoral candidates are guided in a process where they conduct high-quality research through the dissertation.

Dissertation studies often make novel and meaningful contributions to their fields (Auerbach, 2011; Lovitts, 2008). These studies serve to indicate how new researchers are thinking about the field (as in Rutkienė & Teresevičienė, 2010), how doctoral advisors are guiding their students (as in Hilmer & Hilmer, 2011), and how the baseline for research quality is considered by the field (as in Colwell, 1969; Shirmer, 2008). Systematic reviews of these studies are needed to identify patterns and to determine whether and how a field is advancing. Without such reviews, the ability of doctoral advisors to effectively guide their students is severely limited, which ultimately impedes the advancement of the field. Our search for such a review of mathematics education technology dissertation studies revealed no reviews currently available (Ronau et al., 2012).

We therefore set out to examine mathematics education technology dissertations by considering the types of technologies addressed, how outcomes have or have not changed, and how the quality of the studies has changed between 1968 and 2009 by considering two broad questions, separated into eight specific sub-questions.

1. What is the scope of mathematics education technology dissertations?
    a. How has the portion of mathematics education technology dissertations within all mathematics education dissertations changed over time?
    b. What types of technologies have been studied in mathematics education technology dissertations?
    c. What types of outcomes have been associated with the use of mathematics education technology dissertations? Have these outcomes changed over time, and if so, how?
2. What is the quality of mathematics education technology dissertation research?
    a. How consistently and to what degree are mathematics education technology dissertations aligned to explicit theoretical connections?
    b. How consistently are research design and validity issues addressed by mathematics education technology dissertations?
    c. How consistently do mathematics education technology dissertations explicitly describe the reliability and validity/trustworthiness of the outcome measures?
    d. How do mathematics education technology dissertations compare to other mathematics education technology literature in terms of providing explicit information about theoretical connections, design and validity, and measurement reliability and validity/trustworthiness?

e. How well do mathematics education technology dissertations provide explicit theoretical connection, design and validity, and measurement reliability and validity/trustworthiness information across research types?

Such information is important for minimizing unintended duplication and ensuring that new technologies receive the appropriate amount of attention in mathematics education literature. The information presented through this present study is especially timely given the vast increase of technologies that can be used to teach and learn mathematics available in the last decade (e.g., cell phones with graphing calculator applications) as well as the emergence of frameworks to explain how teacher knowledge about technology develops (e.g., Technological Pedagogical Content Knowledge [TPACK]; Mishra & Koehler, 2006; Niess, 2005) and the development of standards to guide how teachers and students use technology in mathematics education (e.g., Association of Mathematics Teacher Educators, 2009; Common Core State Standards for Mathematical Practice, 2010).

Background

To lay a foundation for the present study, we searched for studies that analyzed dissertations. Search terms such as “dissertation,” “mathematics,” and “education” in a wide array of education databases (e.g., Academic Search Premier, ERIC, PsychInfo, and Education Full Text, ProQuest Research Library, ProQuest Digital Dissertations) produced no studies that had examined mathematics education technology dissertations. We, therefore, turned to studies examining dissertation research from other fields as the foundational body of literature for the present study. We noted that the lack of such research leaves a gap in the field that limits the ability of researchers to focus on issues advancing the quality of the field, thus providing an important impetus for the current study.

Dissertation Quality

Publication of dissertation data in subsequent articles expands the influence of dissertation research; without subsequent publications, few researchers attend to the results from the unpublished dissertations (Conn, 2008; Thanheiser, Ellis, & Herbel-Eisnmann, 2012). Conn (2008) speculated that researchers may disregard dissertation research because of a perceived lack in quality, a notion which can be seen in Shakeshaft (1982). Shakeshaft examined 114 dissertations that studied the role of women as education administrators. Her framework looked at the topics researched, the role of the researcher (principal investigator and institution characteristics), research design, data sources, research strategies, data analysis methodology, and the quality of the research. Quality of the research was measured through an examination of eight characteristics, which are summarized below:

1. The degree to which the abstract discussed the sample, design, statistical analyses, problem statement, and findings.
2. The degree to which the literature review was independent of discussions about methodologies and instruments.
3. The use of a probability sample and the degree to which the limitations of the sampling plan were discussed.
4. The degree to which the validity and reliability of the research instrument were addressed.
5. The degree to which practical significance was differentiated from statistical significance.
6. The choice of a survey instrument such as a questionnaire was justified and appropriate.

7. The degree to which sexist language was embedded in the study or instruments.
8. The degree to which the findings aligned with the results and could contribute to the literature (Shakeshaft, 1982).

Shakeshaft (1982) concluded that the quality of her sample of dissertations was generally low, averaging 75.61 out of a possible 100 points across eight categories with a standard deviation of 7.24 points. While the content of the dissertations was limited to women's roles in educational administration, and the subjective nature of these quality measures limited the reliability and validity of the findings, the framework provides a foundation for considering systematic ways to evaluate the quality of dissertations.

The methods for examining quality used by Adams and White (1994) were more subjective than those of Shakeshaft (1982). They measured the quality of 830 dissertations across six fields (public administration, management, planning, criminology, social work, and women's studies) by applying a framework that consisted of (a) the existence of a framework to guide the study, (b) obvious flaws in the design or methodology (e.g., "sample that was egregiously too small with which to draw reasonable conclusions, generalization of findings from a single case study, the use of an obviously inappropriate statistic,…, blatant errors in logic," p. 567), (c) relevance to theory or practice, and (d) topic importance to the field. We noted that the criterion *obvious* flaws is highly subjective, as is whether a sample is *egregiously* too small, whether a statistic is *obviously* inappropriate, or whether a chain of logic has a *blatant* error. With these levels of subjectivity, the validity and reliability of their findings may be subject to more questions than those of a more empirical approach. We noted, however, that even with such threats to validity in their measurement, their finding, that dissertations lack sufficient theoretical connections to make significant contributions to their fields, was consistent with Shakeshaft's (1982) findings on dissertation quality in women as education administrators. Additionally, their framework offered a distinct way to think about which constructs are important for assessing quality.

Other studies examining dissertation quality offered less detailed frameworks, but as a body of literature pursuing the same goal (i.e., measuring dissertation quality), their findings were consistent with those of Shakeshaft (1982) and Adams and White (1994): Dissertation quality is not sufficient to make original, meaningful contributions to the field. For example, Hallinger (2011) examined 130 dissertations within the field of education leadership reporting mixed results. He found that the use of research methodologies had improved over a thirty year period, but the conceptual frameworks and methodology development were insufficient for expanding the knowledge base of the field.

In contrast, Karadağ (2011) focused his examination of the quality of education dissertations on the way each study addressed the validity and reliability of the instruments used in the study. He examined 211 education dissertations conducted at universities in Turkey and found that instrument validity and reliability was either under-developed or entirely missing. Like Shakeshaft (1982), Adams and White (1994), and Hallinger (2011), Karadağ concluded that dissertation quality was insufficient for advancing the field.

Rutkienė and Teresevičienė (2010) evaluated the research quality of an unspecified number of dissertation studies from universities in Lithuania. They examined research quality (defined as the ability of a study to produce valid and reliable findings) by considering the design of the research, the statistical analysis methods prior to data collection, and the validity and reliability of instruments. They concluded that doctoral students typically misunderstood the steps needed to produce a high quality study. For example, experimental methods were often confused with quasi-experimental methods, threats to validity were not recognized (or they were recognized but not addressed), and samples were not identified as being convenience samples. We noted that although these studies recognized the role of the dissertation committee to mentor the new researcher, they did not find that such mentoring resulted in high quality studies.

We especially noticed a consistent lack of positive findings for dissertation quality. Only Hallinger (2011) noted improvements in the research methodology quality, but this finding was overshadowed by another finding, that dissertations lack sufficient attention to conceptual frameworks, which limits their connectivity and thereby usefulness to the literature knowledge base. These studies consistently noted a lack of theoretical connections, attention to the content of the field, and a lack of proper planning and implementation of the investigation.

The findings from these various studies on dissertations led us to consider how dissertations in mathematics education technology research address the scope of the field (Questions 1a-1c). We also wanted to examine the quality of the research (Questions 2a-2e), but we sought a measure that would be less subjective than those offered by previous research and less prone to our own personal biases. We therefore developed and applied a Quality Framework (QF) to evaluate the degree to which mathematics education technology dissertations addressed the critical components for producing reliable, valid, and useful findings.

Quality Framework (QF)

We designed the QF (Figure 1) to capture how well papers identified important information for determining the credibility and usefulness of the results (Rakes, 2012). The QF and its accompanying quality measure represents an effort to discuss the quality of reporting in a way that honors a wide array of purposes and methods, making it an important first step toward establishing a reliable and valid method for discussing the quality of mathematics education technology dissertations. The quality measure is reported as a percent of points within each research type; therefore, it can be used to determine how well dissertations met the quality criteria within its methodology type. That is, mixed methodology studies can be compared to other mixed methodology studies, but not with quantitative studies. In general, the QF cannot be used to assess whether one type of methodology is better than another, but it can compare how well studies met the criteria within their own methodology types. The percent comparisons, however, do permit us to rank studies according to quality across methodology types. We therefore used the QF to examine dissertation quality based on the chosen methodology and did not evaluate the appropriateness of the methodology choice.

[Insert Figure 1 Here]

QF divides quality indicators into categories based on literature about research quality in dissertations. Hallinger (2011) identified four weaknesses in the quality of dissertations: lack of clearly articulated conceptual frameworks for studying the constructs of interest, lack of theoretical models, valid and reliable instrumentation, and reliance on weak research designs for making causal inferences. Based on these weaknesses, we examined a wide array of texts to identify the kinds of information that should be reported within each of these categories. Because theoretical models and conceptual frameworks are often the same or highly overlapping (Britt, 1997), we combined them into a single category for the QF, *Theoretical Connections*. This category consists of both literature support and conceptual framework connections, and it measures the degree to which a study follows the second of Shavelson and Towne’s (2002) principles of scientific research, linking research to relevant theory. Regarding Hallinger’s (2011) category of research designs for making causal inferences, we recognized that requiring causal designs as a measure of quality is too restrictive because, as Schoenfeld (2006) explained, such designs have limited ability to detect important outcomes in mathematics education such as those described by the National Research Council (2001): conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition. We therefore avoided rating one design as being stronger than another. We recognize that some designs provide better evidence for making causal inferences, but making causal linkages is only one purpose found in scholarly literature (e.g., a study might explore theoretical possibilities or examine factors leading to a phenomenon). We therefore looked at designs through the lens of Shavelson and Towne's (2002) third and fourth principles, using methods that permit direct investigation of a question and providing a coherent and explicit chain of reasoning. As a result, we revised Hallinger's (2011) criterion for our QF to be *Design and Validity* rather than Research Designs for Causal Inference.

In addition to the importance of design, Shadish, Cook, and Campbell (2002) pointed out that the trade-offs inherent within choices made to minimize potential threats to the validity of the design strongly influence the strength of the design (i.e., its ability to detect the phenomenon of interest). For example, they explained that choices made to increase internal validity often reduce the generalizability (i.e., external validity) of the study. We used the Shadish et al. (2002) framework to categorize validity threats (internal, external, construct, and statistical conclusion) and the specific validity threats within each category (e.g., selection threats to internal validity).

The third category of the QF (Figure 1) aligned exactly with Hallinger's (2011) *Valid and Reliable Instrumentation*. Using the categories defined by Urbina (2004), we considered the validity of instruments based on whether the study reported evidence to support the content, construct, concurrent criterion, predictive criterion, convergent, and/or discriminant validity for each measure. We also considered whether the study reported internal consistency, alternate forms, split half, test-retest, and/or inter-rater reliability for each measure. Each of the three QF categories (Theoretical Connections, Design and Validity, Valid and Reliable Instrumentation) consist of specific types of key information and an assigned number of quality points. The following sections describe these categories and their point values in more detail.

Theoretical connections. The Theoretical Connections category consisted of the degree of literature support and connections to theoretical frameworks. Hallinger (2011) found the lack of explicit connections to existing theory to be a consistent weakness in dissertations. Congdon and Dunham (1999) considered such theoretical connections to be critical for research papers in general, and Shakeshaft (1982) asserted that such connections are also critical for dissertations. “Because of the quantity of research already done and the expectation that this is an inquiry ripe for further research, it becomes important to synthesize what has been undertaken, not only to know what has been done, but also to see in what direction the research is moving” (p. 5).

We applied a set of definitions to minimize the subjectivity of our coding of the quality of theoretical connections, and all coding was cross-checked by a second coder to ensure that the definitions were applied consistently. A paper was considered “well grounded in literature” and received two points if it presented evidence for how the literature base guided the development of the purpose. It was considered “partially grounded” and received one point if the paper presented a literature base but did not explicitly connect it to the development of the purpose. If it did not present a literature base, the paper was considered “not grounded” and received zero points.

Within Theoretical Connections, points were also assigned for the degree of connections to one or more theoretical frameworks. A “well connected” paper (2 points) presented a theoretical framework and described how the framework guided the purpose/procedures and, for research papers, the interpretation of results. A "partially connected" paper (1 point) presented a theoretical framework and either how it guided the development of the purpose/procedures, and for research papers, the interpretation of results (but not both). A "not connected" paper (0 points) either did not present a theoretical framework or presented one but did not describe how it guided the development of the purpose/procedures nor was it used to interpret results. Figure 1 summarizes the point breakdown for Theoretical Connections.

Design and validity. Points for the other two categories, *Design and Validity* and *Valid and Reliable Instrumentation*, were assigned based on the presence or absence of particular characteristics (0 points if absent, 1 point if present). The characteristics identified under *Design and Validity* were based on the design frameworks presented by Shavelson and Towne (2002), Shadish et al. (2002), Teddlie and Tashakkori (2009), Creswell (2007), and Patton (2002). We first decided whether the study reported on quantitative data, qualitative data, or both (i.e., mixed methods) as described by Teddlie and Tashakkori (2009). For each type of study, we looked for an explicit purpose statement as well as research questions or hypotheses based on Shavelson and Towne's (2002) first principle, posing questions that can be investigated empirically. Designs to collect quantitative data were coded according to Shadish et al.'s (2002) categories: single-group quasi-experiments, multi-group quasi-experiments, randomized experiments, and regression discontinuity designs. Quantitative papers could receive up to three points for the design. Studies using a control group received one point. Studies conducting true (randomized) experiments or regression discontinuity designs each received 2 additional points. Quasi-experiments were also able to receive the full three points: Random and purposive (probabilistic) sampling strategies were assigned 2 points, and if they included a control group, they scored at the same level as true experiments (3 points). Although we recognize that such a strategy may overrate the quality of design for some quasi-experiments, we preferred such an outcome over the possibility of underrating strong quasi-experimental designs. Qualitative designs were divided into six categories based on the frameworks presented by Creswell (2007) and Patton (2002): biography, narrative/historical, phenomenology, grounded theory, ethnography, and case study. Because each qualitative design has inherent strengths and weaknesses for particular settings and none is considered inherently superior, all qualitative designs were assigned 1 point. Studies for which the qualitative design could not be determined were assigned 0 points.

Teddlie and Tashakorri (2009) described the importance of sampling techniques as a component of the design. They divided sampling strategies into probability and non-probability sampling, and they described potential justifications for using any particular strategy except for convenience sampling. Convenience and unclear sampling strategies were therefore assigned one point whereas other sampling strategies received two points.

To consider whether threats to validity were addressed in a study, we applied the four categories described by Shadish et al. (2002): internal, external, construct, and statistical conclusion threats to validity. By definition, statistical conclusion threats were considered not applicable to qualitative studies. We recognized that authors could address a threat to validity without explicitly stating that they did so or that they could use different terminology than the Shadish et al. (2002) categories; such studies received the full point for each validity threat addressed.

Valid and reliable instrumentation. We considered the validity of instruments based on whether the study reported evidence to support the content, construct, concurrent criterion, predictive criterion, convergent, and/or discriminant validity for each measure (as defined by Urbina, 2004). We also considered whether the study reported internal consistency, alternate forms, split half, test-retest, and/or inter-rater reliability for each measure. For qualitative studies, we considered whether the study addressed one or more of the ten strategies for establishing trustworthiness described by Patton (2002): persistent observation, triangulation, peer debriefing, negative case analysis, referential adequacy, member checks, thick description, dependability audit, confirmability audit, or reflective journaling.

Methodology

This section shares the paper selection criteria for our sample (dissertations), techniques and sources of the systematic literature search we used to find appropriate papers for review, the process and tool used in coding those papers, and the analysis of the results. This paper is part of a larger study that analyzed dissertations, journal articles, book chapters, full books, reports, and master's theses. All papers were coded using the QF and the protocol described above. The present study's focus on dissertations provides a unique window into mathematics technology education research.

Literature Search

We identified dissertations for the present study as part of a larger literature search, in which we followed a systematic process based on the techniques outlined by Cooper, Hedges, and Valentine (2009) and Lipsey and Wilson (2001); for example, defining constructs prior to coding, defining keywords prior to conducting the literature search, defining a coding process, training coders, and cross-checking results. To obtain the overall sample, a wide array of databases was searched, using terms to restrict the sample, based on two inclusion criteria: (1) The paper must examine a technology-based intervention (e.g., technology, calculators, computers), and (2) The focus of the paper must target the learning of a mathematics concept or procedure (e.g., mathematics, algebra, geometry, visualization, representation). We searched the following database platforms (and the databases within those platforms): EBSCOWeb (ERIC, Academic Search Premier, PsychInfo, Primary Search Plus, Middle Search Plus, Education Administration Abstracts), JSTOR (limited to the following disciplines: Education, Mathematics, Psychology, and Statistics), OVID, ProQuest (Research Library, Dissertations & Theses, Career & Technical Education), and H. W. Wilson Web (Education Full Text). From the papers we identified through this search, we examined the bibliographies to identify potentially relevant papers that were missed in our searches.

Search terms were organized into three categories: technology, mathematics, and education. For technology, we used search terms such as technology, calculate* (* is a wild card), software, Sketchpad, Geogebra, Wingeom, Cabri, TI, digital, dynamic, virtual, applet, web, Excel, spreadsheet, PowerPoint, tablet, computer, podcast, distance learning, CBL, CBR, probe, handheld, hand-held, hand held, visualization, 3d, 3-d, or robot. For education, we used search terms such as education, teach, learn, class, school, student, college, or train*. For mathematics, we used search terms such as math, geometry, algebra, fraction, rational, number, integer, variable, function, equation, expression, calc, probability, statistics, discrete, matrix, coordinate, or transform. Variants of each term were tried (e.g., plural vs. singular, prefixes vs. whole words) to prevent papers from being missed. We also consulted with colleagues to identify any papers that may not have been indexed by a database but might be relevant (e.g., unpublished reports). Because the overall study included a purpose of exploring the way mathematics education technology research has changed over time, we did not set date restrictions on the sample.

For the larger literature search (including papers other than dissertations), we identified 5,488 potentially relevant papers from 1968 to 2009. Based on an initial examination of titles and abstracts of these papers, the sample was reduced to 1,421 potentially relevant papers. At least two authors considered the titles and abstracts prior to their being removed from the sample. Of the 1,421 remaining papers, the full text was examined by at least two authors, resulting in 184 papers being excluded from the sample as not relevant (i.e., did not address mathematics, technology, or education). We also excluded 27 papers that were unavailable or were only available in a non-English language. Within the final sample of 1,210 papers, 480 were dissertations.

Coding Tool and Process

The coding of the 480 dissertations was part of the larger project and began with the development of a coding form to identify the areas of potential interest within the study. The coding tool was refined through an iterative process. First, we piloted the tool on three papers with two coders. Refinements based on the results of this pilot test were examined with all six researchers coding the same, original three papers. This process was repeated through three more iterations of refinement and the coding of 27 more papers (i.e., 30 papers were coded by all six authors).

The final coding tool captured a wide array of information for the larger project: basic study information (e.g., type of paper, purpose, content area, grade level of sample), research questions or hypotheses, frameworks, data sources, types of technology, National Council of Teachers of Mathematics (NCTM, 2000) *Principles*, and outcomes. For papers with a focus on teacher knowledge, TPACK standards and levels were coded. The type of research, design details, measurement reliability and validity/trustworthiness, and threats to validity were coded for all research studies. To examine dissertations for the present study, we focused on the categories within the QF, types of technology, and outcomes. We followed a grounded theory approach to identify the types of technology and outcomes. The set of predetermined criteria and the emergent type categories were formed into a coding framework. We then presented our coding framework at multiple conferences (e.g., Rakes, Wagener, & Ronau, 2010; Ronau, Rakes, Niess, & Wagener, 2010). Based on feedback from the conference participants, we revised labels, and added potential categories not already found in the literature.

To carry out the coding of the sample, each coder was paired with each of the other coders (i.e., six coders = 15 coding teams) so that each study was coded and cross-checked. The new coding design created a counter-balanced design with all six coders, providing a way to maximize construct validity and inter-rater reliability of the coding. To maximize construct validity, we carefully defined the constructs in a coding handbook. The coding for every paper was checked by a second coder to ensure that the coding adhered to the agreed-upon definitions.

This double-coding process also allowed us to maximize our inter-rater reliability. All coders were trained to follow a prescribed process for discussing issues raised by the second coder, and common issues were recorded and discussed by the pair and by the full team at regular bi-weekly or monthly full-group meetings. With 26 categories to be agreed upon for each paper and 1,210 papers in the sample, there were a total of 31,460 possible disagreements. Overall, we had 2,684 comments, concerns, questions, and/or disagreements, but there were 28,776 data points with no issues. We therefore computed an inter-rater agreement of 91.5%, from which we concluded that the inter-rater reliability for the overall project was high. For the 480 dissertations, the same 26 categories applied, yielding 12,480 potential disagreements. With 1,188 comments, concerns, questions, and/or disagreements, there were 11,292 data points with no issues. We therefore computed an inter-rater agreement of 90.5%, from which we concluded that the inter-rater reliability for the dissertation subsample was also high. Because our inter-rater agreements revolved around adherence to the codebook definitions, we concluded that our fidelity to those definitions was high and that the construct validity of the coding was also high.

Analytic Methods

The present study identified two broad questions and eight specific sub-questions to guide the analysis of how new mathematics education technology researchers are mentored to identify particular types of technology or content areas, to apply various research methodologies, and to report sufficient information to support study quality. The first broad question examined the scope of mathematics education technology dissertations, specifically the proportion of mathematics education dissertation research focused on technology, the types of technologies studied, and the outcomes found in mathematics education technology dissertations. The second broad question examined the quality of that research, looking specifically at the theoretical connections, design and validity, and reliable and valid measurement through the QF as described. We organized our results according to the research questions.

We adopted a grounded theory approach to identify themes in the sample, allowing the data to provide the technology and outcome categories rather than imposing a pre-determined framework. To identify the types of technologies and outcomes studied in mathematics education technology dissertations and to compute the quality of dissertation research (Question 1 and its sub-questions), the dissertation sample was analyzed using descriptive statistics such as percentages, counts, means, and standard deviations. To determine how outcomes and technology foci have or have not changed over time, regression analyses were conducted. To compare dissertation quality to research published in journal articles and other publication types (books, edited books, book chapters, master's theses, conference proceedings, and reports) and across research types, analysis of variance (ANOVA) and multivariate analysis of variance

(MANOVA) were used (Question 2 and its sub-questions). Because both qualitative and quantitative methods were used to complement one another, we considered the present study to have a mixed methods design following a triangulation design (as in Creswell & Plano Clark, 2007).

Results

The present study sought to examine mathematics education technology dissertations to determine the degree to which new researchers in mathematics education technology are being prepared to choose topics important for advancing the field and to conduct high quality research. The first broad question and its sub-questions focused on describing the current state of the field and how the field has changed over time. Time spans were aggregated to five-year periods to provide sufficient detail to identify patterns while maintaining the readability of tables and figures. The second question and its sub-questions focused on the consistency with which mathematics education technology dissertations reported evidence supporting connections to existing theory, design and validity, and valid and reliable instrumentation as a measure of quality, across publication (dissertations, journals) and research types (qualitative, quantitative, mixed methods).

Scope of Mathematics Education Technology Dissertations

Below we share the results of research question 1 which addressed the scope, of mathematics education technology dissertations, the proportion of this set compared with all mathematics education dissertations, the types of technologies studied, and the types of outcomes pursued by these studies.

Portion of the field. Our inquiry into the scope of mathematics education technology dissertations began by an examination of the number of mathematics education technology dissertations published as a proportion of the total dissertations published in mathematics education, thereby considering the proportion of mathematics education dissertation research devoted to technology. From 1965 to 1989, the proportion of mathematics education dissertations studying technology ranged from 0.0% (0 out of 651 in 1970-74) to 4.8% (16 out of 336 in 1985-89). The introduction of the graphing calculator by Casio (the FX-7000G in 1986) and Texas Instruments (the TI-81 in 1990) provided students with more capabilities for studying mathematics, teachers with more possibilities for exploring complex mathematics in inquiry-based learning, and researchers with more opportunities to study their use in mathematics (as described in Demana & Waits, 1992; Waits & Demana, 1998). From 1990-94, the proportion of dissertations studying technology quadrupled to 20.1% (69 out of 343 mathematics education dissertations) and remained steady throughout the 1990s (100 out of 501 mathematics education dissertations in 1995-1999, 20.0%; Figure 2).

[Insert Figure 2 Here]

In 2000-2004, the proportion of mathematics education technology dissertations increased to approximately 27.3% (117 out of 429 mathematics education dissertations) and remained steady throughout the 2000s (164 out of 607 mathematics education dissertations, 27.0%; Figure 2). From these data, the increased focus on technology in mathematics education is clear. As new technologies emerge to provide more flexibility in how mathematics is taught and learned, new researchers in mathematics education have become increasingly interested in the appropriate ways to use the technology to the greatest effect on a wide array of outcomes. We consider this finding to be an important indicator to the field and not a foregone conclusion. If, for example, we had found that the focus on technology had not increased as it did at particular times, the finding may have indicated that the field is slow to incorporate new technologies. Instead, we found that the field is quick to adapt, and we considered this adaptation to be a positive characteristic.

Technologies. We identified four broad technology categories within the sample dissertations: calculators (n = 175), computer software (n = 268), Internet technologies (n = 112), and other technology (n = 148). These four categories were further broken down into sub-categories (see Table 1). Because some dissertations studied more than one type of technology, the number of technology types addressed was more than the number of dissertations, 703 technology types addressed in the 480 dissertations for an average of 1.5 technology types per dissertation.

[Insert Table 1 about here]

Taking the results from Figure 2 and Table 1 together as a whole, we concluded that the number and types of technology dissertations varied widely over the last four decades but followed a discernible pattern. Few dissertations before 1989 studied technology (Figure 2), and the pre-1989 mathematics education dissertations that did address technology mostly studied non-scientific calculators (see Table 1). We found only one dissertation that studied mathematics education technology before 1975. Not surprising given the technology available at the time, pre-1980 dissertations focused on non-scientific calculators. Graphing calculators started being the focus of calculator studies in the late 1980's and seemed to peak by the end of the 1990's coinciding with the spike in education technology dissertations in that time period, which agrees with the observations of Demana and Waits (1992). In the late 1990's, studies of general graphing calculator use declined while studies of specific advanced features (e.g., symbolic calculus) of the graphing calculator or more advanced graphing calculators (e.g., TI-89) increased. Studies addressing the use of computer software in mathematics education increased also, from an average of less than one per year in the 1980's to an average of almost 10 per year in the 1990's and 16 per year in the 2000's. These patterns of technology types studied are demonstrated in Figure 3.

[Insert Figure 3 about here]

Internet technology studies also showed an increase; however, this change started about a decade after the computer studies, coinciding roughly with the emergence of the technologies in the market. Studies on distance learning dominated the topic of choice with 48 out of a possible 112 studies, or more than half of the Internet studies. Thirty-three of these studies were completed at the collegiate level.

Outcomes. We identified six student outcomes in the sample: achievement, conceptual learning, procedural learning, orientation (i.e., affective domain as in Schoenfeld, 1985, 2011), discernment (i.e., cognitive domain as in Ronau & Rakes, 2011), and learning behavior (as in Battista, 2002). We also found six teacher knowledge outcomes: knowledge of subject matter, knowledge of pedagogy, knowledge of orientation (affective domain), knowledge of discernment (cognitive domain), knowledge of individual context, and knowledge of environmental context (as in Davis, 2007). Our pre-analysis of the set of studies identified three additional teacher outcomes: teacher orientation (affective domain), teaching choices, and professional activities that we added to the coding tool. The pre-analysis also revealed outcomes beyond those for students and teachers such as instrument analysis, theory development, literature reviews, and tool development, which we therefore coded and included in the category *Other Outcomes*. We originally intended to use TPACK as a framework to help analyze teacher knowledge with respect to technology; but our sample contained only three dissertations that addressed TPACK so this analysis is not included in the present study.

The sample of 480 dissertations reported 1,083 outcomes, and multiple outcomes became more frequent than single outcomes in the 1990s and 2000s. The most number of outcomes from a single study in this set of dissertations was eight (*n* = 2 dissertations). A total of 306 dissertations included multiple outcomes, the largest proportion being multiple student outcomes (e.g., student achievement and orientation).

[Insert Table 2 about here]

Student outcomes accounted for 685 (*n* = 203 dissertations) of the 1,083 total number of outcomes, teacher knowledge for 166 outcomes (*n* = 142 dissertations), other teacher outcomes for 197 (*n* = 116 dissertations), and “other” for 35 outcomes (research to practice, instrument analysis, tool development, theory development, literature review; *n* = 27 dissertations). Achievement accounted for 275 of the 685 student outcomes (40.1%), and orientation for 180 outcomes (26.2%). Student discernment (cognitive domain) was studied the least at only 31 outcomes (4.7%). Pedagogical knowledge accounted for the largest proportion of teacher knowledge outcomes, 55 out of the 166 teacher knowledge outcomes (33.1%). Teacher knowledge of orientation (affective domain) accounted for the second highest number of outcomes, 35 of the 166 teacher knowledge outcomes (21.1%). We found this result somewhat surprising, having expected subject matter knowledge to be the most studied given the growing emphasis on mathematics knowledge for teaching over the last two decades (e.g., Ball, 1998; Hill & Ball, 2009; Hill, Schilling, & Ball, 2004).

There is growing recognition that teacher knowledge outcomes should be coupled with student outcomes to assess the degree that improvement in the teacher results in improved student learning (Chow, 2012). We therefore examined how many dissertations made such connections for mathematics education technology. Of the 80 dissertations that examined teacher knowledge, connections to student outcomes were made in 27 (33.8%). Of the 134 dissertations that examined teacher behavior outcomes, 45 (33.6%) also addressed student outcomes.

Using simple linear regression, we investigated how the number of studies addressing both teacher and student outcomes changed over time, both in terms of the number of teacher-with-student outcomes, and in terms of the proportion of the number of teacher outcomes addressed in each year. The number of teacher-with-student outcomes increased between 1968 and 2009, $b = 0.054$, $t(42) = 3.297$; however, the proportion of teacher-with-student outcomes out of the number of teacher outcomes addressed each year remained stable between 1968 and 2009, $b = 0.005$, $t(42) = 1.821$. We therefore concluded that although the field has begun to recognize the importance of making connections between student and teacher outcomes, more effort is needed to train new researchers on the importance of making such connections.

Quality of Mathematics Education Technology Dissertations

The five sub-questions for Research Question 2 focus on the quality of mathematics education technology dissertations in terms of consistency over time, comparability to research published in journals, and differences across research. To study these questions, we applied our QF, which was informed by Hallinger's (2011) categories. Quality measures were computed from the QF as percentages of the number of points possible for the relevant design: 6 points for literature reviews and theory development papers, 12 for qualitative designs, 15 for quantitative designs, and 17 for mixed methods (see Figure 1). The "Other" category shown in Figure 1 was not used in this analysis because papers using Other research types were not relevant to this analysis. Additionally, literature reviews and theory development papers did not have scores for the Reliable and Valid Measures category, and thus, they are also not included in this analysis.

Table 3 presents the average percent of possible QF points earned for Theoretical Connections, Design and Validity, and Reliable and Valid Measures by publication and research types over time.

[Insert Table 3 Here]

We used a multivariate analysis of variance (MANOVA) to examine potential differences in percent of quality points earned. The dependent variables for the MANOVA were the three QF categories (Theoretical Connections, Design and Validity, and Reliable and Valid Measures); the fixed factors were Time Span (five-year intervals), Research Type (quantitative, qualitative, mixed methods), and Publication Type (dissertations and journals). Because the three dependent variables were highly collinear with the overall quality measure (R^2 = .992), we did not include the overall quality measure as a dependent variable or run separate analyses for it, following the advice of Stevens (2012) for avoiding redundancy in such a situation. Time spans prior to 1985 were not included in the analysis due to insufficient sample sizes for meaningful comparisons.

Statistical assumptions. To address possible threats to statistical conclusion validity, we examined the three assumptions of MANOVA outlined by Stevens (2012): independence of observations, multivariate normality, and homogeneity of covariances (*sphericity*). We determined that the observations were independent because each observation was based on a unique study, and we did not include in the sample dissertations that were subsequently published in journal articles.

A statistical test for multivariate normality is not currently available (Stevens, 2012), so we tested univariate normality, a necessary condition for multivariate normality, for each of the 74 groups (3 dependent variables by 2 publication types by 3 research types by 6 time spans = 108 potential groups with 34 groups with no data; e.g., no mixed methods dissertations or journals in 1985-1989) with the Shapiro-Wilk test of normality (*W*). Of the 74 groups, two (2.7%) indicated statistically significant deviation from the normal distribution at the .05 level, 11 (14.9%) at the .01 level, and 38 (51.4%) at the .001 level. Overall, 51 of the 74 groups (68.9%) indicated statistically significant deviation from the normal distribution; we therefore concluded that we could not assume multivariate normality. Because MANOVA has been found to be robust for Type I error but attenuates statistical power (i.e., increased Type II error, Stevens, 2012), we did not consider non-normality to be an issue unless we found non-significance.

For the third assumption (homogeneity of covariance), Box's M test was statistically significant, $M = 241.089$, $p < .001$, indicating heterogeneous covariance. Because the sample sizes across the 74 groups were unequal (ranging from one study to 72 studies), we followed Stevens' (2012) advice and identified the largest and smallest covariance matrix determinants. The largest determinant was for quantitative studies in journals published between 2005 and 2009 ($n = 42$), $det = 0.00057$; the smallest determinant was for qualitative studies published in journals between 1995 and 1999 ($n = 3$), $det = -1.00 \times 10^{-9}$. According to Stevens (2012), having the larger determinant with the larger sample size, as is the case here, results in a conservative MANOVA test (i.e., increased Type II error). Because two of the three statistical assumptions indicated increased Type II error, we noted the need to consider interpreting near-significant results, which we defined as having $p < .10$.

MANOVA results. Wilks' Lambda revealed multivariate significance for the main effects of Research Type and Publication Type (Table 4). No non-significant results were near significant ($p > .5$).

[Insert Table 4 Here]

We therefore considered only the between-subjects effects for Research Type and Publication Type to avoid Type I error. We found that Research Type and Publication Type were statistically significant for all three dependent variables (Table 5).

[Insert Table 5 Here]

Although time span approached statistical significance for Theoretical Connections (Table 5), we chose not to interpret the main effect of Time Span for Theoretical Connections because the multivariate effect of Time Span indicated no statistical significance ($p > .5$; Table 4).

Instead, we focused on statistically significant differences in research and publication types on each of the three dependent variables. For Theoretical Connections, we found that qualitative research earned a higher percent of QF points than both quantitative and mixed methods research (Table 6). The magnitude of difference (standardized mean difference effect size, Cohen's d) was small ($|d| \leq 0.3$ standard deviations; Lipsey & Wilson, 2001) between qualitative and mixed methods and medium ($0.3 < |d| \leq 0.67$ standard deviations; Lipsey & Wilson, 2001) between qualitative and quantitative research.

[Insert Table 6 Here]

For Design and Validity, quantitative research earned a higher percent of QF points than qualitative research, which earned a higher percentage than mixed methods research. The magnitude of difference was large ($|d| > 0.67$ standard deviations; Lipsey & Wilson, 2001) between quantitative and qualitative and between qualitative and mixed methods research. The magnitude of difference between quantitative and mixed methods was small. For Reliable and Valid Measures, quantitative research earned a higher percent of QF points than qualitative research, but no differences were indicated for mixed methods research from quantitative or qualitative research. The magnitude of this difference was medium.

Dissertations earned a statistically significant higher percent of QF points than journals for all three dependent variables (Table 6; Figure 4). The magnitudes of these differences were large for all three dependent variables.

[Insert Figure 4 Here]

Discussion

By examining the scope of the research (Research Question 1), we found that new researchers have consistently focused on emerging technologies for their dissertations. The number of technology studies for some types of software, such as tutorial software, was high in number but low in replication, as indicated in the Other Technology and the Miscellaneous subcategory (Table 1). Moreover, a number of graphing calculator studies left the type of calculator itself unspecified, making a determination of whether the findings are applicable to a particular calculator difficult. Such conditions limit researchers' ability to synthesize dissertation findings, thereby limiting the ability of dissertations to advance the literature base to the fullest degree.

Attention to how teacher outcomes influence student outcomes has been slower to take hold, as shown by the near zero growth in the number of teacher outcomes connected to student outcomes and in their proportion of the total number of teacher outcomes studied

From the present study, essentially a survey of dissertation research in mathematics education technology, we make inference about the type of guidance doctoral students are receiving with regard to research quality. By comparing the quality of dissertation research (new researchers) to research published in journals (ostensibly by more experienced researchers), we found that new researchers report information about Theoretical Connections, Design and Validity, and Reliable and Valid Measures more than experienced researchers (Table 3). Even so, the percentage of overall QF points earned for any five-year period never exceeded 66.5%, less than two-thirds of the components called for by the field, and the percentage of QF points earned has remained relatively stable over this time period. That dissertation research scored higher than journals was not especially surprising, given the lack of space constraints and the close supervision of experienced researchers in the dissertation process. The magnitude of the difference, however, was quite surprising to us (Table 6). Why do experienced researchers not report the components that they mentor their students to address? Although space constraints may explain some of the lack of important details in journals, the potential causes for such a divergence are unclear from the available data. From our own experiences developing the present study, we recognize that important details are easy to leave out of a report even though they were attended to. We found that our external readers were extremely helpful in recognizing what we had missed.

The QF provided a useful structure for analyzing the quality of dissertations that synthesized research on what constitutes important information for quality, but we note that it represents a bare minimum of standards for quality. For the present study, we chose to emphasize reliability and validity, an emphasis lacking in previous dissertation syntheses, at the expense of looking more in-depth and drawing judgments about the appropriateness of a strategy. We consider this trade-off to be worthwhile, laying the foundation for future analyses to add depth to the QF criteria. Researchers who wish to apply this tool to future studies should attend to several features of the present study to maximize its validity.

First, we explicitly defined the constructs in a coding handbook to strengthen construct validity. A coding design was used so that each coder was paired with each of the other coders on the research team to strengthen the validity and reliability of the coding results. Second, coders were trained to follow a prescribed process for discussing issues raised by the second coder, and common issues were recorded and discussed by the pair and by the full team at regular bi-weekly or monthly full-group meetings. Third, if a more detailed quality analysis is desired (e.g., how well a quality characteristic was addressed), careful attention to avoiding bias in assigning point values must be given. For example, we were unable to find guidance for weighting characteristics such as neglecting relevant types of reliability versus relevant types of validity. The number of outcomes, data sources, measures, frameworks, or research questions needed to be considered optimal has not been studied to our knowledge (i.e., more is not necessarily better). In the present study, dissertations in our sample varied in number of research questions between one and 41.

In its present form the QF proved to be a useful tool to analyze the quality of studies on the basis of scientific reporting. We were able to synthesize our coding results to provide a detailed profile of the quality of dissertations. Moreover, our results on the quality of dissertations were similar to that of previous researchers such as Shakeshaft (1982).

The analysis of dissertation quality by research type provided an only slightly different perspective. Qualitative studies addressed theoretical connections more consistently than either quantitative or mixed methods, giving them a greater ability to enhance the quality of the field and to help researchers build new studies on a solid foundation. Quantitative studies reported design and validity issues more consistently than both qualitative and mixed methods studies. Additionally, quantitative studies reported measure reliability and validity information more consistently than qualitative studies.

In general, our results indicated that the mathematics education technology research community should demand more of itself. Doctoral students as well as their mentors can play a pivotal role in such improvement by (a) attending to the QF categories as they plan, design, implement, and complete their dissertation studies and (b) demanding that dissertations represent a minimum bar for quality rather than a maximum. The mathematics education technology research community must in turn begin to demand greater quality in its published studies, both through how researchers write about their own studies and how they review the works of others. By insisting on accepting only papers that explicitly include key information necessary to support the scientific basis of a study, as illustrated by the QF, the research community would not only improve the quality of the literature in the field, but also would provide better examples for novice researchers as they plan their initial studies.

References

Adams, G. B., & White, J. D. (1994). Dissertation research in public administration and cognate fields: An assessment of methods and quality. *Public Administration Review, 54*, 565-576.

Association of Mathematics Teacher Educators. (2009). *Mathematics TPACK (Technological Pedagogical Content Knowledge) Framework.* A position paper adopted by the Board of Directors June 19, 2009. Retrieved from http://amte.net/sites/all/themes/amte/resources/MathTPACKFramework.pdf

Auerbach, S. (2011). "It's not just going to collect dust on a shelf:" Faculty perceptions of the applied dissertation in the new California State University (CSU) Ed.D. programs leadership education from within a feminist ethos. *Journal of Research on Leadership Education, 6*, 59-82.

Ball, D. L. (1988). Research on teaching mathematics: Making subject matter knowledge part of the equation. East Lansing, MI: National Center for Research on Teacher Education. (Retrieved from http://www.eric.ed.gov/PDFS/ED301467.pdf)

Battista, M. T. (1999). The mathematical mis-education of America's youth: Ignoring research and scientific study in education. *Phi Delta Kappan, 80*, 424-433.

Battista, M. T. (2002). Learning geometry in a dynamic environment. *Teaching Children Mathematics, 8*, 333-337.

Britt, D. W. (1997). *A conceptual introduction to modeling: Quantitative and qualitative perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

Chow, W. Y. (2012). Effective teachers and effective teaching. In E. Albro (Ed.), *Request for applications: Education research grants. CFDA 305A* (pp. 18-22). Washington, DC:

Institute of Education Sciences, National Center for Education Research. Retrieved from http://ies.ed.gov/funding/pdf/2013_84305A.pdf

Colwell, R. J. (1969) *A critique of research studies in music education*. Urbana, IL: Illinois University. Retrieved from www.eric.ed.gov/ERICWebPortal/detail?accno=ED035100

Common Core State Standards for Mathematical Practice. (2010). *Mathematics→Standards for mathematical practice* [Website]. Retrieved from http://www.corestandards.org/Math/Practice

Congdon, J. D., & Dunham, A. E. (1999). Defining the beginning: The importance of research design. *IUCN/SSC Marine Turtle Specialist Group*, 4, 1-5. Retrieved from http://mtsg.files.wordpress.com/2010/07/14-defining-the-beginning.pdf.

Conn, V. S. (2008). The light under the bushel basket: Unpublished dissertations. *Western Journal of Nursing Research, 30*, 537-539.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York, NY: Russell Sage Foundation.

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

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Davis, B. (2007). Learners within contexts that learn. In T. Lamberg & L. R. Wiest (Eds.), *Proceedings of the 29th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 19-32). Stateline (Lake Tahoe), NV: University of Nevada, Reno.

Demana, F., & Waits, B. K. (1992). A computer for all students. *Mathematics Teacher, 85,* 94-95.

Ellington, A. J. (2003). A meta-analysis of the effects of calculators on students' achievement and attitude levels in precollege mathematics classes. *Journal for Research in Mathematics Education, 34,* 433-463.

Ellington, A. J. (2006). The effects of non-CAS graphing calculators on student achievement and attitude levels in mathematics: A meta-analysis. *International Journal of Instructional Media, 106*, 16-26.

Hallinger, P. (2011). A review of three decades of doctoral studies using the Principal Instructional Management Rating scale: A lens on methodological progress in educational leadership. *Educational Administration Quarterly, 47*, 271-306.

Hembree, R., & Dessart, D. J. (1986). Effects of hand-held calculators in precollege mathematics education: A meta-analysis. *Journal for Research in Mathematics Education, 17*, 83-99.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal, 105,* 11-30.

Hill, H., & Ball, D. (2009). The curious - and crucial - case of mathematical knowledge for teaching. *Phi Delta Kappan, 91,* 68-71.

Hilmer, M. J., & Hilmer, C. E. (2011). Is it where you go or who you know? On the relationship between students, Ph.D. program quality, dissertation advisor prominence, and early career publishing success. *Economics of Education Review, 30,* 991-996.

Hoffer, T. B., Welch, V., Webber, K., Williams, K., Lisek, B., Hess, M, Loew, D., & Guzman-Barron, I. (2009). *Doctorate recipients from United States universities: Summary Report 2005.* Washington, DC: National Science Foundation, National Institutes of Health, U.S.

Department of Education, National Endowment for the Humanities, U.S. Department of Agriculture, National Aeronautics and Space Administration. Retrieved from http://www.nsf.gov/statistics/doctorates/pdf/sed2005.pdf

Karadağ, E. (2011). Instruments used in doctoral dissertations in educational sciences in Turkey: Quality of research and analytical errors. *Educational Sciences: Theory & Practice, 11*, 330-334.

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

Lovitts, B. E. (2008). The transition to independent research: Who makes it, who doesn't, and why. *The Journal of Higher Education, 79*, 296-325.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*, 1017-1054.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press. Retrieved from http://www.nap.edu/catalog.php?record_id=9822

National Science Foundation Survey of Earned Doctorates Tabulation Engine. (2010). *NSF Survey of Earned Doctorates (SED) tabulation engine.* [Website]. Retrieved from https://ncses.norc.org/NSFTabEngine/#TABULATION

Niess, M. L. (2005). Preparing teachers to teach science and mathematics with technology: Developing a technology pedagogical content knowledge. *Teaching and Teacher Education, 21*, 509-523.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Rakes, C. R. (2012, February). Research in mathematics educational technology: Study overview. In C. R. Rakes (Chair), *A structured inquiry of research in mathematics educational technology: Findings and implications*. Symposium presented at the meeting of the Association of Mathematics Teacher Educators, Fort Worth, TX.

Rakes, C. R., Wagener, L., & Ronau, R. N. (2010, January). *New directions in the research of technology-enhanced education.* Paper presented at the annual meeting of the Association of Mathematics Teacher Educators, Irvine, CA.

Ronau, R. N., & Rakes, C. R. (2011). Aspects of teacher knowledge and their interactions: A comprehensive framework for research. In R. N. Ronau, C. R. Rakes, & M. L. Niess (Eds.), *Education technology, teacher knowledge, and classroom impact: A research handbook on frameworks and approaches.* Hershey, PA: IGI Global.

Ronau, R. N., Rakes, C. R., Bush, S., Driskell, S. O., & Pugalee, D. (2012, February). Dissertation studies in mathematics education technology: 1968-2009. In C. R. Rakes (Chair), *A structured inquiry of research in mathematics education technology: Findings and implications*. Symposium presented at the meeting of the Association of Mathematics Teacher Educators, Fort Worth, TX.

Ronau, R. N., Rakes, C. R., Niess, M. L., & Wagner, L. (2010, April). *Research in mathematics instructional technology: Current trends and future demands.* Paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

Rutkienė, A., & Teresevičienė, M. (2010). Improvement of experiment planning as an important precondition for the quality of education research. *The Quality of Higher Education, 2010*, 88-107.

Schoenfeld, A. H. (1985). *Mathematical problem solving.* Orlando, FL: Academic Press.

Schoenfeld, A. H. (2006). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher*, *35*, 13-21.

Schoenfeld, A. H. (2011). *How we think: A theory of goal-oriented decision making and its educational applications*. New York, NY US: Routledge/Taylor & Francis Group.

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

Shakeshaft, C. (1982, March). *A framework for studying schools as work settings for women leaders.* Paper presented at the annual meeting of the American Education Research Association, New York, NY. Retrieved from http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED216441&ERICExtSearch_SearchType_0=no&accno=ED216441

Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, D.C.: National Research Council, National Academy Press.

Shirmer, B. R. (2008). How effectively are we preparing teacher educators in special education? The case of deaf education. *American Annals of the Deaf*, *153*, 411-419.

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences.* (5th ed.). New York, NY: Taylor & Francis Group.

Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction. *Phi Delta Kappan*, *79*, 14-21.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating* *quantitative and qualitative approaches in the social and behavioral sciences*. Los Angeles, CA: Sage Publications.

Thanheiser, E., Ellis, A., & Herbel-Eisenmann, B. (2012). From dissertation to publication in JRME. *Journal for Research in Mathematics Education*, *43*, 144-158.

Thurgood, L., Golladay, M. J., & Hill, S. T. (2006). *U.S. doctorates in the twentieth century* (NSF 06-319). Washington, DC: National Science Foundation, Division of Science Resources Statistics. Retrieved from http://www.nsf.gov/statistics/nsf06319/pdf/nsf06319.pdf

Urbina, S. (2004). *Essentials of psychological testing*. Hoboken, NJ: John Wiley & Sons.

Waits, B. K., & Demana, F. (1998). *The role of graphing calculators in mathematics reform*. Columbus, OH: Ohio State University. Retrieved from http://www.math.osu.edu/~waits.1/papers/roleofgraphcalc.pdf

Table 1

Number of Technologies Addressed by Mathematics Education Technology Dissertations over Time

Technology Type	1965-69 (*n* = 1)	1970-74 (*n* = 0)	1975-79 (*n* = 8)	1980-84 (*n* = 5)	1985-89 (*n* = 16)	1990-94 (*n* = 69)	1995-99 (*n* = 100)	2000-04 (*n* = 117)	2005-09 (*n* = 164)	Total (*n* = 480)
Calculators	**0**	**0**	**8**	**5**	**5**	**28**	**51**	**40**	**38**	**175**
Non-Scientific	0	0	8	4	3	6	5	3	5	34
Scientific	0	0	0	1	1	3	4	4	1	14
Graphing (Type Specified)	0	0	0	0	0	6	14	7	12	39
Graphing (Type Not Specified)	0	0	0	0	1	10	19	15	12	57
Advanced Features	0	0	0	0	0	3	9	11	8	31
Computer Software	**0**	**0**	**0**	**0**	**10**	**49**	**53**	**73**	**83**	**268**
Geometry	0	0	0	0	1	1	8	15	12	37
Tutorial	0	0	0	0	3	16	23	30	37	109
Spreadsheet	0	0	0	0	1	3	4	3	7	18
Algebra	0	0	0	0	0	8	3	8	8	27
Graphing	0	0	0	0	3	8	5	1	1	18
Statistics & Statistics Instruction	0	0	0	0	0	4	1	4	3	12
Games and Puzzles	0	0	0	0	0	6	4	4	3	17
Presentation	0	0	0	0	0	1	4	4	9	18
Testing	0	0	0	0	2	1	1	2	3	9
Applets (Computer-Based)	0	0	0	0	0	1	0	2	0	3
Internet	**0**	**0**	**0**	**0**	**2**	**3**	**9**	**34**	**64**	**112**
Distance Learning	0	0	0	0	2	1	6	13	26	48
WebQuests, Web Sites, Wikis	0	0	0	0	0	0	1	10	10	21
Virtual Manipulatives	0	0	0	0	0	0	1	2	7	10
Social Networking	0	0	0	0	0	0	1	5	7	13
Testing/Tutorial/Puzzle	0	0	0	0	0	1	0	3	11	15
Applets (Internet-Based)	0	0	0	0	0	1	0	1	3	5
Other Technology Tools	**1**	**0**	**0**	**0**	**8**	**23**	**34**	**26**	**56**	**148**
General Technology Use	0	0	0	0	0	1	3	2	15	21
Calculator Presence	0	0	0	0	1	3	2	0	0	6
Computer Presence	0	0	0	0	4	13	9	5	5	36
Computer Programming	0	0	0	0	3	5	5	1	0	14
Interactive Whiteboard	0	0	0	0	0	0	0	0	9	9
Probeware	0	0	0	0	0	0	4	4	1	9
Online Courseware	0	0	0	0	0	0	0	3	3	6
Email	0	0	0	0	0	0	1	3	1	5
Miscellaneous[a]	1	0	0	0	0	1	10	8	22	42
5-Year Span Total	**1**	**0**	**8**	**5**	**25**	**103**	**147**	**173**	**241**	**703**

Note. Bolded rows indicate Technology Type totals.

[a] Miscellaneous represents Other technology tools addressed 4 times or less (e.g., Blackboard®, chat rooms, digital curriculum materials, and document sharing).

Table 2

Number of Outcomes for Mathematics Education Technology Dissertation Studies over Time

Outcome	1965-69 (*n* = 1)	1970-74 (*n* = 0)	1975-79 (*n* = 8)	1980-84 (*n* = 5)	1985-89 (*n* = 16)	1990-94 (*n* = 69)	1995-99 (*n* = 100)	2000-04 (*n* = 117)	2005-09 (*n* = 164)	Outcome Total
Student	**1**	**0**	**14**	**8**	**17**	**102**	**142**	**173**	**228**	**685**
Achievement	1	0	8	5	9	39	57	65	91	275
Learning (Conceptual)	0	0	1	0	1	14	23	26	21	86
Learning (Procedural)	0	0	0	0	0	5	11	10	9	35
Orientation	0	0	4	3	4	30	36	41	62	180
Discernment	0	0	0	0	0	5	5	9	12	31
Learning Behavior	0	0	1	0	3	9	10	22	33	78
Teacher Knowledge	**0**	**0**	**0**	**0**	**2**	**28**	**46**	**38**	**52**	**166**
of Subject Matter	0	0	0	0	1	4	9	8	7	29
of Pedagogy	0	0	0	0	0	10	15	13	17	55
of Discernment	0	0	0	0	0	3	5	4	8	20
of Orientation	0	0	0	0	1	7	9	7	11	35
of Individual	0	0	0	0	0	1	4	2	3	10
of Environment	0	0	0	0	0	3	4	4	6	17
Teacher Orientation and Behavior	**0**	**0**	**0**	**1**	**6**	**39**	**48**	**44**	**59**	**197**
Orientation	0	0	0	1	3	24	30	24	28	110
Teaching Choices/Practice	0	0	0	0	0	13	13	14	24	64
Professional Activities	0	0	0	0	0	1	4	4	4	13
Classroom Activity	0	0	0	0	3	1	1	2	3	10
Other Outcomes	**0**	**0**	**0**	**1**	**4**	**6**	**6**	**6**	**12**	**35**
Research to Practice	0	0	0	0	1	0	1	0	0	2
Instrument Analysis	0	0	0	0	3	5	3	5	8	24
Theory Development	0	0	0	0	0	0	1	0	0	1
Literature Review	0	0	0	0	0	0	1	0	0	1
Author Opinion	0	0	0	0	0	0	0	1	0	1
Tool Development	0	0	0	0	0	0	0	0	1	1
Miscellaneous Outcomes	0	0	0	1	0	1	0	0	3	5
Total Outcomes	**1**	**0**	**14**	**10**	**29**	**175**	**242**	**261**	**351**	**1,083**

Note. Values represent the number of relevant outcomes. Bolded rows indicate Outcome totals.
[a]*n* represents the number of dissertations published during the five-year span.

Table 3

Percent of Possible Points for Overall Quality, Theoretical Connections, Design and Validity, and Reliable and Valid Measures, by Research and Publication Type over Time

	Overall Quality										**Theoretical Connections**									
Research by Publication Type	1965-69	1970-74	1975-79	1980-84	1985-89	1990-94	1995-99	2000-04	2005-09	Overall	1965-69	1970-74	1975-79	1980-84	1985-89	1990-94	1995-99	2000-04	2005-09	Overall
Dissertation	**60.0**	**—**	**64.2**	**54.7**	**62.9**	**63.6**	**63.3**	**63.5**	**66.5**	**64.4**	**25.0**	**—**	**56.3**	**35.0**	**68.8**	**67.4**	**75.8**	**75.2**	**80.2**	**74.8**
Quantitative	60.0	—	64.2	54.7	68.5	68.6	65.6	64.4	67.6	66.3	25.0	—	56.3	35.0	75.0	63.6	64.5	66.5	71.9	66.6
Qualitative	—	—	—	—	50.0	52.1	59.3	64.3	62.9	60.9	—	—	—	—	56.3	70.8	86.1	88.6	90.8	85.8
Mixed Methods	—	—	—	—	52.9	62.0	63.5	61.2	67.5	64.2	—	—	—	—	50.0	71.6	82.5	74.2	84.0	79.2
Journal	**—**	**—**	**20.0**	**—**	**57.8**	**42.8**	**45.9**	**44.5**	**50.0**	**47.2**	**—**	**—**	**0.0**	**—**	**58.3**	**37.5**	**45.8**	**54.8**	**59.6**	**55.1**
Quantitative	—	—	20.0	—	57.8	43.3	45.9	41.5	54.0	48.6	—	—	0.0	—	58.3	25.0	43.1	36.4	57.1	47.2
Qualitative	—	—	—	—	—	41.7	44.4	45.6	48.1	46.5	—	—	—	—	—	62.5	58.3	68.1	71.0	68.8
Mixed Methods	—	—	—	—	—	—	47.1	47.1	41.6	44.3	—	—	—	—	—	—	50.0	47.5	40.4	44.2
5-Year Average	**60.0**	**—**	**59.3**	**54.7**	**62.1**	**62.0**	**59.9**	**56.5**	**60.8**	**59.5**	**25.0**	**—**	**50.0**	**35.0**	**67.1**	**65.0**	**70.0**	**67.7**	**73.1**	**69.3**

	Design and Validity										**Reliable and Valid Measures**									
Research by Publication Type	1965-69	1970-74	1975-79	1980-84	1985-89	1990-94	1995-99	2000-04	2005-09	Overall	1965-69	1970-74	1975-79	1980-84	1985-89	1990-94	1995-99	2000-04	2005-09	Overall
Dissertation	**77.8**	**—**	**70.8**	**71.1**	**62.5**	**64.0**	**61.4**	**60.2**	**63.3**	**62.5**	**50.0**	**—**	**50.0**	**20.0**	**54.2**	**55.1**	**45.8**	**51.6**	**50.8**	**50.3**
Quantitative	77.8	—	70.8	71.1	66.7	71.1	68.7	65.6	69.8	68.8	50.0	—	50.0	20.0	63.6	67.1	53.5	55.0	49.3	54.2
Qualitative	—	—	—	—	54.2	47.2	49.4	52.9	48.7	50.1	—	—	—	—	25.0	29.2	35.2	50.0	50.0	43.5
Mixed Methods	—	—	—	—	50.0	61.8	61.7	60.0	65.1	62.5	—	—	—	—	66.7	50.0	44.4	47.9	53.5	49.8
Journal	**—**	**—**	**33.3**	**—**	**63.0**	**46.3**	**49.1**	**43.9**	**49.8**	**47.6**	**—**	**—**	**0.0**	**—**	**33.3**	**33.3**	**31.3**	**22.3**	**28.7**	**26.8**
Quantitative	—	—	33.3	—	63.0	52.8	50.6	48.0	56.9	53.2	—	—	0.0	—	33.3	37.5	30.6	22.7	34.5	30.6
Qualitative	—	—	—	—	—	33.3	38.9	38.9	40.9	39.6	—	—	—	—	—	25.0	33.3	20.8	24.2	22.9
Mixed Methods	—	—	—	—	—	—	50.0	53.0	48.5	50.4	—	—	—	—	—	—	33.3	26.7	20.5	24.4
5-Year Average	**77.8**	**—**	**66.7**	**71.1**	**62.6**	**62.6**	**59.0**	**54.2**	**58.7**	**58.3**	**50.0**	**—**	**44.4**	**20.0**	**50.9**	**53.3**	**43.0**	**40.8**	**43.2**	**43.7**

Note. All values are average percentages. Quantitative includes single subject designs and meta-analyses. Qualitative includes action research and design experiments. Bolded rows indicate Dissertation, Journal, and 5-Year Average Totals.

Table 4

Wilks' Lambda Multivariate Results for MANOVA of Quality by Time Span, Publication Type, and Research Type

Effect	Λ	F	Hypothesis DF	Error DF
Time Span	0.983	0.88	12	1648.595
Research Type	0.905	10.63***	6	1246.000
Publication Type	0.920	18.04***	3	623.000
Time Span * Research Type	0.967	0.89	24	1807.491
Time Span * Publication Type	0.993	0.37	12	1648.595
Research Type * Publication Type	0.992	0.82	6	1246.000
Time Span * Research Type * Publication Type	0.981	0.82	15	1720.23

*$p < .05$; **$p < .01$; ***$p < .001$

Table 5

Between-Subjects Effects of Quality by Time Span, Publication Type, and Research Type

Dependent Variable	Source of Variance	DF	Mean Square	F	p
Theoretical Connections	Time Span	4	.156	2.288	.059
	Research Type	2	.427	6.259	.002
	Publication Type	1	2.424	35.547	< .001
	Time Span * Research Type	8	.100	1.469	.165
	Time Span * Publication Type	4	.006	.081	.988
	Research Type * Publication Type	2	.098	1.443	.237
	Time Span * Research Type * Publication Type	5	.089	1.303	.261
	Error	625	.068		
Design and Validity	Time Span	4	.010	.423	.792
	Research Type	2	.433	18.674	< .001
	Publication Type	1	.541	23.346	< .001
	Time Span * Research Type	8	.020	.872	.540
	Time Span * Publication Type	4	.010	.431	.786
	Research Type * Publication Type	2	.018	.756	.470
	Time Span * Research Type * Publication Type	5	.017	.744	.591
	Error	625	.023		
Reliable and Valid Measures	Time Span	4	.007	.061	.993
	Research Type	2	.344	3.025	.049
	Publication Type	1	1.741	15.282	< .001
	Time Span * Research Type	8	.051	.444	.895
	Time Span * Publication Type	4	.057	.504	.733
	Research Type * Publication Type	2	.049	.431	.650
	Time Span * Research Type * Publication Type	5	.106	.928	.462
	Error	625	.114		

Table 6

Pairwise Comparisons of Quality by Research and Publication Types

Dependent Variable	Type I	Type J	Mean Difference (I-J)	SE	p	Cohen's d
Theoretical Connections	Mixed Methods	Qualitative	-0.100	0.053	.063	-0.208
	Mixed Methods	Quantitative	0.064	0.048	.188	0.411
	Qualitative	Quantitative	0.164	0.041	< .001	0.616
	Dissertation	Journal	0.216	0.039	< .001	0.694
Design and Validity	Mixed Methods	Qualitative	0.113	0.031	< .001	1.013
	Mixed Methods	Quantitative	-0.050	0.028	.074	-0.208
	Qualitative	Quantitative	-0.164	0.024	< .001	-1.125
	Dissertation	Journal	0.116	0.023	< .001	0.851
Reliable and Valid Measures	Mixed Methods	Qualitative	0.104	0.069	.134	0.342
	Mixed Methods	Quantitative	-0.018	0.063	.768	-0.061
	Qualitative	Quantitative	-0.122	0.053	.023	-0.358
	Dissertation	Journal	0.208	0.050	< .001	0.707

Mixed Methods (up to 17 pts)

Quantitative (up to 15 pts)

Theoretical Connections (up to 4 pts)
- **Literature Support (≤ 2 pts)**
  - Well Grounded (2 pts)
  - Partially Grounded (1 pt)
  - Not Grounded (0 pts)
- **Conceptual Framework Connections (≤ 2 pts)**
  - Well Connected (2 pts)
  - Partially Connected (1 pt)
  - Not Connected (0 pts)

Design and Validity (up to 9 pts)
- **Purpose Statement (1 pt)**
- **Research Questions/Hypotheses (1 pt)**
- **Design (up to 3 pts)**
  - Randomized Experiment (2 pts)
  - Regression Discontinuity Design (2 pts)
  - Quasi-Experimental Design with:
    - Sampling Strategies Unclear (1 pt)
    - Convenience Sample (1 pt)
    - Other Sampling Strategies (2 pts)
  - Use of Control Group (1 pt)
- **Threats to Validity Addressed (up to 4 pts)**
  - Internal (1 pt)
  - External (1 pt)
  - Construct (1 pt)
  - Statistical Conclusion (1 pt)

Valid and Reliable Instrumentation (up to 2 pts)
- **Reliability (1 point)**
  - Internal Consistency
  - Split Half
  - Inter-Rater
  - Test-Retest
  - Alternate Forms
- **Validity (1 point)**
  - Content
  - Construct
  - Concurrent Criterion
  - Discriminant
  - Predictive Criterion
  - Convergent

Qualitative (up to 12 pts)

Theoretical Connections (up to 4 pts)
- **Literature Support (≤ 2 pts)**
  - Well Grounded (2 pts)
  - Partially Grounded (1 pt)
  - Not Grounded (0 pts)
- **Conceptual Framework Connections (≤ 2 pts)**
  - Well Connected (2 pts)
  - Partially Connected (1 pt)
  - Not Connected (0 pts)

Design and Validity (up to 5 pts)
- **Purpose Statement (1 pt)**
- **Research Questions/Hypotheses (1 pt)**
- **Design (1 pt)**
  - Biography
  - Phenomenology
  - Historical/Narrative
  - Grounded Theory
  - Ethnography
  - Case Study
- **Threats to Validity Addressed (≤ 3 pts)**
  - Internal (1 pt)
  - External (1 pt)
  - Construct (1 pt)

Valid and Reliable Instrumentation (up to 2 pts)
- **Reliability (1 point)**
  - Internal Consistency
  - Inter-Rater
- **Validity (1 pt)**
  - Persistent Observation
  - Member Checks
  - Triangulation
  - Thick Description
  - Peer Debriefing
  - Dependability Audit
  - Negative Case Analysis
  - Confirmability Audit
  - Referential Adequacy
  - Reflective Journal

Theory Development Papers and Literature Reviews (up to 6 pts)

Theoretical Connections (up to 4 pts)
- **Literature Support (up to 2 pts)**
  - Well Grounded (2 pts)
  - Partially Grounded (1 pt)
  - Not Grounded (0 pts)
- **Conceptual Framework (up to 2 pts)**
  - Well Connected (2 pts)
  - Partially Connected (1 pt)
  - Not Connected (0 pts)

Design and Validity (up to 2 pts)
- **Purpose Statement (1 pt)**
- **Design (1 pt)**
  - Biography
  - Phenomenology
  - Grounded Theory
  - Ethnography
  - Historical/Narrative
  - Case Study

Other (up to 5 pts)

Theoretical Connections (up to 4 pts)
- **Literature Support (up to 2 pts)**
  - Well Grounded (2 pts)
  - Partially Grounded (1 pt)
  - Not Grounded (0 pts)
- **Conceptual Framework (up to 2 pts)**
  - Well Connected (2 pts)
  - Partially Connected (1 pt)
  - Not Connected (0 pts)

Design and Validity (up to 1 pt)
- **Purpose Statement (1 pt)**

Figure 1. Categories included in the Quality Framework (QF) and points available within each category. Quantitative includes meta-analyses and single-subject research. Qualitative includes action research and design experiments. Other includes non-research papers and descriptions of mathematics education technology development

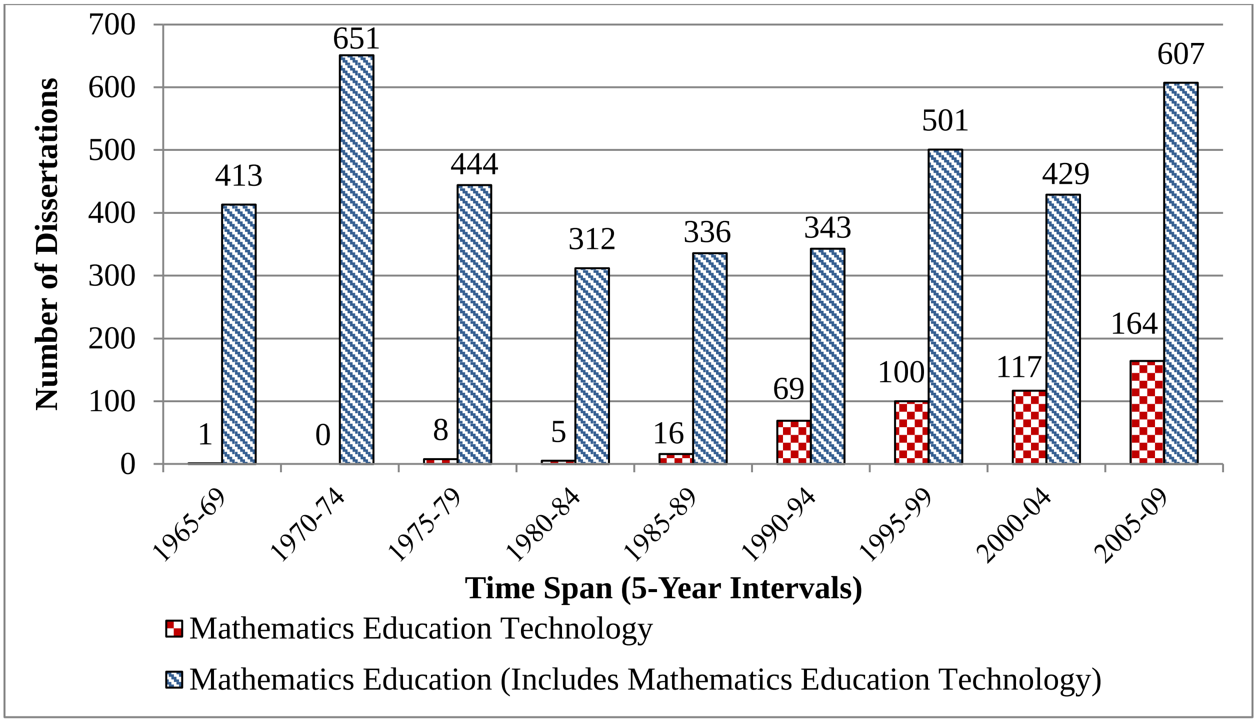

Figure 2. Mathematics education and mathematics education technology doctorates awarded by Time Span.

Note. Mathematics education data from 1965-1995 from Thurgood, Golladay, and Hill (2006). Mathematics education data from 1995-2005 from Hoffer et al. (2009). Mathematics education data from 2006-2009 from National Science Foundation Survey of Earned Doctorates Tabulation Engine (2010).

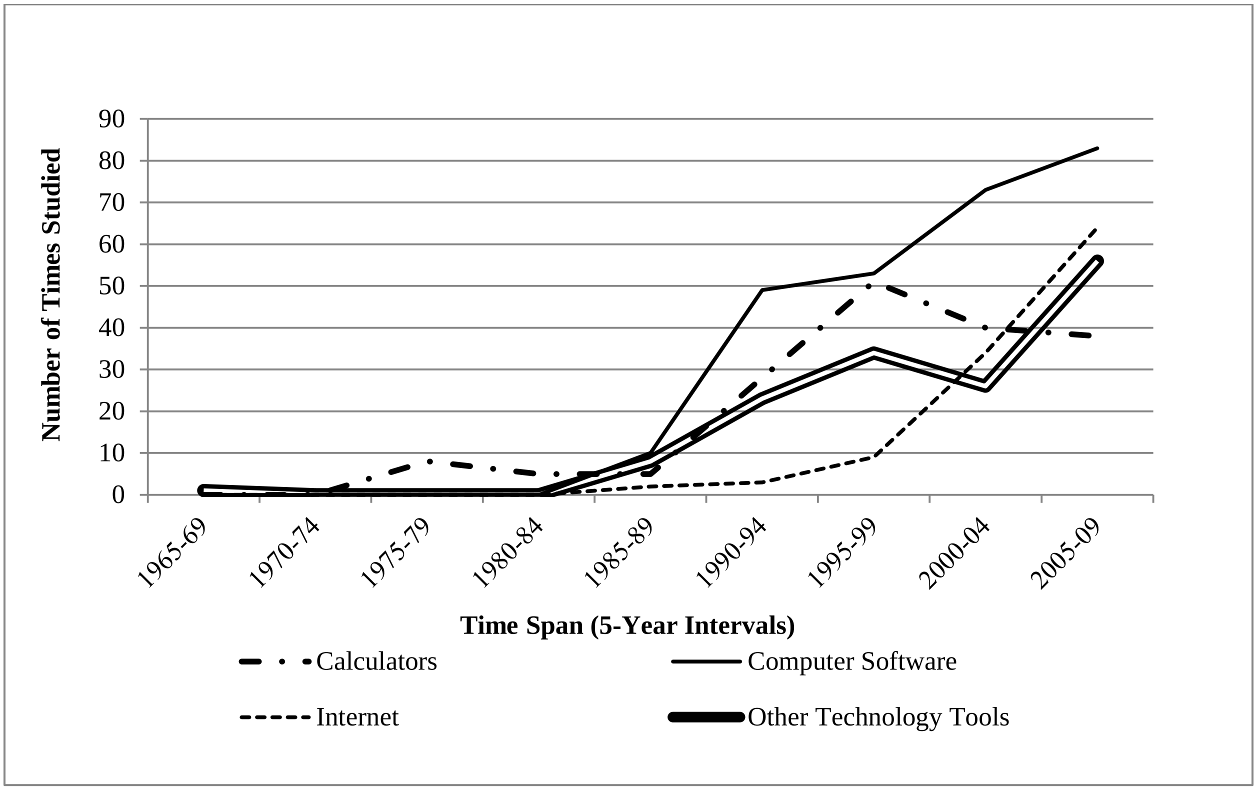

Figure 3. Number of times technology types were studied by dissertations over time.

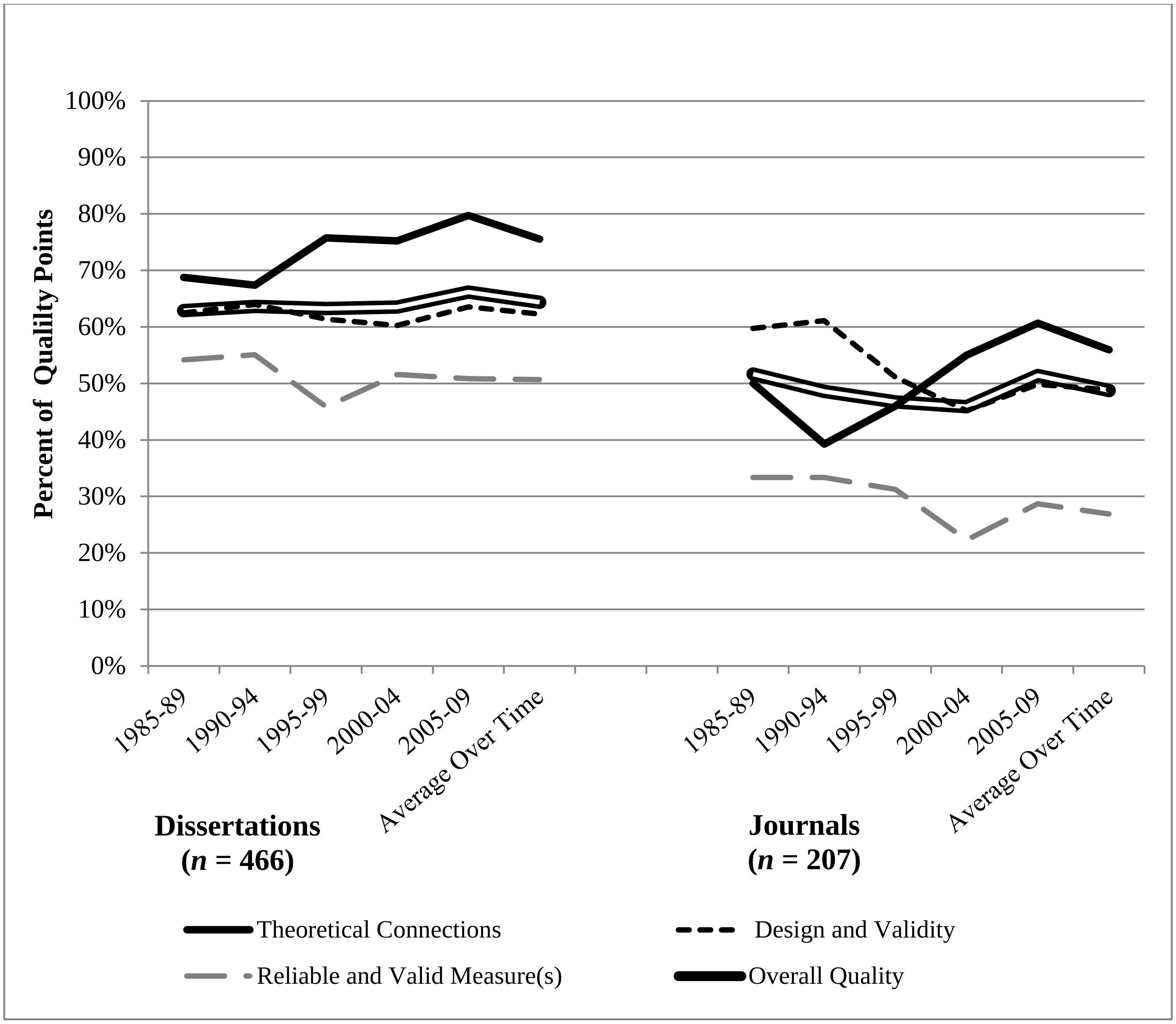

Figure 4. Percent of quality points earned for dissertations and journals by QF category over time. *Note.* Five-year spans for 1965-1984 are not shown due to insufficient sample sizes.